研究日报

2026年5月24日 星期日

武汉光谷 · 阴 · 26°C · 24—28°C

路漫漫其修远兮，吾将上下而求索。

第 1 版 / 7

UrbanComp Lab 学习资料库（https://research.urbancomp.dev/）

THIS EDITION

五个方向的当日进展

Broadening Access to Transportation Safety Data with

从自然语言接入交通安全部署，到多模态地球观测基础模型跃升——地理大模型正迈向可验证、可执行、可治理的新阶段。

Transportation safety analysis requires integrating crash records, roadway attributes, and geospatial data through GIS-based workflows, but access remains uneven across agencies and community stakeholders. Technical prerequisites create a gap between analytical tools central to safety planning and the practitioners able to use them. Local agencies, school committees, and residents may have safety concerns but limited capacity to retrieve, filter, map, and analyze relevant data. Generative AI offers a way to narrow this divide, but its public-sector use raises questions about reliability, reproducibility, and governance. This paper presents a schema-grounded natural language interface for transportation safety analysis, using a large language model (LLM) to interpret user intent while preserving deterministic, reviewable execution against an authoritative database. User queries are translated into structured semantic frames, validated by a rule-based layer, compiled into a typed directed acyclic graph of spatial operations, and executed against a PostGIS database. This bounded design separates language interpretation from deterministic execution, keeping results reproducible and schema-grounded while removing access barriers. The framework is evaluated using a statewide Massachusetts transportation safety database integrating crash records, roadway attributes, and geospatial layers including schools, bus stops, crosswalks, and municipal boundaries. All queries executed successfully; the validation layer corrects errors in 29% of evaluation queries, reflecting the gap between flexible natural language and strict schema-grounded requirements. The results suggest that combining natural language accessibility with deterministic execution is a practical direction for broadening access to transportation safety data, with implications for trustworthy AI in public-sector planning.

编者按：本期头版聚焦地理智能体（Geo-Agent）演进的核心张力：一边是自然语言界面降低专业数据使用门槛的务实突破，一边是多模态基础模型对空间本质表征能力的持续深化。

TREND OVERVIEW

趋势综述：地理智能体时代：空间理解的范式迁移与可信落地。

近期研究聚焦于将多源异构地理数据（如HSI、多光谱LiDAR、OSM图结构）融入GeoLLM/GeoFM训练与评估框架，并通过构建专用基准（如GS-QA、CrossViewBench、OSM+）支撑空间推理与智能体行为建模。

近期研究聚焦于突破模态孤立建模局限，转向场景中心（scene-centered）或几何接地（geometric grounding）的统一表征与联合生成；方法重心从两两模态翻译转向多模态基础模型预训练、跨模态流匹配与令牌级几何自适应。

近期研究聚焦于将轨迹数据作为多模态协同决策与动态资源配置的底层支撑，问题重心从静态建模转向时空耦合的实时响应与跨域（空-地-网）协同优化。

DIRECTION PULSE

1 地理大模型与地理智能体

近期研究聚焦于将多源异构地理数据（如HSI、多光谱LiDAR、OSM图结构）融入GeoLLM/GeoFM训练与评估框架，并通过构建专用基准（如GS-QA、CrossViewBench、OSM+）支撑空间推理与智能体行为建模。

2 多源多模态地理数据

近期研究聚焦于突破模态孤立建模局限，转向场景中心（scene-centered）或几何接地（geometric grounding）的统一表征与联合生成；方法重心从两两模态翻译转向多模态基础模型预训练、跨模态流匹配与令牌级几何自适应。

3 轨迹数据与城市交通研究

近期研究聚焦于将轨迹数据作为多模态协同决策与动态资源配置的底层支撑，问题重心从静态建模转向时空耦合的实时响应与跨域（空-地-网）协同优化。

4 复杂网络、韧性城市与地理模拟

近 30 天该方向累计出现 0 条相关内容，重点集中在 复杂网络、韧性城市与地理模拟。

5 城市感知、街景感知与空间优化

近 30 天该方向累计出现 0 条相关内容，重点集中在 城市感知、街景感知与空间优化。

HIGHLIGHTS

- 自然语言接口首次实现交通安全部署级确定性执行，语言理解与PostGIS空间运算严格解耦。
- 首个融合星载高光谱与多源遥感的地球观测基础模型SpectralEarth-FM发布，支持异构谱维联合预训练。
- ArchSIBench首次系统定义建筑空间智能五大维度，揭示VLMs在布局理解与功能配置上的显著认知缺口。
- GeoX开创无标注自博弈框架，以可执行程序为载体驱动VLM自主习得地理空间逻辑。

地理大模型与地理智能体

UrbanComp Lab 学习资料库（https://research.urbancomp.dev/）

近期研究聚焦于将多源异构地理数据（如HSI、多光谱LiDAR、OSM图结构）融入GeoLLM/GeoFM训练与评估框架，并通过构建专用基准（如GS-QA、CrossViewBench、OSM+）支撑空间推理与智能体行为建模。

近30天 17 | 近7天 15 | 来源 1 | 论文 20

趋势信号

- hyperspectral imagery（HSI）正被系统性纳入地球观测基础模型的多模态预训练，以弥补其在现有EO-FM中的表征缺失
- 面向地理空间问答（geospatial QA）和跨视角空间推理的专用基准数据集（GS-QA、CrossViewBench）被密集提出，强调多源信息融合与空间谓词多样性
- 十亿级结构化地理图数据集OSM+发布，旨在支撑城市级地理智能体的可扩展性验证与真实拓扑建模
- 检索增强型时空建模（如Bridge）开始用于解决地理智能体在冷启动区域的动态决策问题，体现对'记忆-推理-行动'闭环的初步探索

核心观点

- 地理大模型的核心挑战不在单纯扩大参数规模，而在于构建与地理本体（如空间关系、尺度、投影、拓扑）对齐的表示与推理机制
- 多模态融合必须超越图像-文本对齐，需显式建模地理传感器特性（如HSI光谱响应、LiDAR几何-光谱耦合、OSM语义-拓扑耦合）
- 地理智能体的有效性高度依赖于高质量、结构化、任务对齐的基准数据集，而非通用LLM微调范式
- 空间智能（Spatial Intelligence）被明确定义为跨视角、跨模态、跨尺度的一致性推理能力，且需通过指令微调与显式对齐机制（如CrossViewer）实现

RESEARCH IDEA

GS-QA基准中多源推理问题在OSM+十亿级图上失效

GS-QA中依赖OSM与Wikipedia双源协同的空间谓词问答，在OSM+十亿级道路图结构上因缺乏Wikipedia语义锚点与图结构-文本对齐接口而无法执行多跳空间推理

为什么现在值得做：城市计算与应急响应系统亟需在超大规模路网中执行带语义约束的空间查询（例如'距最近三甲医院500米内且周边有无障碍公交站的社区'），而当前无公开数据集支持该类问题在十亿级图上的端到端评估。

关键难点

- OSM+未发布节点级Wikipedia实体ID映射表，需重建跨源对齐索引
- GS-QA中Wikipedia文本引用未标准化，存在重定向、消歧与版本漂移问题
- 十亿级图上执行多跳空间谓词推理需图数据库原生支持方向性边过滤与坐标系一致的几何计算，当前Neo4j/PostGIS插件均不满足

建议切入

- 基于OSM+节点经纬度与Wikipedia地理坐标模板（如{{Coord}}）进行粗粒度空间匹配，生成初始实体链接候选集
- 利用CrossViewBench中跨视角实体对齐模块微调一个轻量级实体链接模型，以解决Wikipedia页面消歧
- 将GS-QA问题模板编译为Cypher+ST_Geometry混合查询语言，在OSM+导出的PostGIS实例中实现空间谓词与文本谓词的联合执行

REPRESENTATIVE ITEMS

ARXIV

SpectralEarth-FM

Earth observation (EO) foundation models (FMs) are increasingly trained on multisensor data, spanning multispectral imagery (MSI), synthetic aperture radar (SAR), and derived geospatial layers, but hyperspectral imagery (HSI) remains underrepresented. Conversely, existing hyperspectral FMs are trained on HSI alone, leaving joint pretraining and fusion of HSI with co-located EO sensors unexplored. We introduce SpectralEarth-FM, a hierarchical transformer for multisensor EO input with heterogeneous spectral dimensionality. The architecture combines spectral tokenization for hyperspectral inputs, sensor-specific encoders, a cross-sensor fusion module, and a shared hierarchical encoder, enabling joint processing of HSI and lower-channel observations. To pretrain SpectralEarth-FM, we curate SpectralEarth-MM, a dataset that co-locates HSI from three spaceborne sensors (EnMAP, EMIT, DESIS) with Sentinel-2, Landsat-8/9 optical imagery, Landsat land surface temperature (LST), and Sentinel-1 SAR, over common geographic footprints. It comprises approximately 2M globally distributed locations, 25M georeferenced patches, and over 40TB of data. Pretraining uses a Joint-Embedding Predictive Architecture (JEPA)-style objective that matches representations between global views and single-sensor local views from the same location. We evaluate SpectralEarth-FM on hyperspectral downstream tasks and standard EO benchmarks following the PANGAEA protocol, achieving state-of-the-art results across both evaluation settings.。

ARXIV

GS-QA：一个面向地理空间问答的基准数据集

大型语言模型（LLMs）的近期进展显著提升了问答（QA）任务的性能。为应对QA系统评估的挑战，学界已提出若干标准化基准。本工作聚焦于地理空间问答（geospatial QA）问题，其中存在大量以空间数据库或其他形式组织的地理空间数据。

ARXIV

基于多光谱LiDAR与深度学习的三维土地利用/土地覆盖（LULC）分

土地利用/土地覆盖（LULC）分类对国家三维测绘、地理空间分析及可持续规划至关重要。多光谱（MS）LiDAR可同步获取空间-光谱信息，而深度学习（DL）支持三维点云语义分割；然而，其应用受限于缺乏公开可用的城市与郊区多光谱LiDAR数据集，且现有数据集尚未与国家测绘与地籍机构（NMCAs）的分类体系对齐。本研究通过提出L1与L2两级NMCA对齐的LULC分类体系，并构建一个新型基准多光谱LiDAR数据集，填补上述空白。

ARXIV

OSM+：面向城市级实验的十亿级 OpenStreetMap 数据集

道路网络数据蕴含丰富的城市信息，但处理全球范围的OpenStreetMap（OSM）数据计算开销巨大，且生成的图结构往往难以统一，不利于下游任务的基准评测。现有图学习基准未能体现真实世界道路网络的十亿级规模及其独特拓扑特性，导致模型可扩展性研究不足。

近期研究聚焦于突破模态孤立建模局限，转向场景中心（scene-centered）或几何接地（geometric grounding）的统一表征与联合生成；方法重心从两两模态翻译转向多模态基础模型预训练、跨模态流匹配与令牌级几何自适应。

近30天 20 | 近7天 17 | 来源 4 | 论文 24

趋势信号

- MetaEarth-MM提出scene-centered joint modeling以替代孤立pairwise translation
- SpectralEarth-FM强调将hyperspectral imagery（HSI）纳入多模态EO foundation models预训练，弥补其当前缺失
- GeoWeaver引入token-adaptive geometric evidence allocation，将几何建模前置于推理阶段
- FlowGS结合流匹配（flow matching）与2D高斯溅射（Gaussian Splatting）实现连续尺度遥感图像生成，体现生成范式向高效、可微、几何感知演进

核心观点

- 多模态地理数据融合不能仅依赖语义对齐，必须显式建模物理几何结构与空间约束
- hyperspectral imagery（HSI）是当前多模态地球观测预训练中显著缺失的关键模态
- ‘GeoMultimodal’正从数据拼接走向联合表征学习，核心挑战在于跨传感器、跨分辨率、跨物理量纲的对齐与先验建模
- 生成式方法（如扩散、流匹配）在遥感多模态任务中正被重新定位为表征学习与重建的统一接口，而非单纯图像合成工具

RESEARCH IDEA

HSI-MSI-SAR联合预训练在城市场景中因光谱-几何解耦失效

SpectralEarth-FM提出的HSI-MSI-SAR联合预训练范式在城市场景中无法保持光谱判别性与空间结构一致性，因其将HSI嵌入与MSI/SAR特征对齐置于同一语义层级，未建模城市地物的局部光谱混叠与三维遮挡导致的几何畸变耦合效应

为什么现在值得做：城市更新与低碳治理亟需融合亚米级SAR建筑形变、HSI材料成分与MSI时序变化的诊断能力；FlowGS等连续尺度重建方法已提供跨分辨率对齐基础，使HSI超分后与SAR/MSI的空间配准具备操作可行性。

关键难点

- 需构建城市典型地物（如玻璃幕墙、光伏板、植被屋顶）的HSI-MSI-SAR三元组配准数据集，现有公开数据集无同步采集保障
- 光谱混叠（如阴影区HSI信噪比骤降）与几何畸变（如SAR侧视导致的建筑物位移）的联合退化建模缺乏可微分物理先验
- SpectralEarth-FM冻结的HSI编码器无法反向传播几何梯度，需重设计可导的光谱-几何双流投影头

建议切入

- 基于旧金山UAM走廊空域建模中使用的细粒度地面出行数据，反演典型城区三维点云与材质分布，驱动HSI-SAR合成数据生成
- 复用GeoWeaver的令牌自适应几何证据分配机制，将其扩展至HSI频段维度，为每个波段通道动态检索对应几何抽象（如法向量、遮挡掩码
- 在SpectralEarth-FM主干中插入轻量级光谱-几何解耦模块：前馈路径保留光谱判别性，残差路径注入几何约束梯度

REPRESENTATIVE ITEMS

ARXIV

MetaEarth-MM

Multi-modal remote sensing images are vital for Earth observation, yet complete paired observations are often scarce in practice. Existing generative methods commonly address this problem through isolated pairwise modality translation, but their versatility and scalability remain limited as the number of modalities and generation tasks increases. Here, we develop a generative foundation model MetaEarth-MM for multi-modal remote sensing imagery, enabling paired joint generation and any-to-any translation across five modalities within a unified model. Recognizing the intrinsic scene consistency underlying multi-modal observations, we introduce a scene-centered joint modeling paradigm in MetaEarth-MM. Unlike previous methods that rely on direct appearance-level cross-modal mapping, our model organizes the generation around the underlying scene content. Specifically, MetaEarth-MM adopts a decoupled architecture that first infers a latent scene representation from available observations, and then generates target modalities conditioned on this intermediate state. To support training, we further construct EarthMM, a large-scale dataset comprising 2.8 million multi-resolution global images with 2.2 million aligned pairs. Extensive experiments demonstrate that MetaEarth-MM not only exhibits strong generative capability and robust generalization across diverse generation tasks, but also supports downstream tasks at both data and representation levels, highlighting its potential as a general foundation model for cross-modal Earth observation. The code and dataset will be available at https://github.com/YZPioneer/MetaEarth-MM.。

ISPRS JOURNAL OF PHOTOGRAMMETRY AND REMOTE SENSING

基于视觉模型引导与门控注意力的鲁棒无检测器多模态图像匹配

出版日期：2026年8月；来源：《ISPRS摄影测量与遥感杂志》（ISPRS Journal of Photogrammetry and Remote Sensing），第238卷；作者：唐腾飞、韩志强、彭涛、陈金浩、叶远鑫。

ARXIV

SpectralEarth-FM

Earth observation (EO) foundation models (FMs) are increasingly trained on multisensor data, spanning multispectral imagery (MSI), synthetic aperture radar (SAR), and derived geospatial layers, but hyperspectral imagery (HSI) remains underrepresented. Conversely, existing hyperspectral FMs are trained on HSI alone, leaving joint pretraining and fusion of HSI with co-located EO sensors unexplored. We introduce SpectralEarth-FM, a hierarchical transformer for multisensor EO input with heterogeneous spectral dimensionality. The architecture combines spectral tokenization for hyperspectral inputs, sensor-specific encoders, a cross-sensor fusion module, and a shared hierarchical encoder, enabling joint processing of HSI and lower-channel observations. To pretrain SpectralEarth-FM, we curate SpectralEarth-MM, a dataset that co-locates HSI from three spaceborne sensors (EnMAP, EMIT, DESIS) with Sentinel-2, Landsat-8/9 optical imagery, Landsat land surface temperature (LST), and Sentinel-1 SAR, over common geographic footprints. It comprises approximately 2M globally distributed locations, 25M georeferenced patches, and over 40TB of data. Pretraining uses a Joint-Embedding Predictive Architecture (JEPA)-style objective that matches representations between global views and single-sensor local views from the same location. We evaluate SpectralEarth-FM on hyperspectral downstream tasks and standard EO benchmarks following the PANGAEA protocol, achieving state-of-the-art results across both evaluation settings.。

ARXIV

基于流的高斯溅射方法用于连续尺度遥感图像超分辨率

高分辨率遥感图像（RSI）对地球观测应用至关重要，但其获取常受限于传感器能力与成本。近年来，生成式超分辨率（SR）方法——尤其是扩散模型——取得了显著进展；然而，这类方法通常需耗时的迭代推理（40 – 1000步），且在连续尺度SR场景中灵活性有限。为解决上述问题，我们提出FlowGS，一种面向任意尺度RSI超分辨率的生成式重建框架。

轨迹数据与城市交通研究

UrbanComp Lab 学习资料库（https://research.urbancomp.dev/）

近期研究聚焦于将轨迹数据作为多模态协同决策与动态资源配置的底层支撑，问题重心从静态建模转向时空耦合的实时响应与跨域（空-地-网）协同优化。

近30天 16 | 近7天 13 | 来源 4 | 论文 18

趋势信号

- UAM走廊中引入动态单向车道分配，依据时变空域需求对车道进行激活、停用或方向反转
- 自动驾驶VLA模型开始摒弃长链自然语言推理接口，转向可自动导出的单步元动作（meta-actions）作为轨迹条件化决策接口
- ISAC系统中利用HDBSCAN聚类热点引导无人机轨迹规划，并结合Soft Actor-Critic联合优化轨迹、天线位置与波束赋形
- 3D Gaussian Splatting训练被重构为轨迹依赖型核外优化问题，仅缓存当前相机批次可见的高斯基元工作集

核心观点

- 轨迹不仅是运动记录，更是连接物理移动、语义意图与系统控制的结构化接口
- 多式联运效率提升高度依赖首程-中程-末程的全链路轨迹分解与协同建模
- 动态性（如空域车道方向切换、无人机轨迹重规划、元动作序列生成）已成为提升系统资源利用率与服务鲁棒性的关键设计原则
- 轨迹条件化建模正从‘后验分析’转向‘前摄干预’：即以轨迹为输入驱动调度、通信、感知等子系统的联合优化

RESEARCH IDEA

轨迹方法跨城市迁移的首要失稳环节

轨迹方法迁移到另一座城市或极端天气场景后，最先失稳的通常不是模型结构，而是采样方式、路网约束和行为机制的变化。

为什么现在值得做： 城市空中交通（UAM）走廊中的动态车道分配以提升多式联运门到门出行效率 与 FUSE：面向车载与机器人SLAM系统的统一状态估计框架 已经提供了可复用的变量、数据或模型入口，这使得问题不再停留在概念层面，可以直接构造成小规模验证。

关键难点

- 需构建中小城市典型路网拓扑约束下的轨迹特征敏感性分析框架
- 原始论文未公开聚类特征权重配置，无法直接复现低密度条件下的特征缩放策略
- 缺乏公开的中小城市带语义标签的异构轨迹数据集用于验证场景类别一致性

建议切入

- 基于OSM提取全国100个县级市路网参数（车道数、交叉口密度、平均路段长度），构建低密度路网基准测试集
- 在SceneSelect原始特征空间中引入拓扑归一化项：将空间邻近度按路段平均长度重标度，交互频次按车辆密度加权
- 采用FUSE框架中提出的退化感知校正模块，对低信噪比聚类中心进行鲁棒性增强

REPRESENTATIVE ITEMS

ARXIV

城市空中交通（UAM）走廊中的动态车道分配以提升多式联运门到门出行效

本文将城市空中交通（UAM）走廊中的动态单向车道分配建模为一个离散时间混合整数线性规划（MILP）问题，该模型可根据双向空域需求的时变特性，对车道进行激活、停用或方向反转。我们基于细粒度地面出行数据建模需求，将每趟出行分解为包含首程、中程与末程的多式联运序列，并通过垂直起降机场侧调度模型对由UAM承运的中程段进行路径规划。以旧金山湾区为案例，在康特拉科斯塔县与硅谷之间部署一条跨多区域的UAM走廊。

ARXIV

DiagEval

评估大语言模型（LLM）生成的交互式软件不仅需静态分析，还需实际执行。核心难点在于：正确性是潜在用户界面（UI）状态转移图上的图级可达性质，而 GUI 评估器仅能观测单次执行轨迹。因此，一次失败的 rollout 仅排除一条已实现路径，导致失败归因在评估器侧执行错误与真实软件缺陷之间存在歧义。

ARXIV

LVDrive：基于潜在视觉表征增强的视觉-语言-动作自动驾驶模型

视觉-语言-动作（VLA）模型已成为端到端自动驾驶的一种有前景的框架。然而，现有VLA模型通常依赖稀疏的动作监督，未能充分利用其强大的场景理解与推理能力。近期通过世界模型引入密集视觉监督的尝试，往往过度强调像素级图像重建，而忽视了语义上有意义的场景表征学习。

ARXIV

DriveMA：以单步元动作重构驾驶视觉-语言-动作模型（Drivi

驾驶视觉-语言-动作模型（Driving VLAs）通常引入自然语言推理作为端到端规划的中间接口，但以推理为中心的接口面临三个实际瓶颈：高质量推理标注难以获取；紧凑型模型难以生成与理解长推理链；推理延迟显著增加。本文重新审视Driving VLAs中语言接口的设计，表明简洁的单步元动作（meta-actions）是一种简单而有效的替代方案，可取代冗长的自然语言推理。

近 30 天该方向累计出现 0 条相关内容，重点集中在 复杂网络、韧性城市与地理模拟。

近30天 0 | 近7天 0 | 来源 0 | 论文 0

趋势信号

- 近 7 天新增 0 条，近 90 天累计 0 条。
- 内容结构以论文 0 条、资讯 0 条为主。
- 来源覆盖 0 个渠道，显示该方向具备持续输入。

核心观点

- Broadening Access to Transportation Safety Data with Generative AI: A Schema-Grounded Framework for Spatial Natural Language Queries：Transportation safety analysis requires integrating crash records, roadw...
- SpectralEarth-FM: Bringing Hyperspectral Imagery into Multimodal Earth Observation Pretraining：Earth observation (EO) foundation models (FMs) are increasingly trained...
- ArchSIBench: Benchmarking the Architectural Spatial Intelligence of Vision-Language Models：Architectural spatial intelligence, the ability to recognize and infer a...

REPRESENTATIVE ITEMS

ARXIV

Broadening Access to

Transportation safety analysis requires integrating crash records, roadway attributes, and geospatial data through GIS-based workflows, but access remains uneven across agencies and community stakeholders. Technical prerequisites create a gap between analytical tools central to safety planning and the practitioners able to use them. Local agencies, school committees, and residents may have safety concerns but limited capacity to retrieve, filter, map, and analyze relevant data. Generative AI offers a way to narrow this divide, but its public-sector use raises questions about reliability, reproducibility, and governance. This paper presents a schema-grounded natural language interface for transportation safety analysis, using a large language model (LLM) to interpret user intent while preserving deterministic, reviewable execution against an authoritative database. User queries are translated into structured semantic frames, validated by a rule-based layer, compiled into a typed directed acyclic graph of spatial operations, and executed against a PostGIS database. This bounded design separates language interpretation from deterministic execution, keeping results reproducible and schema-grounded while removing access barriers. The framework is evaluated using a statewide Massachusetts transportation safety database integrating crash records, roadway attributes, and geospatial layers including schools, bus stops, crosswalks, and municipal boundaries. All queries executed successfully; the validation layer corrects errors in 29% of evaluation queries, reflecting the gap between flexible natural language and strict schema-grounded requirements. The results suggest that combining natural language accessibility with deterministic execution is a practical direction for broadening access to transportation safety data, with implications for trustworthy AI in public-sector planning.。

ARXIV

SpectralEarth-FM

Earth observation (EO) foundation models (FMs) are increasingly trained on multisensor data, spanning multispectral imagery (MSI), synthetic aperture radar (SAR), and derived geospatial layers, but hyperspectral imagery (HSI) remains underrepresented. Conversely, existing hyperspectral FMs are trained on HSI alone, leaving joint pretraining and fusion of HSI with co-located EO sensors unexplored. We introduce SpectralEarth-FM, a hierarchical transformer for multisensor EO input with heterogeneous spectral dimensionality. The architecture combines spectral tokenization for hyperspectral inputs, sensor-specific encoders, a cross-sensor fusion module, and a shared hierarchical encoder, enabling joint processing of HSI and lower-channel observations. To pretrain SpectralEarth-FM, we curate SpectralEarth-MM, a dataset that co-locates HSI from three spaceborne sensors (EnMAP, EMIT, DESIS) with Sentinel-2, Landsat-8/9 optical imagery, Landsat land surface temperature (LST), and Sentinel-1 SAR, over common geographic footprints. It comprises approximately 2M globally distributed locations, 25M georeferenced patches, and over 40TB of data. Pretraining uses a Joint-Embedding Predictive Architecture (JEPA)-style objective that matches representations between global views and single-sensor local views from the same location. We evaluate SpectralEarth-FM on hyperspectral downstream tasks and standard EO benchmarks following the PANGAEA protocol, achieving state-of-the-art results across both evaluation settings.。

ARXIV

ArchSIBench

Architectural spatial intelligence, the ability to recognize and infer architectural space, is fundamental to tasks such as robot navigation, embodied interaction, and 3D scene understanding and generation. Although extensive research has evaluated the basic spatial skills of Vision-Language Models (VLMs) such as relative orientation, distance comparison, and object counting, these tasks cover only the most elementary levels of spatial cognition and largely overlook higher-level cognition of architectural space, including layout understanding, circulation patterns, and functional zoning. In this work, we present ArchSIBench, a Benchmark for Architectural Spatial Intelligence based on the perspectives from architecture, cognitive science, and psychology. ArchSIBench covers five core dimensions: perception, reasoning, navigation, transformation, and configuration, comprising 17 fine-grained subtasks. Through careful manual annotation by experts with architectural backgrounds, we construct 3,000 question-answer pairs to enable comprehensive evaluation of architectural spatial intelligence. Based on ArchSIBench, we evaluate various VLMs and find that the architectural spatial intelligence of most models shows significant differences from human baselines; additionally, models exhibit substantial variability across capability dimensions. Some state-of-the-art models can approach the level of human evaluators without architectural training. However, a clear gap remains compared to human evaluators with architectural training, particularly in spatial transformation and configuration reasoning. We believe that ArchSIBench will provide important insights and systematic resources for measuring and advancing the architectural spatial intelligence of VLMs. The dataset and code are available at https://huggingface.co/datasets/ArchSIBench/ArchSIBench.。

近 30 天该方向累计出现 0 条相关内容，重点集中在 城市感知、街景感知与空间优化。

近30天 0 | 近7天 0 | 来源 0 | 论文 0

趋势信号

- 近 7 天新增 0 条，近 90 天累计 0 条。
- 内容结构以论文 0 条、资讯 0 条为主。
- 来源覆盖 0 个渠道，显示该方向具备持续输入。

核心观点

- Broadening Access to Transportation Safety Data with Generative AI: A Schema-Grounded Framework for Spatial Natural Language Queries：Transportation safety analysis requires integrating crash records, roadw...
- SpectralEarth-FM: Bringing Hyperspectral Imagery into Multimodal Earth Observation Pretraining：Earth observation (EO) foundation models (FMs) are increasingly trained...
- ArchSIBench: Benchmarking the Architectural Spatial Intelligence of Vision-Language Models：Architectural spatial intelligence, the ability to recognize and infer a...

REPRESENTATIVE ITEMS

ARXIV

Broadening Access to

Transportation safety analysis requires integrating crash records, roadway attributes, and geospatial data through GIS-based workflows, but access remains uneven across agencies and community stakeholders. Technical prerequisites create a gap between analytical tools central to safety planning and the practitioners able to use them. Local agencies, school committees, and residents may have safety concerns but limited capacity to retrieve, filter, map, and analyze relevant data. Generative AI offers a way to narrow this divide, but its public-sector use raises questions about reliability, reproducibility, and governance. This paper presents a schema-grounded natural language interface for transportation safety analysis, using a large language model (LLM) to interpret user intent while preserving deterministic, reviewable execution against an authoritative database. User queries are translated into structured semantic frames, validated by a rule-based layer, compiled into a typed directed acyclic graph of spatial operations, and executed against a PostGIS database. This bounded design separates language interpretation from deterministic execution, keeping results reproducible and schema-grounded while removing access barriers. The framework is evaluated using a statewide Massachusetts transportation safety database integrating crash records, roadway attributes, and geospatial layers including schools, bus stops, crosswalks, and municipal boundaries. All queries executed successfully; the validation layer corrects errors in 29% of evaluation queries, reflecting the gap between flexible natural language and strict schema-grounded requirements. The results suggest that combining natural language accessibility with deterministic execution is a practical direction for broadening access to transportation safety data, with implications for trustworthy AI in public-sector planning.。

ARXIV

SpectralEarth-FM

Earth observation (EO) foundation models (FMs) are increasingly trained on multisensor data, spanning multispectral imagery (MSI), synthetic aperture radar (SAR), and derived geospatial layers, but hyperspectral imagery (HSI) remains underrepresented. Conversely, existing hyperspectral FMs are trained on HSI alone, leaving joint pretraining and fusion of HSI with co-located EO sensors unexplored. We introduce SpectralEarth-FM, a hierarchical transformer for multisensor EO input with heterogeneous spectral dimensionality. The architecture combines spectral tokenization for hyperspectral inputs, sensor-specific encoders, a cross-sensor fusion module, and a shared hierarchical encoder, enabling joint processing of HSI and lower-channel observations. To pretrain SpectralEarth-FM, we curate SpectralEarth-MM, a dataset that co-locates HSI from three spaceborne sensors (EnMAP, EMIT, DESIS) with Sentinel-2, Landsat-8/9 optical imagery, Landsat land surface temperature (LST), and Sentinel-1 SAR, over common geographic footprints. It comprises approximately 2M globally distributed locations, 25M georeferenced patches, and over 40TB of data. Pretraining uses a Joint-Embedding Predictive Architecture (JEPA)-style objective that matches representations between global views and single-sensor local views from the same location. We evaluate SpectralEarth-FM on hyperspectral downstream tasks and standard EO benchmarks following the PANGAEA protocol, achieving state-of-the-art results across both evaluation settings.。

ARXIV

ArchSIBench

Architectural spatial intelligence, the ability to recognize and infer architectural space, is fundamental to tasks such as robot navigation, embodied interaction, and 3D scene understanding and generation. Although extensive research has evaluated the basic spatial skills of Vision-Language Models (VLMs) such as relative orientation, distance comparison, and object counting, these tasks cover only the most elementary levels of spatial cognition and largely overlook higher-level cognition of architectural space, including layout understanding, circulation patterns, and functional zoning. In this work, we present ArchSIBench, a Benchmark for Architectural Spatial Intelligence based on the perspectives from architecture, cognitive science, and psychology. ArchSIBench covers five core dimensions: perception, reasoning, navigation, transformation, and configuration, comprising 17 fine-grained subtasks. Through careful manual annotation by experts with architectural backgrounds, we construct 3,000 question-answer pairs to enable comprehensive evaluation of architectural spatial intelligence. Based on ArchSIBench, we evaluate various VLMs and find that the architectural spatial intelligence of most models shows significant differences from human baselines; additionally, models exhibit substantial variability across capability dimensions. Some state-of-the-art models can approach the level of human evaluators without architectural training. However, a clear gap remains compared to human evaluators with architectural training, particularly in spatial transformation and configuration reasoning. We believe that ArchSIBench will provide important insights and systematic resources for measuring and advancing the architectural spatial intelligence of VLMs. The dataset and code are available at https://huggingface.co/datasets/ArchSIBench/ArchSIBench.。

研究日报

2026年5月24日 星期日

武汉光谷 · 阴 · 26°C · 24—28°C

路漫漫其修远兮，吾将上下而求索。

第 1 版 / 7

UrbanComp Lab 学习资料库（https://research.urbancomp.dev/）

THIS EDITION

五个方向的当日进展

Broadening Access to Transportation Safety Data with

从自然语言接入交通安全部署，到多模态地球观测基础模型跃升——地理大模型正迈向可验证、可执行、可治理的新阶段。

Transportation safety analysis requires integrating crash records, roadway attributes, and geospatial data through GIS-based workflows, but access remains uneven across agencies and community stakeholders. Technical prerequisites create a gap between analytical tools central to safety planning and the practitioners able to use them. Local agencies, school committees, and residents may have safety concerns but limited capacity to retrieve, filter, map, and analyze relevant data. Generative AI offers a way to narrow this divide, but its public-sector use raises questions about reliability, reproducibility, and governance. This paper presents a schema-grounded natural language interface for transportation safety analysis, using a large language model (LLM) to interpret user intent while preserving deterministic, reviewable execution against an authoritative database. User queries are translated into structured semantic frames, validated by a rule-based layer, compiled into a typed directed acyclic graph of spatial operations, and executed against a PostGIS database. This bounded design separates language interpretation from deterministic execution, keeping results reproducible and schema-grounded while removing access barriers. The framework is evaluated using a statewide Massachusetts transportation safety database integrating crash records, roadway attributes, and geospatial layers including schools, bus stops, crosswalks, and municipal boundaries. All queries executed successfully; the validation layer corrects errors in 29% of evaluation queries, reflecting the gap between flexible natural language and strict schema-grounded requirements. The results suggest that combining natural language accessibility with deterministic execution is a practical direction for broadening access to transportation safety data, with implications for trustworthy AI in public-sector planning.

编者按：本期头版聚焦地理智能体（Geo-Agent）演进的核心张力：一边是自然语言界面降低专业数据使用门槛的务实突破，一边是多模态基础模型对空间本质表征能力的持续深化。

TREND OVERVIEW

趋势综述：地理智能体时代：空间理解的范式迁移与可信落地。

近期研究聚焦于将多源异构地理数据（如HSI、多光谱LiDAR、OSM图结构）融入GeoLLM/GeoFM训练与评估框架，并通过构建专用基准（如GS-QA、CrossViewBench、OSM+）支撑空间推理与智能体行为建模。

近期研究聚焦于突破模态孤立建模局限，转向场景中心（scene-centered）或几何接地（geometric grounding）的统一表征与联合生成；方法重心从两两模态翻译转向多模态基础模型预训练、跨模态流匹配与令牌级几何自适应。

近期研究聚焦于将轨迹数据作为多模态协同决策与动态资源配置的底层支撑，问题重心从静态建模转向时空耦合的实时响应与跨域（空-地-网）协同优化。

DIRECTION PULSE

1 地理大模型与地理智能体

近期研究聚焦于将多源异构地理数据（如HSI、多光谱LiDAR、OSM图结构）融入GeoLLM/GeoFM训练与评估框架，并通过构建专用基准（如GS-QA、CrossViewBench、OSM+）支撑空间推理与智能体行为建模。

2 多源多模态地理数据

近期研究聚焦于突破模态孤立建模局限，转向场景中心（scene-centered）或几何接地（geometric grounding）的统一表征与联合生成；方法重心从两两模态翻译转向多模态基础模型预训练、跨模态流匹配与令牌级几何自适应。

3 轨迹数据与城市交通研究

近期研究聚焦于将轨迹数据作为多模态协同决策与动态资源配置的底层支撑，问题重心从静态建模转向时空耦合的实时响应与跨域（空-地-网）协同优化。

4 复杂网络、韧性城市与地理模拟

近 30 天该方向累计出现 0 条相关内容，重点集中在 复杂网络、韧性城市与地理模拟。

5 城市感知、街景感知与空间优化

近 30 天该方向累计出现 0 条相关内容，重点集中在 城市感知、街景感知与空间优化。

HIGHLIGHTS

- 自然语言接口首次实现交通安全部署级确定性执行，语言理解与PostGIS空间运算严格解耦。
- 首个融合星载高光谱与多源遥感的地球观测基础模型SpectralEarth-FM发布，支持异构谱维联合预训练。
- ArchSIBench首次系统定义建筑空间智能五大维度，揭示VLMs在布局理解与功能配置上的显著认知缺口。
- GeoX开创无标注自博弈框架，以可执行程序为载体驱动VLM自主习得地理空间逻辑。

UrbanComp Lab 学习资料库（https://research.urbancomp.dev/）

近期研究聚焦于将多源异构地理数据（如HSI、多光谱LiDAR、OSM图结构）融入GeoLLM/GeoFM训练与评估框架，并通过构建专用基准（如GS-QA、CrossViewBench、OSM+）支撑空间推理与智能体行为建模。

近30天 17 | 近7天 15 | 来源 1 | 论文 20

趋势信号

- hyperspectral imagery（HSI）正被系统性纳入地球观测基础模型的多模态预训练，以弥补其在现有EO-FM中的表征缺失
- 面向地理空间问答（geospatial QA）和跨视角空间推理的专用基准数据集（GS-QA、CrossViewBench）被密集提出，强调多源信息融合与空间谓词多样性
- 十亿级结构化地理图数据集OSM+发布，旨在支撑城市级地理智能体的可扩展性验证与真实拓扑建模
- 检索增强型时空建模（如Bridge）开始用于解决地理智能体在冷启动区域的动态决策问题，体现对'记忆-推理-行动'闭环的初步探索

核心观点

- 地理大模型的核心挑战不在单纯扩大参数规模，而在于构建与地理本体（如空间关系、尺度、投影、拓扑）对齐的表示与推理机制
- 多模态融合必须超越图像-文本对齐，需显式建模地理传感器特性（如HSI光谱响应、LiDAR几何-光谱耦合、OSM语义-拓扑耦合）
- 地理智能体的有效性高度依赖于高质量、结构化、任务对齐的基准数据集，而非通用LLM微调范式
- 空间智能（Spatial Intelligence）被明确定义为跨视角、跨模态、跨尺度的一致性推理能力，且需通过指令微调与显式对齐机制（如CrossViewer）实现

RESEARCH IDEA

GS-QA基准中多源推理问题在OSM+十亿级图上失效

GS-QA中依赖OSM与Wikipedia双源协同的空间谓词问答，在OSM+十亿级道路图结构上因缺乏Wikipedia语义锚点与图结构-文本对齐接口而无法执行多跳空间推理

为什么现在值得做：城市计算与应急响应系统亟需在超大规模路网中执行带语义约束的空间查询（例如'距最近三甲医院500米内且周边有无障碍公交站的社区'），而当前无公开数据集支持该类问题在十亿级图上的端到端评估。

关键难点

- OSM+未发布节点级Wikipedia实体ID映射表，需重建跨源对齐索引

建议切入

- 基于OSM+节点经纬度与Wikipedia地理坐标模板（如{{Coord}}）进行粗粒度空间匹配，生成初始实体链接候选集

REPRESENTATIVE ITEMS

ARXIV

SpectralEarth-FM

Earth observation (EO) foundation models (FMs) are increasingly trained on multisensor data, spanning multispectral imagery (MSI), synthetic aperture radar (SAR), and derived geospatial layers, but hyperspectral imagery (HSI) remains underrepresented. Conversely, existing hyperspectral FMs are trained on HSI alone, leaving joint pretraining and fusion of HSI with co-located EO sensors unexplored. We introduce SpectralEarth-FM, a hierarchical transformer for multisensor EO input with heterogeneous spectral dimensionality. The architecture combines spectral tokenization for hyperspectral inputs, sensor-specific encoders, a cross-sensor fusion module, and a shared hierarchical encoder, enabling joint processing of HSI and lower-channel observations. To pretrain SpectralEarth-FM, we curate SpectralEarth-MM, a dataset that co-locates HSI from three spaceborne sensors (EnMAP, EMIT, DESIS) with Sentinel-2, Landsat-8/9 optical imagery, Landsat land surface temperature (LST), and Sentinel-1 SAR, over common geographic footprints. It comprises approximately 2M globally distributed locations, 25M georeferenced patches, and over 40TB of data. Pretraining uses a Joint-Embedding Predictive Architecture (JEPA)-style objective that matches representations between global views and single-sensor local views from the same location. We evaluate SpectralEarth-FM on hyperspectral downstream tasks and standard EO benchmarks following the PANGAEA protocol, achieving state-of-the-art results across both evaluation settings.。

ARXIV