



南京大學  
NANJING UNIVERSITY



智能科学与技术学院  
School of Intelligence Science and Technology

# 面向空间推理的多模态推理大模型

团队成员：黄振宇，付文博、吴阗、何文京  
指导老师：郭兰哲

时间：2025.12.11

# 目 CO 眾 NTS



智能科学与技术学院  
School of Intelligence Science and Technology

1

选题缘由与研究基础

2

实施计划与研究方法

3

建设目标与经费预算



01

# 选题缘由与研究定位



智能科学与技术学院  
School of Intelligence Science and Technology

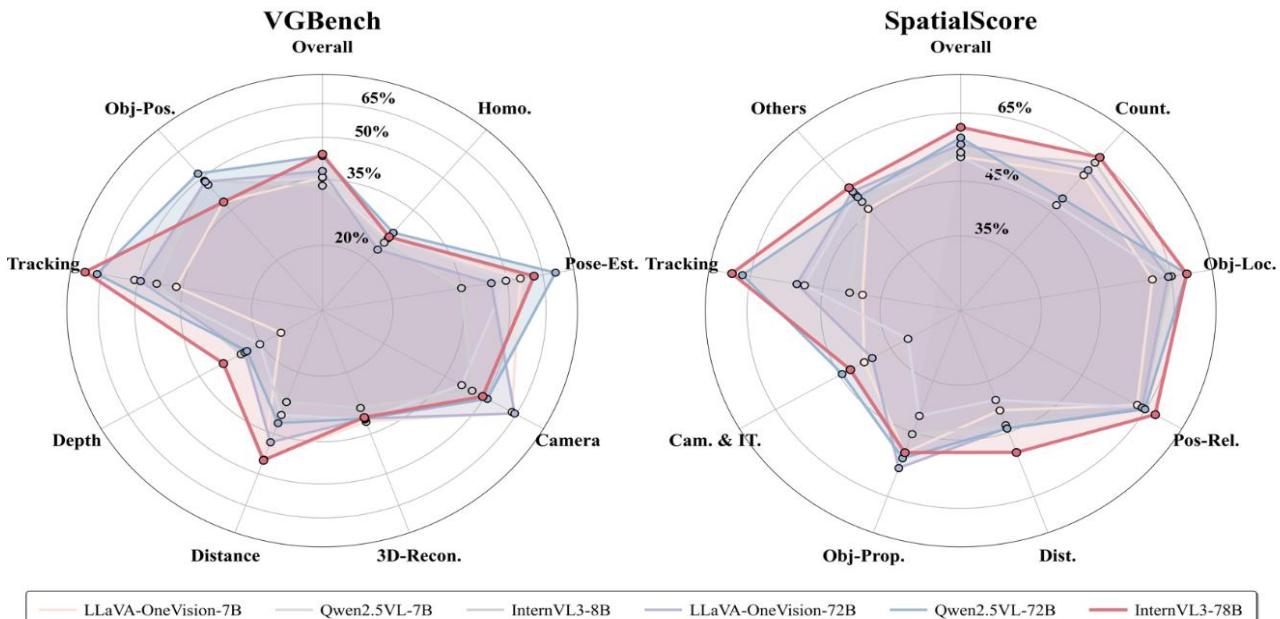


## 现状与痛点

优异的语义理解与生成



精确的空间感知与推理



(b) Current MLLMs still struggle on spatial reasoning, especially visual geometry perception

图片来自：SpatialScore: Towards Unified Evaluation for Multimodal Spatial Understanding

## 解决思路

引擎一（外接知识）：基于“规划-执行”的显式推理范式

核心机制：提示词工程与CoT任务分解  
+ 几何工具调用

解决痛点：对齐空间概念与语义理解，  
确保逻辑闭环

引擎二（内在理解）：轻量级3D空间知识注入与模型结构微调

核心机制：轻量级SFT空间常识注入  
+ 为空间推理任务优化模型架构

解决痛点：增强对空间知识的内在理解，  
提升规划准确率

# 项目定位：从基础评估到分层强化的闭环研究



南京大学  
NANJING UNIVERSITY



智能科学与技术学院  
School of Intelligence Science and Technology

## 相关研究

SpatialScore /  
VGBench: 领域内极具挑战性的空间推理评估基准。

SpatialAgent: 基于 Plan-Execute 与 ReAct 范式的前沿推理框架。

## 项目定位

训练目标: 产出空间推理能力优于基线的模型原型。

开源贡献: 提供空间感知增强的微调数据构建流程与示例数据集

## 预期成果

## 初始评估

基于最新的开源模型 (InternVL/Qwen) 搭建基础系统。

量化分析现有模型在空间推理上的具体弱势点。

1. 免训练范式: 优化 Plan-Execute 链, 参考 ReAct 范式, 提升模型灵活性与工具调用准确度。
2. 模型微调与架构改进: 引入轻量级 SFT, 增加 3D 理解与知识融合模组, 提升内在直觉。

## 核心创新





02

## 实施计划与核心创新



智能科学与技术学院  
School of Intelligence Science and Technology

# 实施计划：“三阶段式”技术路线



南京大学  
NANJING UNIVERSITY



智能科学与技术学院  
School of Intelligence Science and Technology



## 基线搭建与空间诊断

部署基线：复现 SpatialAgent 推理范式作为基线。

量化痛点：精准定位“空间感知与推理盲点”。



## 核心双引擎架构构建

免训练范式增强：构建思维链，调用几何工具。

3D 理解注入：基于 SpatialScore 进行轻量 SFT，增加空间理解与融合模块



## 系统效能评估与复盘

闭环验证：对比基线，验证性能提升。

消融实验：解析双引擎各自的贡献度。



## 阶段1矩阵

阶段目标

部署基线系统：复现 SpatialAgent 推理范式。

量化核心痛点：精准测定现有 MLLM 在空间推理任务上的准确率短板。

实验配置

评估基准：SpatialScore（涵盖视觉几何感知子任务）。

待测模型：选用 Qwen3-VL 等开源 SOTA 模型。

诊断维度

细粒度分析：物体计数、相对距离、3D 姿态等维度分析

失败模式挖掘：“看不清”（感知错误）  
or“算不对”（推理错误）。

产出与价值

明确推理链所需的关键工具组合。  
构建错误分析体系并筛选高难样例。



## 阶段2矩阵

引擎 I——显式推理

Plan-Execute 推理链与 ReAct 范式：引入并改进 SpatialAgent 机制。

引擎 II——隐式感知

轻量级 3D 知识注入：利用 SpatialScore 筛选的高难样本进行 SFT。  
模型架构微调优化：尝试引入空间感知与融合模块，增强模型空间推理能力

关键技术支撑

提示词工程、工具挂载、参数高效微调（LoRA 等）、针对空间推理的架构微调的现有研究

架构协同价值

双层增强闭环：感知增强与推理增强形成互补闭环。  
高效能突破：以低算力成本，实现空间任务鲁棒性与精度的提升。



03

## 进度安排与预期成果



智能科学与技术学院  
School of Intelligence Science and Technology



## 项目进度规划

第一阶段：基线构建  
(第1-3月)

复现 SpatialScore 评测流程，量化模型“空间推理盲点”，筛选 SFT 训练样本。

第二阶段：架构探索  
(第4-7月)

构建 Plan-Execute 推理链与 ReAct 范式（引擎 I），并行开展轻量级 3D 知识微调与模型架构探索（引擎 II）。

第三阶段：评估与分析  
(第8-10月)

完成双引擎消融实验，验证性能提升，撰写论文并整理代码开源。



## 可行性保障基石



### 核心论文研读

深研空间推理方面的论文，确立双引擎技术路线。



### 技术经验积累

熟练掌握 InternVL/Qwen 等基座模型部署，具备 SFT 微调与工具链集成能力。



### 资源支持保障

依托实验室 GPU 算力资源，使用开源的 SpatialScore/ScanNet 等数据集。



## 项目负责人

黄振宇

## 团队成员

付文博

何文京

吴阗

### • 扎实的数理与算法基石

系统修读《离散数学》、《概率论与数理统计》、《最优化方法导论》等核心课程，具备严谨的逻辑思维与数学建模能力，精通 Python 等编程语言，在数据结构、算法分析与设计方面拥有坚实基础，具备复杂问题的代码实现能力。

### • 熟练的工程实践能力

掌握 PyTorch 深度学习框架，具备模型构建与调试能力，熟悉 Hugging Face 体系，拥有开源 LLM/MLLM 的本地部署、运行、微调（SFT/LoRA）等经验。

### • 丰富的科研训练背景

项目成员在大二提前进入实验室课题组学习，定期参与学术组会探讨前沿技术与研究方向，协助完成复杂 Benchmark 构建与评测工作，多次通过 arXiv、顶会（CVPR/ICCV/ACL）等渠道获取并复现高质量文献。

# 预期成果交付与经费投入规划



南京大学  
NANJING UNIVERSITY



智能科学与技术学院  
School of Intelligence Science and Technology



技术原型：产出优化后的可运行模型原型。



学术成果：撰写 CCF C 级及以上会议论文一篇，



工程化：GitHub 开源代码和详细实验报告。



开支科目	预算经费	主要用途	经费占比
计算资源/算力租赁	4000	用于基线复现、VGBench 评测，以及轻量级 SFT 微调所需的 GPU 资源租凭（如 A100 / 4090）。	40%
实验数据与工具	2000	获得/清洗 3D 场景数据、筛选难例样本，并支付必要的商业 API 调用费用。	20%
硬件设备升级	2000	升级 SSD 存储与内存，用于高效管理和预处理大规模 3D 数据集，提高本地实验效率。	20%
文献与学术交流	1000	用于订阅顶会论文库、参加 AI / 计算机视觉 相关学术会议或研讨会，跟进前沿研究。	10%
文档与其他	1000	包括打印技术报告、论文投稿版面费、专利申请费等及其他支出。	10%



南京大學  
NANJING UNIVERSITY



智能科学与技术学院  
School of Intelligence Science and Technology

请各位领导老师批评指正！