

王佳豪

16602536798 | jiahaowang@smail.nju.edu.cn | wang-jiahao.github.io

教育背景

直博 - 南京大学 - 智能科学与技术学院 - 计算机科学与技术 - 刘佳恒老师

2027 年 9 月入学

本科 - 南京大学 - 智能软件与工程学院 - 软件工程 (智能化软件)

2023 年 9 月 - 2027 年 6 月

研究兴趣

多模态大模型及其智能体, 重点关注大模型后训练与评测, 以及 Multimodal Agent。

实习经历

阿里巴巴 - 大模型 Agentic Coding

2026 年 7 月 - 2026 年 10 月

背景与目标: 面向真实环境中的复杂 Agentic 任务, 研究长程规划、多模态感知、工具调用及任务交付的自动化评测。

我的职责: 负责 Agentic 任务数据构建、评测实验及失败轨迹分析, 重点开展 ALE、Claw-Eval 及其多模态任务评测。

- Claw-Eval 多模态 Agentic 数据合成:** 设计 Asset Grounding → Task Synthesis → Verifier Generation → Trajectory Distillation 四阶段 Pipeline; 针对视频 QA、网页复刻与文档提取, 构建时序事件、视觉布局及结构化表格证据, 自动合成复杂任务与评分规则, 并通过 Agent Rollout 蒸馏高质量执行轨迹。
- ALE 长程任务评测:** 基于 Agents' Last Exam (ALE), 接入 Codex CLI、Claude Code 等 Agent Harness, 覆盖 CLI/GUI 混合操作、多模态工具调用及复杂产物交付; 重点评估长程规划、执行效率及任务完成质量。
- 执行轨迹与失败分析:** 收集 Agent 工具调用、执行日志及产物, 分析 Planning、Tool Use、Context Management 等环节的失败模式; 对比不同模型与 Harness 的能力差异, 指导任务筛选、难度分级及评测优化。

论文

OmniCap-IF | 第一作者 | KDD 2027 D&B Track 在投 | 快手可灵合作研究 | Qwen3.8-Omni 官方评测采用

背景与目标: 针对全模态视频描述中复杂指令遵循困难、评测体系缺失的问题, 构建 OmniCap-IF, 并通过高质量数据合成与后训练提升模型的指令遵循及全模态理解能力。

我的职责: 第一作者, 独立负责从研究设计、数据构建、评测框架开发到模型后训练及实验分析的完整流程。

- Benchmark 与评测:** 设计涵盖格式与内容的 50 类约束, 构建 1,920 条复杂指令; 结合规则校验、Temporal IoU 与 QA-based LLM Judge 实现双轨评测, 揭示 Format-Content Tradeoff。
- 训练数据 Pipeline:** 搭建视频筛选 → 约束感知指令合成 → 内容/格式解耦生成的三阶段流水线, 从 14K 高质量视频构建 54K 指令微调数据; 通过负向约束过滤、子任务分解与响应聚合, 提升复杂指令数据质量。
- 模型后训练:** 基于 Qwen2.5-Omni-3B/7B, 采用 LoRA 完成 OmniCaptioner-IF 系列模型的 SFT; 7B 模型在 OmniCap-IF 上 CSR 从 49.19% 提升至 70.73% (+21.54 pp), ISR 从 2.34% 提升至 11.46% (+9.12 pp)。
- 外部 Benchmark 泛化:** 7B 模型在 IF-VidCap 上 CSR 从 56.49 提升至 61.20, Omni-Cloze Total 从 12.90 提升至 32.01, UGC-VideoCap Avg. 从 57.70 提升至 72.97。

AVSCap | 共同一作 | EMNLP 2026 | 快手可灵合作研究

背景与目标: 针对全模态视频描述中的非语音音频遗漏及跨模态事件关联不足问题, 构建音视频协同数据与细粒度 Benchmark, 通过 SFT 和 GRPO 提升 Audio-Visual Synergy 能力。

我的职责: 共同第一作者, 主要负责数据构建 Pipeline、Benchmark、奖励方案设计及模型评测与实验分析。

- 训练数据 Pipeline:** 设计 Decoupled-then-Fused 数据合成流程, 分别提取视觉事件与 Speech、Music、SFX 信息, 通过时序对齐、跨模态事件绑定及自动化质量校验, 构建 AVSCap-130K 高质量三模态训练数据。
- Benchmark 与奖励设计:** 构建包含 1,226 个人工标注视频的 AVSCapBench, 设计 Visual、Audio、Synergy 事件级 Recall 评测协议; 参与设计结合长度约束、语音保真与跨模态协同的 GRPO 混合奖励机制。
- 模型评测与泛化分析:** 负责消融实验及跨基准评测分析; AVSCap-7B 在 UGC-VideoCap 上将平均分从 57.70 提升至 82.40 (+24.70), Daily-Omni Caption-to-QA 从 13.40% 提升至 66.60% (+53.20 pp)。

AdaptiveOmniAgent | 合作作者 | ICLR 2027 在投 | 快手可灵合作研究

- 参与长视频理解智能体研究，提出基于问题粒度的 Global / Local 双模式动态路由，结合音视频事件定位、多模态工具调用与迭代规划；在 OmniVideoBench 上将 Gemini-3.0-Flash 准确率从 55.10% 提升至 62.30% (+7.20 pp)，10–30 分钟视频提升 11.14 pp，平均推理步数从 5.1 降至 4.3。

论文发表

注：* 表示共同第一作者

► 第一作者 / 共同第一作者

[2027 KDD D&B 在投] **J. Wang***, A. Ping*, Y. Wang*, Y. Zhang, S. Li, et al. OmniCap-IF: Benchmarking and Improving Instruction Following Abilities for Omni-Video Captioning.

[2026 EMNLP Main] Y. Wang*, **J. Wang***, J. Tang*, Y. Zhang, Z. Cao, et al. AVSCap: Orchestrating Audio-Visual Synergy for Omni-modal Video Captioning.

► 其他合作论文

[2026 ICML] Z. Cao*, T. Wang*, J. Wang*, ..., **J. Wang**, et al. T2AV-Compass: Towards Unified Evaluation for Text-to-Audio-Video Generation.

[2027 ICLR 在投] L. Zhu, X. Ma, T. Liu, ..., **J. Wang**, et al. StartupBench: Benchmarking General-Purpose Agents on Market-Validated End-to-End Workflows.

[2027 ICLR 在投] Y. Peng, Q. Gu, L. Zhu, ..., **J. Wang**, et al. xDailyBench: Benchmarking LLMs on Professional Consultation for Real-Life Problems.

[2027 ICLR 在投] L. Zhu, H. Li, Y. Yao, ..., **J. Wang**, et al. Socratic Skill Learning: Interleaving Teaching and Practice for Transferable Agent Skills.

项目

朝曦计划（全校仅 43 人入选，其中 15 人入选国自然本科生项目）、2项国家级大创、3项发明专利

技术能力

编程：Python、C++、PyTorch、Docker、Linux、Git

大模型：Transformer、SFT、PPO/DPO/GRPO、vLLM、Multimodal LLM、LLM Agent

语言：CET-6：534、CET-4：580