

柯冠舟

基本信息

邮箱: guanzhouk@gmail.com 主页: [Link](#) GitHub: [Guanzhou-Ke](#) 谷歌学术: [Link](#)

研究方向: 多模态理解, 缺失多模态数据生成 求职意向: 26届毕业生, 多模态理解与生成/Agent 工程师或实习生

研究经历

多模态理解与缺失数据生成 博士课题 2022年9月-2026年6月

- 背景与目标: 多模态数据在现实场景中易受设备故障、隐私保护等因素影响, 导致模态信息不完整, 严重影响模型性能, 常见于医疗诊断、推荐系统等任务中。本工作旨在探索模态间的交互关系与共享语义, 通过理解与建模观测模态, 实现缺失模态的语义补全与数据恢复。
- 成果1: 提出表征解耦与共享建模框架。利用模态掩码建模与解耦知识蒸馏方法, 实现模态间共享表征的无监督建模。实验验证共享表示具备跨模态语义对齐与补全能力, 成果发表于ACM MM'23 [3] 和 CVPR'24 [2]。
- 成果2: 构建知识驱动的无需训练的生成范式。结合预训练模态大模型完成推理式补全, 在多个跨模态缺失场景中显著优于传统对抗生成和条件生成方法。成果发表于CVPR'25 [1]。
- 技术栈: 生成模型, 多模态内容理解与加工, Agentic Workflow, 知识挖掘

医疗报告生成大模型 算法实习生 微软亚洲研究院 2024年2月-2024年10月

- 背景与目标: 当前多模态大模型(如LLaVA)在医疗图文生成任务中广泛应用, 但在生成胸部X光报告时常出现hallucination(幻觉)问题, 表现为虚构病灶、错误诊断等。本工作旨在构建一个医学场景适配的多模态报告生成系统, 提升报告的准确性与医学一致性, 降低幻视发生率。
- 成果1: 基于LLaVA-1.5, 设计了一个端到端的Chest X-ray报告生成系统, 支持图文协同建模与自由生成。提出自适应注意力矩阵机制, 在训练阶段根据医学影像和报告关键字之间相似性动态调整注意力, 强化模型对真实病灶的关注。
- 成果2: 引入医学知识引导(如CheXpert标签或实体对齐)提升文本生成的临床语义一致性。
- 技术栈: 大模型微调, 动态注意力矩阵, Vision-Language Model

多模态新闻伪造检测 算法实习生 中科院自动化所 2023年6月-2023年12月

- 背景与目标: 随着生成式技术的发展, 虚假新闻的表达形式从单一文本向图文音协同生成演变, 传统检测方法难以识别跨模态伪造与模态间语义冲突。本工作旨在构建面向真实世界的多模态新闻伪造检测模型, 充分挖掘图像、音频与文本之间的语义一致性与伪造特征, 提升对复杂多模态伪造的识别能力。
- 成果: 搭建图文音三模态融合检测系统, 集成跨模态表示与对齐模块, 实现新闻内容的结构化建模与伪造识别。提出基于模态解耦的检测方法, 将共享伪造模式与模态特有伪造信息分离建模, 增强模型对隐蔽伪造与协同欺骗的感知能力, 成果发表于CVPR'24 [2]。
- 技术栈: 多模态伪造检测, 跨模态一致性, 图文音融合, 音频数据处理

教育背景

学位	学校	时间	学院	专业	导师
理学博士	北京交通大学	2022-9 至 2026-6	经济管理学院	管理科学与工程	余旻
联合培养博士	新加坡管理大学	2024-10 至 2025-10	计算机与信息系统学院	计算机科学与工程	Shengfeng He(何盛烽)
工学硕士	五邑大学	2019-9 至 2022-6	智能制造学院	系统工程	洪智勇
工学学士	五邑大学	2017-9 至 2019-6	计算机学院	通信工程	

代表论文

- [1] Guanzhou Ke, Shengfeng He, Xiao-Li Wang, Bo Wang, Guoqing Chao, Yuanyang Zhang, Yi Xie, Hexing Su. Knowledge Bridger: Towards Training-free Missing Multi-modality Completion. CVPR 2025. [\[PDF\]](#) / [\[CODE\]](#) (CCF-A)
- [2] Guanzhou Ke, Bo Wang, Xiaoli Wang, Shengfeng He. Rethinking Multi-view Representation Learning via Distilled Disentangling. CVPR 2024. [\[PDF\]](#) / [\[CODE\]](#) (CCF-A)
- [3] Guanzhou Ke, Yang Yu, Guoqing Chao, Xiaoli Wang, Chenyang Xu, Shengfeng He. Disentangling Multi-view Representations Beyond Inductive Bias. ACM MM 2023. [\[PDF\]](#) / [\[CODE\]](#) (CCF-A)
- [4] Guanzhou Ke, et al. "How Far Are We from Generating Missing Modalities with Foundation Models?." arXiv preprint arXiv:2506.03530 (2025). (Under Review)

曾获奖项

- 第十七、十八、十九届“华为杯”全国研究生数学建模比赛全国二等奖, 2020, 2021, 2022年12月
- 第九届蓝桥杯信息竞赛B组全国二等奖, 2018年6月
- 国家奖学金, 2015年12月