

教育背景

- **南京大学 (硕士)** 江苏南京, 2023.09 – 至今
计算机学院 计算机技术专业 导师: 路通, 苏丰 方向: 计算机视觉, 多模态大模型
荣誉: 校级一等学业奖学金, 校级英才奖学金二等, 优秀共青团员, 优秀研究生干部 校内任职: 研究生党支部书记
- **大连理工大学 (本科)** 辽宁大连, 2019.09 – 2023.06
软件学院 软件工程专业 平均成绩: 93.6/100 GPA: 4.55/5.0 成绩排名: 6/383
荣誉: 国家级励志奖学金, 校级一等学习优秀奖学金, 校级精神文明奖学金, 校级社会实践优秀个人二等奖

个人总结

- **技术:** 有 C/C++ 经验, 了解数据结构和算法, Python、Pytorch 的基本技能, 以及一些软件工程
- **科研:** 深入参与多项多模态大模型相关的科研工作, 并参与 InternVL 大型前沿系列模型的探索
- **项目:** 有上海人工智能实验室、字节跳动等企业的实习经历, 具备一定的项目开发和部署能力

实习经历

- **上海人工智能实验室多模态大模型算法工程师实习** 2024.10 – 至今
 - **实习情况:** 研二学年在上海人工智能实验室 (浦江国家实验室) 大模型部门实习, 多模态大模型算法工程师
 - **主要贡献:** 1) 作为主要成员参与完成多模态模型 InternVL 2.5 系列的数据构建、模型探索、模型评测和论文编写等工作; 2) 参与多模态大模型推理能力提升工作 InternVL MPO 中数据集管线搭建、数据构建、实验验证等主要工作; 3) 作为主要成员参与 InternVL 长思维链多模态推理能力扩展工作, 参与长思维链强化学习设计、数据构建、一站式评测框架搭建等主要工作。
- **字节跳动测试开发工程师实习** 2023.01 – 2023.05
 - **实习情况:** 大四学年在字节跳动旗下抖音视界有限公司产品研发和工程架构部实习, 测试开发工程师的开发序列
 - **主要贡献:** 1) 采用 Go 语言 Hertz 框架开发用户流量定时统计项目; 2) 采用 Python 语言 Django 框架开发部门测试数据外泄报警机器人; 3) 团队合作维护和改进了 VR 算法自动化测试平台, 该平台代码量达到近 7 万行, 其中个人修复了十余项 Bug, 并基于消息队列机制实现了批量任务创建、紧急任务中止、任务重建三个子项目。

科研经历

- **InternVL-2.5 技术报告 (共一)** 2024.08 – 2024.12
 - **论文:** Zhe Chen, Weiyun Wang, **Yue Cao**, et al. Expanding Performance Boundaries of Open-Source Multimodal Models with Model, Data, and Test-Time Scaling *arXiv preprint arXiv:2412.05271*, 2024.
 - **方法:** InternVL 2.5 覆盖了从 1B 参数到 78B 参数的先进多模态大语言模型系列, 其中 InternVL2_5-78B 是首个在 MMMU 基准上得分超过 70 的开源模型。在这项工作中, 我们深入研究了模型缩放与性能之间的关系, 系统地探索了视觉编码器、语言模型、数据集大小和测试时间配置的性能趋势。通过对一系列基准的广泛评估, InternVL 2.5 可与 GPT-4o 和 Claude-3.5-Sonnet 等领先的商业模型相媲美。
 - **主要贡献:** 参与完成模型探索、各类数据集构建、模型评测框架搭建、论文编写等主要工作。
- **多模态指令微调数据集 MMInstruct (CCF-A SCIS, 共一)** 2023.11 – 2024.05
 - **论文:** Yangzhou Liu, **Yue Cao**, Zhangwei Gao, et al. MMInstruct: A High-Quality Multi-Modal Instruction Tuning Dataset with Extensive Diversity. *arXiv preprint arXiv:2407.15838*, 2024.

- 方法: 为了解决目前多模态数据集质量低、多样性差的局限性, 我们构造了一个多模型数据集, 包括了 24 个域和 4 种指令类型的 973K 指令数据。为了构造该数据集, 我们创新性地提出了一个利用 GPT-4V、GPT-3.5 和人工修正的指令生成数据引擎。可实现半自动、低成本、多领域的指令数据生成。通过实验, 我们证明在 MMInstruct 上微调多模态大语言模型 (MLLMs) 可显著提高其性能, 从而设定新的先进基准。
- 主要贡献: 参与完成指令生成数据引擎实现、数据集构建、实验验证与论文编写的全流程。
- 多模态大模型推理能力提升 MPO 和 MMPR (CCF-A CVPR 在审) 2024.08 – 2024.11
 - 论文: Weiyun Wang, Zhe Chen, Wenhai Wang, Yue Cao, et al. Enhancing the Reasoning Ability of Multimodal Large Language Models via Mixed Preference Optimization. *arXiv preprint arXiv:2411.10442*, 2024.
 - 方法: 为了解决分布变化对于多模态大型语言模型推理能力特别是思路链 (CoT) 方面的限制, 我们提出了一种混合偏好优化 (MPO) 方法, 以增强 MLLM 的多模态推理能力。同时, 我们开发了一个自动化偏好数据构建管道, 用于创建高质量、大规模的多模态推理偏好数据集 MMPR。实验结果表明, 我们的方法在多个基准测试中均表现出色, 尤其是在多模态推理任务中。我们的 InternVL2-8B-MPO 在 MathVista 上的性能可与 10 倍大的 InternVL2-76B 相媲美。
 - 主要贡献: 主要完成 MMPR 数据集管线搭建、数据集构建、实验验证与论文编写等主要工作。
- 多模态大模型细粒度能力改进 MMFuser (CCF-B ICME 在审, 共一) 2024.03 – 2024.08
 - 论文: Yue Cao, Yangzhou Liu, Zhe Chen, et al. MMFuser: Multimodal Multi-Layer Feature Fuser for Fine-Grained Vision-Language Understanding. *arXiv preprint arXiv:2410.11829*, 2024.
 - 方法: MMFuser 通过简单而高效地集成 Vision Transformers (ViT) 的多层特征, 解决了当前多模态大型语言模型 (MLLMs) 在捕捉复杂图像细节方面的局限性。与依赖 ViT 最后一层特征图的现有 MLLMs 不同, MMFuser 创新地使用语义对齐的深度特征作为查询, 以动态地从浅层特征中提取缺失的细节, 同时保持整体语义对齐。这种方法提供了一种更灵活、更轻量级的新型解决方案。我们的实验表明, 将 MMFuser 应用于现有 MLLMs (e.g. LLaVA-1.5) 上会在各种基准测试中带来显著的改进。
 - 主要贡献: 参与方法创新的前期探索性实验、模型设计、代码编写、实验验证与论文编写的全流程。

项目经历

- 基于多模态可信的行业智能化运维系统研究课题 2024.03 – 2024.12
 - 项目情况: 南京大学-中国移动紫金创新研究院联合项目, 项目经费: 100 万, 核心成员
 - 主要贡献: 1) 负责网络故障预测方面模型研究和部署, 针对预测效率低和实时性不足问题, 提出了多元概率长时序预测模型。2) 参与运维多模态大模型的数据集构建, 并优化模型在细粒度视觉信息方面的理解能力。3) 完成系统的整体部署, 已广泛应用于矿区等多个行业领域。

竞赛经历

2020 年亚太地区大学生数学建模竞赛 国家级二等奖	2020.11
2020 年高教社杯全国大学生数学建模竞赛 国家级二等奖	2020.09
2021 年中国高校大数据挑战赛 国家级二等奖	2021.11
2021 年中国高校计算机大赛网络技术挑战赛 东北赛区一等奖	2021.08