

教育经历

东南大学 人工智能专业，计软智学院 本科在读	2022 – 2026
自动化所 中国科学院 直博生	2026 – 2031

项目经历

OmniReasoner: Thinking with Long Audio-Video via Adaptive Zoom-in | 一作

- 发表情况: 投稿 NeurIPS 2026.
- 项目描述: 我们提出了 OmniReasoner, 一个面向长音视频推理的端到端后训练框架, 使全模态大模型能够先低成本浏览完整音视频, 再根据问题自适应选择关键时间片段进行高保真音视频复查。项目进一步设计了 TimeAnchor 绝对时间锚定机制和 Temporal Augmented Data Engine, 在统一时间轴上对齐音频、视频与局部检索行为, 从而提升模型在长音视频场景下的时序定位、跨模态证据获取和推理回答能力。

Omnivideobench: Towards audio-visual understanding evaluation for omni mllms | 共一第二

- 发表情况: ICLR 2026 [arxiv]
- 项目描述: 我们提出了 OmniVideoBench, 一个面向全模态多模态大模型的音视频协同理解评测基准, 包含来自 628 段多样化视频的 1000 组高质量问答, 并覆盖时序推理、空间定位、计数、因果推断和总结等 13 类问题, 重点评估模型对音频与视觉信息的互补利用和逻辑一致性推理能力。

PixCLIP: Achieving Fine-grained Visual Language Understanding via Any-granularity Pixel-Text Alignment Learning | 共一第二

- 发表情况: ICML 2026 (CCF-A 类会议)。[arxiv]
- 项目描述: 本文提出了 PixCLIP, 这是首个能够同时处理局部视觉提示和长文本输入的模型, 克服了视觉和文本两个维度上的粒度限制。通过自动化构建数据流水线, 我们创建了 LongGRIT 数据集, 包含 150 万条细粒度局部区域描述, 专为提升模型性能而设计。我们提出了一种三元像素-文本对比学习框架, 利用 LLMs 对长文本进行编码, 成功赋予模型将任意图像局部区域与任意长度文本对齐的能力。

A Comprehensive Benchmark for Continual Instruction Tuning of Multimodal LLMs with Chain-of-Thought Reasoning Analysis | 二作

- 发表情况: [arxiv]
- 项目描述: 我们提出了 MLLM-CTBench, 这是一个综合性 MLLMs 持续学习基准测试。

FlowPrune: Accelerating Attention Flow Calculation by Pruning Flow Network | 二作

- 发表情况: NeurIPS 2025 (CCF-A 类会议) [openreview]
- 项目描述: 我们引入了 FlowPrune, 该框架通过应用最大流计算之前对注意力图进行剪枝, 来加速注意力流分析。

STGE-Former: Spatial-Temporal Graph-Enhanced Transformer for EEG-Based Major Depressive Disorder Detection | 一作

- 发表情况: ICASSP 2025 (CCF-B 类会议)
- 项目描述: 我们提出了一种新的时空图增强 Transformer, STGE-Former, 用于基于脑电的抑郁诊断任务

实习经历

华为 ICT 公共开发部-AI 工程师(实习生)	2024.06 – 2024.09
• 主要工作: 多模态大模型训练和推理优化加速; 大模型训练和推理由 GPU 向 NPU 迁移。	

Pinch(AI 初创, 被 Y Combinator 投资) Machine Learning Engineer	2025.01 – 2025.06
• 主要工作: 音色克隆和语音合成模型的研究。面向的场景是 AI 同声传译, 在保证同传过程中翻译准确快速的同时, 还要尽量做到音色一致。	

Shopee LLM Team 预训练-大模型算法实习生	2025.09 – 2026.04
• 主要工作: 多模态 agentic 后训练研究。在实习期间, 我提出了一个面向 Long Audio-Video 推理的 Omni 模型 agentic 后训练框架, 并构建了一个 30K 的后训练数据集, 其中一半数据由我设计的自动化 Data Engine 合成。由于现有框架(VeRL, Slime 等)并不适配 Omni 模型和带有音频输入的工具, 我基于 TRL 实现了 Omni 模型的多轮 tool-use RL 框架, 实现了较为稳定的训练。	

实践经历

东南大学 131AIClub 负责人(2024-2025); 校龙舟队主力队员 (2023-2024)

专业技能

熟悉 Python, 熟练使用 Python 进行数据分析; 熟悉 Pytorch, Transformers 等框架的使用; 熟悉分布式训练/推理, 熟悉 deepspeed, vllm 等训推加速框架; 有英伟达 GPU 以及昇腾 NPU 分布式训练经验, 有 slurm, k8s 集群使用经验; 熟悉 git, 有项目代码管理经验。