

王启年

(+86)189-9358-7909 | eniacyi.wang@outlook.com | eniacyi-wang.github.io

教育经历

上海交通大学 | 电子信息，人工智能学院 | 硕士研究生 2025.09—2028.03

主要研究方向为基于多模态大模型的视频理解，在视觉理解和多模态大模型微调方面有一定的研究和工程经验，有独立开发和维护开源代码的能力

电子科技大学 | 微电子科学与技术，集成电路科学与工程学院 | 工学学士 2021.09—2025.06

项目经历

NeSeg: An Agentic System for Video Segmentation with Negative Hints 2025.10—2026.03
(Core Contributor) [Github]

- 项目背景:** 尽管多模态大模型在视频物体分割上取得了显著进展，现有工作（如 VideoLISA 等）仍依赖 SAM 在隐空间进行掩码解码，且已有的 Agent 式方案仅支持单轮正样本提示，缺乏对分割结果的多轮交互式修正能力。
- 核心创新:** 提出首个支持负样本提示的 Agent 式视频分割系统 NeSeg，使用 VeRL 训练框架，基于 Qwen2.5-VL-7B-Instruct 进行监督微调与 GRPO 强化学习训练；设计了面向分割任务的 SFT 策略与多维度 RL 奖励函数，在 RefCOCO 与 Reason VOS 数据集上性能持平同类模型 OneThinker，达到领域先进水平。
- 数据构造:** 基于 ReVOS、RVOS 与 SA-V 等公开分割数据集，构建了一套自动生成训练数据的流水线，可高效产出 SFT 所需的指令数据与 RL 所需的问题-掩码对；该流水线累计生成了 5k+ 条高质量训练样本。
- 监督微调:** 在 SFT 阶段，引导模型同时学习时序定位、空间定位与正负样本提示能力。通过掩码内随机采样生成正样本点，并利用 DINOv2 特征挖掘困难负样本点，构建了包含正负提示的多样化训练数据。
- GRPO 强化学习:** 为赋予模型多轮修正能力，采用 GRPO 进行训练，并设计了复合奖励函数，涵盖格式奖励、分割准确性奖励及工具调用奖励。模型在多轮交互或达到预设轮次上限后输出最终分割结果，与真值掩码对比产生准确性奖励，从而学会在高效率交互下逐步优化分割质量。

StreamTeller: Perception-First Memory with Event-Driven Visual Evidence 2025.02–Present
(In Progress) (Core Contributor) [Github]

- 项目背景与动机:** 流式视频理解中，现有方法主要集中于 KV Cache 压缩、视觉 token 压缩及文本摘要压缩等，显著降低了流式视频推理成本。然而，现有方法面临历史记忆干扰当前感知以及文本压缩丢失细粒度视觉证据等关键挑战，难以兼顾感知能力、长期事件记忆与视觉细节。
- 核心方法:** 提出事件驱动的层次化多模态交错记忆机制，作为 training-free 插件接入现有多模态大模型。视频流被组织为事件节点，每个节点包含结构化 caption 与可选的视觉证据。通过事件驱动的门控 caption 策略，仅在语义漂移时触发描述以降低开销；利用 caption 作为语义索引，仅在文本无法表达时保留绑定的视觉证据。
- 记忆检索与优势:** 采用感知优先的按需检索策略：当前画面问题由最近帧回答；历史问题依据 caption 锚定检索，仅在需要颜色、位置等细粒度信息时才加载视觉证据。该方法显式分离感知与记忆，在提升长期 recall 的同时有效抑制记忆对当前感知的干扰。

科研经历

Weaver: End-to-End Agentic System Training for Video Interleaved Reasoning

CVPR 2026 Findings (Co-Author) [arXiv]

- 项目背景:** 现有视觉推理方法多依赖纯文本链式思维进行逻辑推演，面临视觉与语言模态表征不匹配的固有局限，且对细粒度视觉线索的感知敏锐度不足，限制了复杂视频场景下的推理能力。
- 项目创新:** 提出了端到端可训练的 Agent 式多模态推理系统 Weaver，首次实现基于工具调用的交错式多模态推理。系统能够在推理过程中自主调用工具逐步提取关键视觉证据，构建真实的视觉-语言交织推理轨迹，在长视频理解任务上达到领域先进水平。
- 核心贡献:** 进行模型的系统性评估与基线方法的公平复现，负责多组对比分析，保障了实验结论的可靠性与可复现性。

技术能力

- 熟练掌握 Python、C/C++、C# 等编程语言
- 熟练掌握 Linux、Shell、Git 等工作流
- 熟练掌握 PyTorch、Huggingface Transformers 等深度学习框架，熟练掌握 VeRL、TRL 等 RL 训练框架，熟悉 vLLM 等推理框架。