刘熠

+86-13006363023 yiliu61richard@gmail.com www.yiliu.uno
Google Scholar：总被引 2500+，h-index 9

基本信息

出生年月：1995.08 籍贯：湖北黄冈 政治面貌：**中共党员** Base：上海

原生多模态方向算法负责人：端侧 VLM 全栈研发（数据合成、课程化模态注入训练、NPU 部署落地）与多模态 Agent 后训练（RL 数据构建、评测体系、业务落地）。**MagicVL-2B**（一作）落地荣耀旗舰 Magic V5；TIP / IJCV / SPL 一作及共一论文 6 篇；ECCV Ego4D 挑战赛双赛道第 1 名。

研究方向：原生多模态大模型、多模态数据合成、RL 后训练、多模态 Agent。

教育经历

中国科学院大学 模式识别与智能系统 博士 2019.09 – 2024.07
深圳先进技术研究院 MMLab@SIAT 博士导师：乔宇教授 / 王亚立研究员 北京/深圳

华中科技大学 自动化理工交叉创新实验班 本科 2015.09 – 2019.07
人工智能与自动化学院 湖北武汉

工作经历

字节跳动 Commercial AI-AIGC **创意素材 Agent 方向负责人（算法）** 2025.11 – 至今

主导广告创意素材 Agent（商品图 + 创意灵感 → 分镜脚本 → 视频生成）的模型研发与全链路落地。

- **Agent 模型建设**：基于大盘投放数据建成 60 万级商品素材片段库与逐帧标注 GT 数据管线；SFT 对齐真实投放画面分布，创意画面设计一致性全面优于 8 个版本的通用大模型（胜率 62.6%–86.5%）。
- **RL 生成指令优化**：面向 Seedance-2.0 构建 KTO 强化学习训练数据与双重奖励机制（AIGC 质检细粒度评分 + 创意还原度门控，防止 reward hacking）；交互/人物 badcase 通过率 +23pp/+26pp，整体质检通过率 16%→28%，人工评估确认画面一致性无退化。
- **评测体系建设**：建立字面相似度、多维度评分、元素级匹配、pairwise 偏好对比 4 类自动评估方法，以 200 例人工双盲标注校准——自动评估与人工判定的一致性 κ=0.59，接近标注员间一致性 0.60；自建推理链路成本约为商用 API 的 1/10。
- **业务落地与治理**：牵头制定 P0 badcase 分类与质检标准，联合质检、生成模型与产品运营团队建设 benchmark 及分层训练数据；规划 Agent 版本演进路线，推动素材达到动态创意上线标准。

荣耀终端 **大模型能力平台部** **多模态理解与生成组组长** 2024.04 – 2025.10

实线管理 10+ 人团队，负责端侧 VLM 技术预研与业务落地，并主导核心训练方案与数据管线的设计实现。

- **端侧多模态大模型**：主导 **MagicVL-2B**（一作）研发——轻量视觉编码器（<100M 参数）+ 课程学习式多阶段模态注入（缓解模态增量训练中的灾难性遗忘）；对比同尺寸 InternVL2.5-2B **功耗降低 41.1%**，并在 MMBench / MMStar / OCRBench / AI2D 等主流基准上领先；落地荣耀旗舰折叠屏 **Magic V5**。
- **多模态数据合成**：负责端侧 VLM 训练数据的清洗与构建（累计 100B token 量级）；建成短视频结构化描述、屏幕知识 GUI 两条数据合成管线，支撑视频描述与 **GUI Agent** 业务。
- **视频理解大模型**：主导短视频细粒度描述模型 **VideoCap-R1**（结构化思维 + RL 后训练，CAREBench 物体 F1 超越 GPT-4o-mini）与长视频高效理解模型 **E-VRAG**（资源受限下的检索增强）研发。

上海人工智能实验室　通用技术研究部　见习研究员（实习） 2022.03 – 2023.12

- 参与视频理解基础模型 **InternVideo** 研发与时序定位等下游任务的对齐训练。
- 参与构建多模态视频评测基准 **MVBench**（CVPR 2024）与长视频理解基准 **LvBench**（IJCV 2025）。

论文成果

[1] MagicVL-2B: Empowering Vision-Language Models on Mobile Devices with Lightweight Visual Encoders via Curriculum Learning, arXiv, 2025 (第1作者)

[2] FineAction: A Fine-Grained Video Dataset for Temporal Action Localization, IEEE Transactions on Image Processing, 2022 (TIP, 中科院1区, IF=13.7, 第1作者, 被引 137)

[3] LvBench: A Benchmark for Long-form Video Understanding with Versatile Multi-modal Question Answering, International Journal of Computer Vision, 2025 (IJCV, 中科院1区, IF=9.3, 共同一作)

[4] MLLM-TA: Leveraging Multimodal Large Language Models for Precise Temporal Video Grounding, IEEE Signal Processing Letters, 2024 (SPL, 中科院2区, IF=3.9, 第1作者)

[5] VideoCap-R1: Enhancing MLLMs for Video Captioning via Structured Thinking, Computer Vision and Image Understanding, 2026 (CVIU, CCF-B, IF=3.5, 共同通讯)

[6] E-VRAG: Enhancing Long Video Understanding with Resource-Efficient Retrieval Augmented Generation, arXiv, 2025 (第1通讯)

[7] MVBench: A Comprehensive Multi-modal Video Understanding Benchmark, Conference on Computer Vision and Pattern Recognition, 2024 (CVPR, CCF-A会议, 被引 1400+, 第6作者)

[8] Dual Masked Modeling for Weakly-supervised Temporal Boundary Discovery, IEEE Transactions on Multimedia, 2023 (TMM, 中科院1区, IF=9.7, 共同一作)

[9] Learning Discriminative Feature Representation for Open Set Action Recognition, ACM International Conference on Multimedia, 2023 (ACM MM, CCF-A会议, 共同一作)

[10] InternVideo: General Video Foundation Models via Generative and Discriminative Learning, arXiv, 2022 (被引 700+, 第9作者)

[11] F2S-Net: Learning Frame-To-Segment Prediction for Online Action Detection, Journal of Real-Time Image Processing, 2024 (JRTIP, 中科院3区, IF=3.0, 第1作者)

荣誉奖项

- **ECCV 2022 Ego4D Episodic Memory Challenge（Looking at Me Track）第 1 名**
- **ECCV 2022 Ego4D Episodic Memory Challenge（Moment Queries Track）第 1 名**
- 2025年　荣耀终端-AI平台部「闪耀之星」
- 2024年　**荣耀终端-研发管理部总裁团队「AI 多模态竞争力突破团队」**
- 2023年　中国科学院大学「三好学生标兵」
- 2020–2022年　中科院深圳先进技术研究院「院长优秀奖」（连续三年）