

邹元昊 | Yuanhao Zou

目标地点: 上海 • yuanhaoz@ucf.edu • 16680807958
个人主页 • GitHub • Google Scholar • LinkedIn

求职方向: 视频理解、长视频推理与时序定位、高效/端侧视觉语言模型、多模态智能体与强化学习。
实习时间: 2027.05 - 2027.09, 每周 5 天, 可连续实习 4 个月; 国内实习无需 CPT 审批。
研究优势: 聚焦 “更少计算、更强时序推理” 的多模态模型; 拥有 ICLR、CVPR、ECCV 等顶会论文与工业研究经历。

教育背景

University of Central Florida (中佛罗里达大学) | 计算机科学博士 2025.08 - 至今
• GPA: 4.0/4.0; 导师: **Prof. Chen Chen** (Center for Research in Computer Vision)
University of Michigan, Ann Arbor (密歇根大学安娜堡分校) | 电子与计算机工程硕士 2023.08 - 2025.05
• GPA: 4.0/4.0
中南大学 | 计算机科学与技术学士 2019.09 - 2023.06
• GPA: 88/100

代表性论文

- NeurIPS 2026 投稿 | 一作** Conditional Multi-Event Temporal Grounding in Long-Form Video
- ECCV 2026 | 二作** CLARITY: Medical World Model for Guiding Treatment Decisions by Simulating Context-Aware Disease Trajectories in Latent Space
- CVPR 2026 Workshop | 共同一作** How Should Video LLMs Output Time? An Analysis of Efficient Temporal Grounding Paradigms
- ICLR 2026 | 一作** A.I.R.: Enabling Adaptive, Iterative, and Reasoning-based Frame Selection for Video Question Answering
- CVPR 2025 | 一作** Alignment, Mining and Fusion: Representation Alignment with Hard Negative Mining and Selective Knowledge Fusion for Medical Visual Question Answering
- CVPR 2025 Workshop | 一作** MVCM: Enhancing Multi-View and Cross-Modality Alignment for Medical Visual Question Answering and Medical Image-Text Retrieval
- Knowledge-Based Systems | 共同一作** HFA-UNet: Hybrid and Full Attention UNet for Thyroid Nodule Segmentation

实习经历

联影智能 (United Imaging Intelligence) | 多模态算法研究实习生 (全职) 2026.05 - 2026.08
• **多模态视频推理智能体:** 面向热成像、深度与多视角视频, 训练基于强化学习的智能体, 解决需要跨模态信息联合建模与复杂推理的视频问答任务。
Axon | 视频理解算法研究实习生 (兼职) 2025.10 - 2025.12
• **高效视频时序定位:** 为 Axon AB4 等端侧设备设计连续时间解码与非自回归定位范式; **1.5B** 模型在 Charades-STA 上达到 **57.1 mIoU**, 超过 **7B** 基线的 27.5, 并降低推理延迟。成果发表于 **CVPR 2026 Workshop (共同一作)**。

科研经历

University of Central Florida | Prof. Chen Chen 课题组 2025.02 - 至今
• **自适应视频帧选择 (A.I.R.):** 以轻量 CLIP 模型引导视觉语言模型迭代选帧, 平均仅分析约 **32 帧** (基线 128 帧), 在 NExT-QA / LongVideoBench 上分别提升最高 **7.0% / 4.5%**, VLM 推理时间降低约 **74%** (162s $\rightarrow$ 42s)。成果发表于 **ICLR 2026 (一作)**。
• **条件多事件时序定位 (CoMET):** 构建包含 **600** 个长视频 (平均 33.8 分钟)、**2,789** 条多条件查询的基准; 提出免训练智能体, F1@0.5 相比 GPT-5 提升 **6.1**。项目与 **Qualcomm** 合作, 投稿 **NeurIPS 2026 (一作)**。
• **高效视觉语言模型:** 研究可部署于端侧设备的紧凑型 VLM, 兼顾视觉感知、时序建模和推理效率。
Stony Brook University | Zhaozheng Yin 教授课题组 2024.02 - 2024.11
• **医学视觉语言模型:** 面向医学 VQA 与图像-报告检索, 提出硬负样本挖掘、表示对齐与选择性知识融合方法; 在 VQA-RAD 上达到 **80.5%** (较 BiomedCLIP 提升 **5.3%**), 在 SLAKE / PathVQA 上达到 **86.7% / 65.3%**。成果包含 **CVPR 2025 一作论文**。
University of Nottingham Ningbo China | Xiangjian He 教授课题组 2022.09 - 2023.08
• **医学图像分割:** 融合 U-Net、Transformer、混合注意力与多尺度特征, 解决小样本甲状腺结节超声分割; 在 DDTI 上达到 **85.2% Dice / 74.6% mIoU**, 较此前 SOTA 提升 **2.4% / 1.0%**, 并在三个公开数据集上验证跨域泛化。成果发表于 **Knowledge-Based Systems (共同一作)**。

技术能力

编程语言: Python, C++ **深度学习:** PyTorch, PyTorch Lightning, Hugging Face Transformers, Accelerate, DeepSpeed, vLLM
工具与基础设施: Linux, Git, Docker, Slurm, 多 GPU / 分布式训练, Weights & Biases **常用库:** NumPy, Pandas, OpenCV

学术服务

审稿: NeurIPS 2026 (4 篇投稿), CVPR 2026 Efficient Deep Learning for Computer Vision Workshop, ECCV 2026 Medical Video Understanding Workshop, IEEE Transactions on Multimedia, The Visual Computer, Journal of Imaging Informatics in Medicine 等。