Yuanhao Zou

Orlando, FL | yuanhaoz@ucf.edu | 734-358-6400| LinkedIn

Education

University of Central Florida, Orlando, PhD. in Computer Science — Aug 2025 – Present

- **Advisor:** Chen Chen
- **Research Interest:** Video Understanding, Efficient Vision Language Model, Video Anomaly Detection.

University of Michigan, Ann Arbor, MS in Electrical and Computer Engineering — Aug 2023 – May 2025

- GPA: 4.0/4.0
- **Coursework:** Computer Vision (A), Machine Learning (A), Advanced Computer Vision (A), Robotic Mathematics (A^+), Large Language Model.

Central South University, China, BE in Computer Science and Technology — Sep 2019 – Jun 2023

- GPA: 88/100
- **Coursework:** Machine Learning, Distributed System & Cloud Computing, Digital Image Processing, Computer Vision, Android Development, Bioinformatics, Computer Architecture, Data Structure & Algorithm.

Main Publications

- [**CVPR 2026 Submission, Second Author**] CLARITY: Medical World Model for Guiding Treatment Decisions by Simulating Context-Aware Disease Trajectories in Latent Space
- [**ICLR 2026 Submission, First Author**] ***A.I.R.***: Enabling Adaptive, Iterative, and Reasoning-based Frame Selection For Video Question Answering
- [**CVPR 2025, First Author**] Alignment, Mining and Fusion: Representation Alignment with Hard Negative Mining and Selective Knowledge Fusion for Medical Visual Question Answering
- [**CVPR 2025 Workshop, First Author**] MVCM: Enhancing Multi-View and Cross-Modality Alignment for Medical Visual Question Answering and Medical Image-Text Retrieval
- [**Knowledge-Based Systems, First Author**] HFA-UNet: Hybrid and Full Attention UNet for Thyroid Nodule Segmentation

Internship

Axon (Part-Time) — Oct. 2025 – Dec. 2025

- **Efficient Vision Language Model:** Developing efficient video anomaly detection model for edge devices like Axon AB4, effectively handling images in low-level and extreme conditions.

Research Experience

University of Central Florida, Prof. Chen Chen — Feb. 2025 – Present

- **Video Understanding:** Frame Selection Model integrates fast lightweight models like CLIP and heavy vision language model, achieving better performance and efficiency. A paper submitted to **ICLR 2026**
- **Efficient Vision Language Model:** Developing efficient vision language models that used on edge devices. A current internship with Axon.

Stony Brook University, Prof. Zhaozheng Yin — Feb 2024 – Nov 2024

- **Medical Vision-Language model:** (1) Addressed challenges in cross-modality understanding and the underutilization of multi-view images in medical radiology datasets, and applied to Medical VQA and Medical Image-Report Retrieval. (2) Developed a unified representation alignment approach with hard negative mining and selective knowledge fusion to enhance Med-VQA performance significantly.
- **Outcomes:** First Author of two research papers to **CVPR 2025**.

University of Nottingham Ningbo China (UNNC), Prof. Xiangjian He Sep 2022 – Aug 2023

- **Medical Image Segmentation:** Developed a deep learning network integrating U-Net and Transformer for hybrid attention and multi-scale fusion modules to address challenges with limited samples and small objects in a new cervical dataset.
- **Outcomes:** a paper published to Journal **Knowledge-Based Systems**.

Other Publications

- [**International Workshop on Advanced Imaging Technology (IWAIT) 2024**] Multimodality semisupervised learning for ophthalmic biomarkers detection.
- [**International Workshop on Advanced Imaging Technology (IWAIT) 2024**] Alignment, Mining and Fusion: Representation Alignment with Hard Negative Mining and Selective Knowledge Fusion for Medical Visual Question Answering.

Technologies

Proficient Technologies: Pytorch, Linux, Lightning, Git