MingYu Zhao

Email: zhao.mingyu@rutgers.edu | Tel: 732-207-3192

EDUCATION

Rutgers University, Graduate School of Arts and Sciences — New Brunswick, NJ
Ph.D. in Computer Science — Expected May 2029

University of Science and Technology of China — Hefei, China
B.Eng. in Artificial Intelligence — Jun 2024

PUBLICATIONS

"" denotes equal contribution.*

First-author Publications

- Y. Zhou*, **Mingyu Zhao***, Z. Wang, D. Gu, B. Guo, R. Ye, L. Han, C. Jin, D. N. Metaxas. "M3-Bench: Multi-Modal, Multi-Hop, Multi-Threaded Tool-Using MLLM Agent Benchmark." Submitted to the *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2026.
- **Mingyu Zhao**, Z. Yang, Y. Zhou, Z. Xia, C. Jin, X. He, C. Neidle, D. N. Metaxas. "MHB: Multimodal Handshape-aware Boundary Detection for Continuous Sign Language Recognition." Funded by the U.S. National Science Foundation (NSF).
- Z. Li*, **Mingyu Zhao***, X. Dong, H. Ling, B. Huang. "CAPAA: Classifier-Agnostic Projector-Based Adversarial Attack." Accepted by the *IEEE International Conference on Multimedia and Expo (ICME)*, 2025.

Other Publications

- C. Jin, Y. Li, **Mingyu Zhao**, S. Zhao, Z. Wang, X. He, L. Han, T. Che, D. N. Metaxas. "Lor-VP: Low-rank Visual Prompting for Efficient Vision Model Adaptation." *International Conference on Learning Representations*, 2025.
- S. Zhang, X. He, D. Liu, Z. Xia, **Mingyu Zhao**, C. Tan, V. Li, B. Liu, D. N. Metaxas, *et al.* "Large Sign Language Models: Toward 3D American Sign Language Translation." *IEEE/CVF Winter Conference on Applications of Computer Vision*, 2026.

RESEARCH EXPERIENCE

Topic: Multi-Modal, Multi-Hop, Multi-Threaded MLLM Agent Benchmark — *Aug. 2025 ~ Nov. 2025*
Instructor: Prof. Dimitris N. Metaxas (Rutgers University)
Goal: Design a **multimodal** benchmark for evaluating **MLLM** tool use under the Model Context Protocol (**MCP**), covering realistic multi-hop and multi-threaded workflows with auditable alignment and interpretable evaluation.

- Designed the benchmark task suite and standardized trajectories for multimodal MCP workflows (visual grounding, cross-tool dependencies, and persistent intermediate resources across steps), supported by an Executor & Judge pipeline with optional lightweight human verification.
- Built an end-to-end evaluation pipeline for **MLLM training/inference** settings with MCP tool integration, enabling controlled assessments across tasks, budgets, and error types to characterize failure modes in realistic tool-use workflows.
- Proposed a structure-aware metric suite with similarity-driven tool-call alignment: embed tool-call signatures with a **sentence encoder** and apply **similarity-bucketed Hungarian matching** to obtain auditable one-to-one correspondences, decoupling semantic fidelity from workflow consistency.

Topic: Multi-Modal Video Temporal Boundary Detection — *Nov. 2024 ~ Jul. 2025*
Instructor: Prof. Dimitris N. Metaxas (Rutgers University) and Prof. Carol Neidle (Boston University)
Goal: Improve **temporal boundary detection** for human-action videos via **multimodal** fusion of skeletal features and handshape cues (evaluated on sign language video datasets).

- Designed a spatio-temporal convolution-based segmentation module with velocity and acceleration features from skeleton sequences, improving robustness around sign boundaries.
- Integrated handshape information by pretraining a handshape classifier over canonical categories and fusing its features into the segmentation stream via cross-attention with a gating mechanism to enhance boundary localization.
- Built an end-to-end Continuous Sign Language recognition pipeline by combining the boundary detector with a recognition model and pre-segmented continuous signs, and validated effectiveness through experiments and ablations.

Topic: CAPAA: Classifier-Agnostic Projector-Based Adversarial Attack — *June. 2024 ~ Oct. 2024*
Instructor: Prof. HaiBin Ling (Stony Brook University) and Prof. BingYao Huang (Southwest University)
Goal: Achieve classifier-agnostic and **pose-agnostic projector-based** adversarial attacks.

- Proposed an algorithm that focuses the adversarial perturbations on the physical adversarial samples themselves, so as to cope with the effects of changing camera poses.
- Designed a novel optimal loss function to generate adversarial attacks against multiple classifiers.
- Validated the effectiveness of the proposed CAPAA by building a camera-projector system in the real world and by evaluating the attack success rate and stealthiness on 10 sets of ablation experiments.

TECHNICAL STRENGTHS

Skills: Python, C/C++, Bash, MATLAB, Node.js, PyTorch, TensorFlow, Git, Vercel, Flask, Slurm, Linux, LaTeX.