Jie Zhu

zhujie4@msu.edu — +1 (202) 758-8919 — Google Scholar — LinkedIn — Github — Personal Website

Research Interests Multimodal: MLLMs (Under-review), Agents (Submitting to CVPR26), VQA (MM23), Multi-modal Biometrics (ICCV25, TPAMI (under review))

PUBLICATIONS

Selected Publications

- **A Quality-Guided Mixture of Score-fusion Experts Framework for Human Recognition**. ICCV 2025.
  Jie Zhu, Yiyang Su, Minchul Kim, Anil Jain, and Xiaoming Liu. [Paper] [GitHub]
  Keywords: Multi-modal, MoE
- **FusionAgent: A Multimodal Agent with Dynamic Model Selection for Human Recognition.** (Submitting to CVPR26)
  Jie Zhu, Yiyang Su, Xiao Guo, Anil Jain, and Xiaoming Liu.
  Keywords: Agents, MLLMs, Refinforcement Learning
- **ReFine-RFT: Improving Reasoning Capability of MLLMs for Fine-grained Visual Understanding.** (Submitting to CVPR 2026)
  Jie Zhu, and Xiaoming Liu. [Paper]
  Keywords: MLLMs, Reinforcement Learning, Visual Understanding

Others

- **Subtoken Image Transformer (SiT) for Generalizable Fine-grained Understanding.** (Under review)
  Jie Zhu, Minchul Kim, Zhizhong Huang, and Xiaoming Liu. [Paper]
  Keywords: Image Tokenization, Fine-grained Understanding
- **ATM: Action Temporality Modeling for Video Question Answering.** ACM MM 2023.
  Junwen Chen, Jie Zhu, and Yu Kong. [Paper][GitHub]
  Keywords: VQA, Action Understanding
- **Person Recognition at Altitude and Range: Fusion of Face, Body Shape and Gait.** (TPAMI (under review))
  Liu Feng, ..., Jie Zhu, et al. [Paper]
  Keywords: Multi-modal
- **Fairness-Sensitive Policy-Gradient Reinforcement Learning for Reducing Bias in Robotic Assistance.** IEEE ROMAN 2024.
  Jie Zhu, Mengsha Hu, Amy Zhang, and Rui Liu. [Paper]
  Keywords: Reinforcement Learning, Fairness

EDUCATION

Michigan State University, United States — Aug 2023 – Apr 2028
Doctor of Philosophy in Computer Science — GPA: 4.0/4.0
Research Areas: Multi-modal, MLLMs, and Biometrics

George Washington University, Washington, DC, United States — Sep 2021 – May 2023
Master of Science in Computer Science — GPA: 3.9/4.0

Northeastern University, Shenyang, China — Aug 2016 – Jun 2020
Bachelor of Science in Computer Science — GPA: 3.2/4.0

RESEARCH EXPERIENCE

FusionAgent: A Multimodal Agent with Dynamic Model Selection for Human Recognition
(Agentic System, MLLM, Refinforcement Learning)

- Developed **FusionAgent**, an agentic MLLM framework that dynamically selects the optimal tool combination for each test sample via *multi-turn reasoning and ReAct design* through GRPO reinforcement fine-tuning. Proposed a *metric-based reward* to supervise model selection.
- Introduced an *Anchor-based Confidence Top-k (ACT)* fusion method for adaptive score integration, achieving **+13.2%** on CCVID, **+7.5%** on MEVID, and **17.7%** on LTCC benchmark over SoTA baselines.

ReFine-RFT: Improving Reasoning Capability of MLLMs for Fine-grained Visual Understanding
(MLLMs, Refinforcement Learning, Visual Understanding)

- Tackled the challenge of fine-grained visual recognition (FGVR) in MLLMs by proposing the **ReFine-RFT** framework using GRPO combined with and the novel **MLLM-based reasoning reward**. Introduced a pipeline to measure the reasoning quality via multiple dimensions.
- Achieved SoTA performance across six FGVR benchmarks, improving average accuracy (**8.2%**) on and reasoning fidelity (**+34.8%**) while maintaining strong generalization to general tasks.

A Quality-Guided Mixture of Score-Fusion Experts Framework for Human Recognition
(Multi-modal, MoE)

- Addressed the challenge of whole-body biometric recognition by proposing the **QME** framework, which uses a *modality-specific Quality Estimator* trained with the proposed *pseudo quality loss* to dynamically weight multiple score-fusion experts (MoE), and a novel *score-triplet loss* to directly align score distributions across modalities.
- Demonstrated substantial improvements over prior methods, achieving **5.6%**, **6.8%**, **6.2%**, **12.3%** overall improvement on CCVID, MEVID, LTCC, and BRIAR benchmarks in challenging multi-modal real-world scenarios.

ATM: Action Temporality Modeling for Video Question Answering. (MM 2023)
(VQA, Action Understanding)

- Tackled the challenge of temporal reasoning in VideoQA by developing **ATM**, an action-centric framework that models fine-grained temporal dynamics through *Action-centric Contrastive Learning (AcCL)* and a *Temporal Sensitivity-aware Confusion (TSC)* loss to mitigate static bias.
- Achieved absolute performance gains of **+2.1%** on NExT-QA and **+5.8%** on TGIF-QA, establishing new SoTA results for temporal reasoning and action understanding in Video QA.

WORK EXPERIENCE

Inter-American Development Bank United States
AI Analytics Consultant - (LLM, Web Design) Jun 2023 – Aug 2023

- Engineered multilingual web scraping pipelines (50+ media outlets) using BeautifulSoup and Scrapy, reducing data collection latency by 40%.
- Built a **ChatGPT-powered dashboard** enabling automated summarization and trend analysis of 10,000+ daily text/video news items.

Research of Institute of Tsinghua, Pearl River Delta Guangzhou, China
AI Engineer (Text-to-Speech) Sep 2020 – Aug 2021

- Designed phoneme-based text normalization pipeline with **Tacotron 2**, improving Mandarin TTS correctness by 15%.
- Constructed proprietary speech dataset of 100,000+ clean/noisy audio samples; filed **14 CN patents** (2 first-inventor, 10 granted).

HONORS & AWARDS

Graduate Tuition Fellowship (3 out of 74) Aug 2022
Faculty Awards of Computer Animation Dec 2021
Third Prize Scholarship Sep 2018 – Jul 2020

ACADEMIC SERVICES

- **Reviewer:** TPAMI 2025; FG 2024-2025; IJCBLRR 2024
- **Teaching Experience:** Computer Animation (Fall 2022), Computer Graphics II (Spring 2023)
- **Others:** Maintained the CVLab website and assisted in preparing research grant proposals.

SELECTED PATENTS

CN113194348B, "Virtual human lecture video generation method, system, device and storage medium". Jul. 2022.
CN113192161B, "Virtual human image video generation method, system, device and storage medium". Oct. 2022.