

Yang Zhongyu

☎ +86 13437555149 ✉ yangzhy21@gmail.com 🌐 [Homepage](#) 📄 [Google Scholar](#) 🐙 [GitHub](#)

RESEARCH INTERESTS

Fields: Multimodal Learning, Vision-Language Models, Generative AI, Image Understanding

Topics: Multimodal Large Language Models; Diffusion-based Image Generation; Retrieval-Augmented Reasoning

Objective: My long-term goal is to build general-purpose multimodal systems that can perceive, reason, and communicate effectively across visual, textual, and behavioral modalities in dynamic, real-world environments.

EDUCATION

- **Lanzhou University (Project 985)** Sept. 2021 - June 2025
B.S. in Mathematics(the Basic Theory Class)(Main major) and Administrative Management (Minor) Lanzhou, China
 - Relevant courses: Mathematical Analysis, Advanced Algebra, C++ Programming, Probability Theory, Ordinary Differential Equations, Numerical Analysis, Microeconomics, Differential Geometry, Functional Analysis, etc.
 - **Thesis:** [Adaptive Multi-task Medical Image Restoration](#), Supervised by [Prof. Yu-Mei Huang](#) and rated as A+ (Top 1%)
- **King Abdullah University of Science and Technology** Dec. 2024 - Present
Remote Research Intern in [Vision-CAIR](#) Group Saudi Arabia
 - Advisor: [Dr. Jun Chen](#) and [Prof. Mohamed Elhoseiny](#)
- **The Chinese University of Hong Kong, Shenzhen** April. 2024 - Nov. 2024
Research Assistant in Laboratory for Intelligent Autonomous Systems (LIAS) at School of Data Science Shenzhen, China
 - Advisor: [Prof. Zhang Ruimao](#)

PATENTS AND PUBLICATIONS

J=Journal, P=Patent, S=Software Copyright, R=Under Review

* Indicates Corresponding Author, † Indicates equal contribution

- [J.1] **Zhongyu Yang**, Ziyue Xue [Analysis and Forecast of GDP of Gansu Province based on ARIMA Model](#). *Chinese Market* (IF=0.6)
- [J.2] Mengying Su, **Zhongyu Yang***, Shujaat Abbas, et al [Toward Enhancing Environment Quality in OECD Countries: Role of Municipal Waste, Renewable Energy, Environment Innovation and Environmental Policy](#). *Renewable energy* (SCI Q1 Top, IF=9.0)
- [J.3] Zhichao Yu, Wenlan Xie, Junjie Guo, **Zhongyu Yang*** [Green Effect of Energy Transition Policy: A quasi-natural Experiment Based on New Energy Demonstration Cities](#). *Finance Research Letters* (SSCI Q1 Top, IF=10.4)
- [P.1] **Zhongyu Yang**. A mathematics teaching system based on virtual reality. (CN116312091A)
- [S.1] **Zhongyu Yang**. Green and Low-carbon Integrated Monitoring Software. (2023SR1355487)
- [S.2] **Zhongyu Yang**. Fully automatic spatial sound field environment perception system. (2024SR0538446)
- [R.1] **Zhongyu Yang**, Zuhao Yang, Yingfang Yuan, et al. [ReChar: Revitalising Characters with Structure-Preserved and User-Specified Aesthetic Enhancements](#). Under review in *Siggraph Asia* 2025. (CVPR 2025 443)
- [R.2] **Zhongyu Yang**[†], Jun Chen[†], Dannong Xu, et al. [WikiAutoGen: Towards Multi-Modal Wikipedia-Style Article Generation](#). Under review in ICCV 2025. [Hugging Face Daily Selection](#).
- [R.3] Dannong Xu[†], **Zhongyu Yang**[†], Jun Chen, et al. [MultiHaystack: Benchmarking Multimodal Reasoning over 2K Images, Videos, and Documents](#). Under review in *NeurIPS* 2025.
- [R.4] Zhengwei Zou[†], **Zhongyu Yang**[†], Xuanming Jiang[†], et al. [EmoRes: Toward User-Agnostic Psychological Support via Topic-Mining Emotional Agent](#). Under review in *EMNLP* 2025.
- [R.5] **Zhongyu Yang**, Junhao Song, Yingfang Yuan, et al. [MERMAID: Multi-perspective Self-reflective Agents with Generative Augmentation for Emotion Recognition](#). Under review in *EMNLP* 2025.

PROJECTS

- **Enhancing Multimodal Model Understanding and Generation** Dec 2024 - Present
Advisor: [KAUST Vision-CAIR Team](#)
 - **Objective:** To enhance the factual grounding and multimodal reasoning of MLLMs via web-scale knowledge retrieval and integration into vision-language inference pipelines.
 - **Methodology:** Developed a multi-agent retrieval strategy for dynamic knowledge injection, enabling context-aware visual grounding and enhanced multimodal question answering.

- Diffusion Model for Reconstructing Chinese Characters via Content-Style Disentanglement**
May 2023 - Sep 2024
 Advisor: *Dr. James Yuan, Heriot-Watt University, UK*
 - Objective:** To design a structure-aware diffusion model for personalized character generation with controllable stylistic prompts and semantic alignment.
 - Methodology:** Introduced cross-attention injection to disentangle content and style in SDXL-based pipelines, preserving glyph topology while enabling flexible multimodal prompt control.
- UNet-Centric MambaMorph: A Comprehensive Visual Mamba Framework Enhanced with Cross-Scan and Semi-Supervised Learning for Medical Segmentation**
Jan. 2024 - Present
 Fundamental Research Funds for Central Universities Research Capacity Improvement Project(Supervisor: Prof.Zhang Wenting)
 - Purpose:** To enhance visual understanding of biomedical images in multimodal settings through improved long-range context modeling.
 - Methods:** Designed a UNet-Mamba hybrid with a novel Cross-Scan module to boost segmentation in weak-label regimes, supporting visual-textual diagnostic pairing.
- FPGA-Based AI Doctor: Deep Learning-Based Clinical Target Delineation for Cervical Cancer**
Mar. 2023 - April. 2024
 National College Student Innovation and Entrepreneurship Training Program(Supervisor: Prof.Wang XingHua)
 - Purpose:** To enable real-time visual perception and decision support for medical diagnosis systems using multi-sensor fusion and vision models.
 - Methods:** Refined U-Net-based architecture with parallel FPGA acceleration and integrated attention modules for enhanced feature extraction and multimodal interpretability.
- Tropical Linear Representation ofInvolute Chinese Monoids**
Mar. 2023 - May. 2024
 National College Student Innovation and Entrepreneurship Training Program(Supervisor: Prof.Zhang Wenting)
 - Purpose:** To introduce and define the tropical linear representation within Chinese monoids of involution.
 - Methods:** The approach encompasses the theoretical establishment of free monoids and rewriting systems, followed by the definition of their tropical linear representations for involution in Chinese monoids.

HONORS AND AWARDS

- Best Wiki Winner**, International Directed Evolution Competition (IDEC) (2024) Top 5%
- Silver Medal**, International Genetically Engineered Machine Competition (iGEM) (2024) Top 15%
- Meritorious Winner**, ICM Mathematical Modeling Competition (2023) Top 6%
- Honorable Mention**, MCM (2023), focus on multimodal data modeling Top 25%
- Best Hardware & Target Design**, IDEC (2023), integrating ML with real-world application Top 1%
- 1st Prize**, National Data Analysis Competition (2022), applied to visual data regression Top 3%
- 1st Prize**, China Big Data Challenge (2022) Top 8%
- Outstanding Student Scholarship**, Lanzhou University (2022, 2024) Top 15%

EXPERIENCE

- SenseTime Research**
Shenzhen, China
 Research Intern, General Perceptual Computing Group
 Feb. 2025 – Present
 - Research on streaming MLLMs, focusing on hallucination mitigation and consistent vision-text alignment.
- Heriot-Watt University**
Edinburgh, UK
 Remote Research Intern in School of Mathematical and Computer Sciences
 March. 2024 – Present
 - Research on diffusion-based image generation and multi-agent systems for task-specialized visual synthesis.
- KAUST**
Saudi Arabia
 Remote Research Intern, Vision-CAIR Group
 Dec. 2024 – Present
 - Exploring multi-agent coordination for web-scale retrieval to augment the reasoning capability of MLLMs.
- CUHK Shenzhen**
Shenzhen, China
 Research Assistant
 Apr. 2024 – Nov. 2024
 - Developed MLLMs’ alignment between visual and textual cues to enable human-centric multimodal generation.
- iFLYTEK Co., Ltd.**
Lanzhou, China
 Data Analysis Assistant Intern in Smart Home Department
 June 2023 - Aug. 2023
 - Developed multimodal recommendation models from multi-source behavior-aware cues.

SKILLS AND SERVICES

- Programming Languages:** Python, R, Stata, Latex
- Languages:** Mandarin(Native), Cantonese(Native), English(Fluent)
- Operation System:** Windows (advanced), Linux(advanced)
- Reviewer Services:** CVPR 2025, ICCV 2025, ICLR 2025; EMFT (Q1), ESPR (Q1), IJER (Q2)