Haonan Zhang

PH.D. STUDENT IN COMPUTER SCIENCE · UESTC

✉ zchiowal@gmail.com | 🏠 Homepage | 🐙 Haonan | 🎓 Google Scholar

Research Interest

My research interests focus on Multi-modal Learning, LLMs. I am currently exploring the exciting and fast-growing field of Vision-Language-Action (VLA) models as part of my recent research.

Education Backgrounds

University of Electronic Science and Technology of China   *Chengdu*

• PH.D. STUDENT IN COMPUTER SCIENCE AND TECHNOLOGY, Supervisors: Prof. Lianli Gao   *Sep. 2022 - present*

University of Electronic Science and Technology of China   *Chengdu*

• MASTER STUDENT IN ELECTRONIC INFORMATION, Supervisors: Prof. Jingkuan Song   *Sep. 2020 - Jun. 2022*

Xidian University   *Xi'an*

• B.ENG. IN COMPUTER SCIENCE AND TECHNOLOGY   *Sep. 2016 - Jun. 2020*

Research Experience

University of Trent. Supervisor, Prof. Nicu Sebe | Visiting Ph.D Student   *May 2025 - present*
Area: *Multimodal LLMs, Vision-Language Action (VLA) Model.*

Alibaba Group, Tongyi Laboratory | Research Intern   *May 2024 - Dec. 2024*
Area: *Multimodal LLMs, Role-playing Agent, Video-Language-Audio Understanding.*

Selected Publications

• **OmniCharacter: Towards Immersive Role-Playing Agents with Seamless Speech-Language Personality Interaction**   *ACL25*

  Haonan Zhang, Run Luo, Xiong Liu, Yuchuan Wu, Ting-En Lin, Pengpeng Zeng, QIANG QU, Feiteng Fang, Min Yang, Lianli Gao, Jingkuan Song, Fei Huang, Yongbin Li

• **MMEvol: Empowering Multimodal Large Language Models with Evol-Instruct**   *ACL25*

  RUN LUO*, **HAONAN ZHANG** (EQUAL CONTRIBUTION)*, LONGZE CHEN*, TING-EN LIN*, XIONG LIU, YUCHUAN WU, MIN YANG, MINZHENG WANG, PENGPENG ZENG, LIANLI GAO, HENG TAO SHEN, YUNSHUI LI, XIAOBO XIA, FEI HUANG, JINGKUAN SONG, YONGBIN LI

• **Text-Video Retrieval with Global-Local Semantic Consistent Learning**   *TIP25*

  HAONAN ZHANG, PENGPENG ZENG, LIANLI GAO, JINGKUAN SONG, YIHANG DUAN, XINYU LYU, HENG TAO SHEN

• **OpenOmni: Large Language Models Pivot Zero-shot Omnimodal Alignment across Language with Real-time Self-Aware Emotional Speech Synthesis**   *arXiv25*

  RUN LUO, TING-EN LIN, **HAONAN ZHANG**, YUCHUAN WU, XIONG LIU, MIN YANG, YONGBIN LI, LONGZE CHEN, JIAMING LI, LEI ZHANG, YANGYI CHEN, HAMID ALINEJAD-ROKNY, FEI HUANG

• **UMP: Unified Modality-aware Prompt Tuning for Text-Video Retrieval**   *TCSVT24*

  HAONAN ZHANG, PENGPENG ZENG, LIANLI GAO, JINGKUAN SONG, HENG TAO SHEN

• **MPT: Multi-grained Prompt Tuning for Text-Video Retrieval**   *ACM MM24*

  HAONAN ZHANG, PENGPENG ZENG, LIANLI GAO, JINGKUAN SONG, HENG TAO SHEN

• **SPT: Spatial pyramid transformer for image captioning**   *TCSVT23*

  HAONAN ZHANG, PENGPENG ZENG, LIANLI GAO, XINYU LYU, JINGKUAN SONG, HENG TAO SHEN

• **Depth-aware sparse transformer for video-language learning**   *ACM MM23*

  HAONAN ZHANG, LIANLI GAO, PENGPENG ZENG, ALAN HANJALIC, HENG TAO SHEN

• **Video Question Answering with Prior Knowledge and Object-sensitive Learning**   *TIP22*

  PENGPENG ZENG, **HAONAN ZHANG**, LIANLI GAO, JINGKUAN SONG, HENG TAO SHEN

• **S^2 Transformer for image captioning**   *IJCAI22*

  PENGPENG ZENG*, **HAONAN ZHANG*** (EQUAL CONTRIBUTION), JINGKUAN SONG, AND LIANLI GAO

Honors & Awards

2025 "Academic Newcomer" Graduate Student Honor Award, UESTC.

2024 Awarded CSC Scholarship, UESTC.

2024 National Scholarship, UESTC.

2024 Shenzhen Stock Exchange Scholarship, UESTC.

2023 First-class Scholarship, UESTC.

2023 Outstanding Graduate Teaching Assistant Award, UESTC.

2022 "Academic Youth" Graduate Student Honor Award, UESTC.

Competitions

2024 The 1st Place Winner in Attribute Recognition track in MMVRAC, ICME 2024.

2024 The 9-th Winner in Challenge of Black-box Adversarial Attacks on Vision Foundation Models in CVPR 2024.

2021 Top-4 on Multi-Modal Video Reasoning and Analyzing Competition, Track-1, ICCV 2021.

Academic Service

Conference: Program Committee of AAAI25, Reviewer for CVPR23,24,25, ICCV25, ICLR25, ECCV24, ACM MM, etc.

Journal: IEEE TIP, IEEE TCSVT, IEEE TMM, etc.

Personal Projects

2023-Present **Awesome-Embodied-Robotics-and-Agent**. A curated list of "Embodied robotics or agent with Vision-Language Models (VLMs) and Large Language Models (LLMs)" research, with **1.4K+ Github stars**.