

Yiyang Huang

huang.yiyan@northeastern.edu | +1 781-873-9395 | LinkedIn

Education

Northeastern University, Boston, US <ul style="list-style-type: none">Ph.D. in Computer Engineering, advised by Prof. Raymond FuResearch Concentration: Multi-Modal LLM, Hallucination, Video LLMGPA: 4.0/4.0	Sep 2024 – present
Xidian University, Xi'an, CN <ul style="list-style-type: none">M.S. in Computer Science, advised by Prof. Xuefeng LiangResearch Concentration: Vision-Language Learning, Transfer LearningThesis: Out-of-Distribution Robustness and Low-Resource Adaptation in Video Analysis	Sep 2021 – Jun 2024
Xidian University, Xi'an, CN <ul style="list-style-type: none">B.Eng. in Intelligence Science and TechnologyThesis: Video Analysis with Contrastive Learning	Sep 2017 – Jun 2021

Experience

Research Intern (On-site) , Adobe Research – San Jose, US <ul style="list-style-type: none">Mentor: Zhaowen WangResponsibilities: Research on Design Understanding with VLM	May 2025 – present
Research Student (On-site) , Kyoto University – Kyoto, JP <ul style="list-style-type: none">Mentor: Prof. Takatsune KumadaResponsibilities: Research on Understanding Social Scene with MLLM	Sep 2023 – Mar 2024

Publications

D-CoDe: Scaling Image-Pretrained VLMs to Video via Dynamic Compression and Question Decomposition <i>Yiyang Huang</i> , Yun Fu	under-review
Addressing Hallucinations in LVLMs via Bias and Vulnerability Mitigation <i>Yiyang Huang</i> , Shi Liang, Yun Fu	under-review
LipReading for Low-resource Languages by Language Dynamic LoRA Shuai Zou, Xuefeng Liang, <i>Yiyang Huang</i> 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [Paper Link]	Mar 2025
CALip: Lipreading using Contrastive and Attribute Learning <i>Yiyang Huang</i> , Xuefeng Liang, Chaowei Fang, 29th ACM International Conference on Multimedia (ACM MM) [Paper Link]	Oct 2021

Research Experience

Training-Free Video Expansion for Image-Based LVLMs <ul style="list-style-type: none">Supervisor: Prof. Raymond FuResearch brief: Identified perception bottlenecks and token overload as key challenges in adapting image-pretrained vision-language models to video tasks. Proposed D-CoDe, a training-free framework that performs dynamic compression of visual input and question decomposition to reduce redundancy and guide	Feb 2025–present
---	------------------

focused, multi-step video understanding.

Addressing Object Hallucination in Large Vision-Language Models

Oct 2024–Jan 2025

- Supervisor: Prof. Raymond Fu
- Research brief: Identified biases and vulnerabilities in the visual encoder of Large Vision-Language Models (LVLMs) that lead to object hallucination. Proposed CLIP-guided Contrastive Decoding (CCD), a training-free method that reallocates attention to visual tokens, integrates noise-derived tokens, and employs adversarial perturbations with contrastive decoding to mitigate hallucinations.

Understanding Social Scene using Visual Cues enhanced Multimodal LLM

Oct 2023–Mar 2024

- Supervisor: Prof. Takatsune Kumada
- Research brief: Enhanced a multimodal LLM with body language cues—including facial expressions, posture, and gestures—to improve comprehension in social scenarios. Combined Chain of Thought (CoT) and cognitive theory to address modal bias in multimodal inputs, enabling more accurate interpretation of characters' emotions and intentions under few-shot in-context learning.

Enhancing Low-Resource Lipreading using Dynamic Language Query

Jan 2023–Jan 2024

- Supervisor: Prof. Xuefeng Liang
- Research brief: Replaced the lipreading model's decoder with a multilingual LLM to enable transferring a pre-trained English lipreading model to low-resource languages. Employed a dynamic language-querying mechanism to adapt query tensors based on inter-language relationships, enhancing transfer learning performance.

Lipreading for Out-of-Distribution Unseen Speaker

Oct 2021–Nov 2022

- Supervisor: Prof. Xuefeng Liang
- Research brief: Proposed treating out-of-distribution (OOD) speaker lipreading as a domain generalization problem by introducing proxy feature distributions to reduce the gap between seen and OOD speakers. Trained a dynamic CNN to learn basic feature distributions from seen speakers and pre-trained a lip auto-encoder to generate combination weights for OOD speakers, enabling the estimation of their proxy feature distributions.

Lipreading using Contrastive and Attribute Learning

Oct 2020–Mar 2021

- Supervisor: Prof. Xuefeng Liang
- Research brief: Proposed an attribute learning module to normalize lip movements across speakers, reducing cross-speaker variation and enabling more consistent visual feature learning. Introduced a contrastive learning module to enhance feature discrimination and mitigate viseme confusion when similar phonemes are spoken.

Technologies

Languages: Python, C++, Matlab, \LaTeX .

Tools and libraries: PyTorch, Numpy, TorchLighting, Skimage, Scikit-learn.

Awards

- Outstanding student, Xidian University, China, 2022.
- National Scholarship, China, 2021.
- Undergraduate Computer Design Competition, 1st prize, China, 2021.
- National robotic competition (RoboMaster), 2nd prize, China, 2019.
- ICRA 2019 AI challenge, 3rd prize, 2019.