Yiyang Huang

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

Education

Northeastern University, Boston, US   Sep 2024 – present
- Ph.D. in Computer Engineering, advised by Prof. Raymond Fu
- Research Concentration: Multi-Modal LLM, Hallucination, Video LLM

Xidian University, Xi'an, CN   Sep 2021 – Jun 2024
- M.S. in Computer Science, advised by Prof. Xuefeng Liang
- Research Concentration: Vision-Language Learning, Transfer Learning
- Thesis: Out-of-Distribution Robustness and Low-Resource Adaptation in Video Analysis

Xidian University, Xi'an, CN   Sep 2017 – Jun 2021
- B.Eng. in Intelligence Science and Technology
- Thesis: Video Analysis with Contrastive Learning

Experience

Research Intern (On-site), Adobe Research – San Jose, US   May 2025 – present
- Mentor: Zhaowen Wang
- Responsibilities: Research on Design Compositional Reasoning with VLM

Research Student (On-site), Kyoto University – Kyoto, JP   Sep 2023 – Mar 2024
- Mentor: Prof. Takatsune Kumada
- Responsibilities: Research on Understanding Social Scene with MLLM

Publications

SHIELD: Suppressing Hallucinations In LVLM Encoders via Bias and Vulnerability Defense   under-review
Yiyang Huang, Shi Liang, Yitian Zhang, Yi Xu, Yun Fu
[Paper]

D-CoDe: Scaling Image-Pretrained VLMs to Video via Dynamic Compression and Question Decomposition   2025
Yiyang Huang, Yizhou Wang, Yun Fu
The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)
[Paper] [Code]

LipReading for Low-resource Languages by Language Dynamic LoRA   2025
Shuai Zou, Xuefeng Liang, *Yiyang Huang*
2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)
[Paper]

CALLip: Lipreading using Contrastive and Attribute Learning   2021
Yiyang Huang, Xuefeng Liang, Chaowei Fang,
29th ACM International Conference on Multimedia (ACM MM)
[Paper]

Research Experience

Layout Composition Reasoning in VLMs   May 2025–present
- Mentor: Zhaowen Wang
- Research brief: Identified the failure to model multi-layer, overlapping compositional relationships in design as the core bottleneck for current VLMs in advanced design understanding tasks. We first construct a specialized dataset by fusing expert rules with automated mining techniques (annotated with the aid of GPT4o), and then

train the VLM using a two-stage "SFT-inject, RL-alig" strategy to systematically endow the model with powerful, multi-level design compositional reasoning capabilities.

Training-Free Video Expansion for Image-Based LVLMs — Feb 2025–May 2025

- Supervisor: Prof. Raymond Fu
- Research 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 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.