

# Yiyang Huang

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

## Education

<b>Northeastern University, Boston, US</b>	Sep 2024 – present
• Ph.D. in Computer Engineering, advised by Prof. Raymond Fu	
• Research Concentration: MLLM, VLM Hallucination, Video Understanding	
<b>Xidian University, Xi'an, CN</b>	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	
<b>Xidian University, Xi'an, CN</b>	Sep 2017 – Jun 2021
• B.Eng. in Intelligence Science and Technology	
• Thesis: Video Analysis with Contrastive Learning	

## Experience

<b>Research Intern (On-site)</b> , Adobe Research – San Jose, US	May 2025 – Nov 2025
• Mentor: Zhaowen Wang; Simon Jenni; Jing Shi;	
• Responsibilities: Research on Design Compositional Understanding with VLM	
<b>Research Student (On-site)</b> , Kyoto University – Kyoto, JP	Sep 2023 – Mar 2024
• Mentor: Prof. Takatsune Kumada	
• Responsibilities: Research on Understanding Social Scene with MLLM	

## Publications

<b>MASON: Compositional Design Layout Understanding in VLMs through Multimodal Alignment and Structural Perception</b>	CVPR under-review
<i>Yiyang Huang, Zhaowen Wang, Simon Jenni, Jing Shi, Yun Fu</i>	
<b>SHIELD: Suppressing Hallucinations In ILM Encoders via Bias and Vulnerability Defense</b>	ICLR under-review
<i>Yiyang Huang, Shi Liang, Yitian Zhang, Yi Xu, Yun Fu</i>	
<i>[Paper]</i>	
<b>D-CoDe: Scaling Image-Pretrained VLMs to Video via Dynamic Compression and Question Decomposition</b>	2025
<i>Yiyang Huang, Yizhou Wang, Yun Fu</i>	
The 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)	
<i>[Paper] [Code]</i>	
<b>LipReading for Low-resource Languages by Language Dynamic LoRA</b>	2025
<i>Shuai Zou, Xuefeng Liang, Yiyang Huang</i>	
2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)	
<i>[Paper]</i>	
<b>CALLip: Lipreading using Contrastive and Attribute Learning</b>	2021
<i>Yiyang Huang, Xuefeng Liang, Chaowei Fang,</i>	
29th ACM International Conference on Multimedia (ACM MM)	
<i>[Paper]</i>	

## Research Experience

---

<b>Compositional Design Layout Understanding in VLMS</b>	May 2025–present
<ul style="list-style-type: none"><li>Mentor: Zhaowen Wang, Simon Jenni, Jing Shi</li><li>Research brief: Developed MASON, a multimodal framework for compositional layout understanding, and built CoDeLayout, the first dataset for multilayered design layouts. The method mitigates semantic drift and structural ambiguity, achieving superior hierarchical layout comprehension with only 30% of training data.</li></ul>	
<b>Training-Free Video Expansion for Image-Based LVLMs</b>	Feb 2025–May 2025
<ul style="list-style-type: none"><li>Supervisor: Prof. Raymond Fu</li><li>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.</li></ul>	
<b>Addressing Object Hallucination in Large Vision-Language Models</b>	Oct 2024–Jan 2025
<ul style="list-style-type: none"><li>Supervisor: Prof. Raymond Fu</li><li>Research brief: Proposed SHIELD, a training-free framework that mitigates object hallucination in Large Vision-Language Models by addressing statistical bias, inherent bias, and vulnerability in visual encoders. The method improves robustness across diverse LViLM families and benchmarks while maintaining strong general-task performance.</li></ul>	
<b>Understanding Social Scene using Visual Cues enhanced Multimodal LLM</b>	Oct 2023–Mar 2024
<ul style="list-style-type: none"><li>Supervisor: Prof. Takatsune Kumada</li><li>Research brief: Integrated body-language cues into a multimodal LLM to improve social-scenario comprehension, using CoT and cognitive-theory-based methods to reduce modal bias and boost few-shot emotion/intent recognition.</li></ul>	
<b>Enhancing Low-Resource Lipreading using Dynamic Language Query</b>	Jan 2023–Jan 2024
<ul style="list-style-type: none"><li>Supervisor: Prof. Xuefeng Liang</li><li>Research brief: Enabled low-resource lipreading transfer by replacing the decoder with a multilingual LLM and adding dynamic language-querying for better cross-language adaptation.</li></ul>	
<b>Lipreading for Out-of-Distribution Unseen Speaker</b>	Oct 2021–Nov 2022
<ul style="list-style-type: none"><li>Supervisor: Prof. Xuefeng Liang</li><li>Research brief: Improved OOD-speaker lipreading by modeling it as domain generalization and generating proxy feature distributions via a dynamic CNN and lip auto-encoder.</li></ul>	
<b>Lipreading using Contrastive and Attribute Learning</b>	Oct 2020–Mar 2021
<ul style="list-style-type: none"><li>Supervisor: Prof. Xuefeng Liang</li><li>Research brief: Proposed an attribute-learning module to normalize cross-speaker lip movements and improve consistency in visual features. Added a contrastive learning module to enhance feature discrimination and reduce viseme confusion between similar phonemes.</li></ul>	

## Technologies

---

**Languages:** Python, C++, Matlab, L<sup>A</sup>T<sub>E</sub>X.

**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.