

# Yiyang Huang

◊ [huang.yiyan@northeastern.edu](mailto:huang.yiyan@northeastern.edu) ◊ [Personal Page](#) ◊ [Google Scholar](#) ◊ [LinkedIn](#)

## EDUCATION

Sep 2024 – Present	<b>Northeastern University, USA</b> Ph.D. in Computer Engineering, advised by <a href="#">Prof. Yun Raymond Fu</a>
Sep 2021 – Jun 2024	<b>Xidian University, China</b> M.S. in Computer Science, advised by <a href="#">Prof. Xuefeng Liang</a>
Sep 2017 – Jun 2021	<b>Xidian University, China</b> B.Eng. in Intelligence Science and Technology

## RESEARCH INTERESTS

**Multimodal LLMs; Efficiency; Reliability; Hallucination Detection & Mitigation; Video Understanding; Layout Understanding**

## EXPERIENCE

<b>Northeastern University</b> SMILE Lab, Research Assistant   Supervisor: <a href="#">Prof. Yun Raymond Fu</a>	Sep 2024 – Present Boston, USA
<ul style="list-style-type: none"><li>Published <b>1 ICLR 2026 paper</b> on mitigating hallucinations in MLLMs via encoder-level inference intervention.</li><li>Identified encoder-side causes of hallucinations and proposed SHIELD, a training-free module (token reweighting/subtraction + adversarial contrastive decoding), reducing hallucinations by 27% on captioning across MLLM families.</li><li>Published <b>1 EMNLP 2025 (Main) paper</b> on scaling image-pretrained MLLMs to video understanding.</li><li>Developed D-CoDe, a plug-and-play framework that scales image-pretrained MLLMs to VideoQA via question decomposition and dynamic token/frame compression, improving first-person video understanding accuracy by 29% under LLaVA-NeXT.</li><li>Authored a comprehensive survey on Video-LLM hallucinations and maintain a curated GitHub repository tracking related papers and benchmarks. <a href="#">[GitHub]</a></li></ul>	
<b>Adobe Research</b> Vision-Language Lab, Research Intern   Mentors: <a href="#">Zhaowen Wang</a> , <a href="#">Simon Jenni</a> , <a href="#">Jing Shi</a>	May 2025 – Nov 2025 San Jose, USA
<ul style="list-style-type: none"><li>Submitted <b>1 CVPR paper</b> and filed <b>1 patent</b> on compositional layout understanding in MLLMs.</li><li>Identified semantic drift and structural ambiguity as key challenges in compositional layout reasoning, and developed MASON with metadata-visual grounding and 3D structural perception, improving compositional matching by 26% over Gemini-2.5-Pro and 15% over GPT-5/GPT-o3 under a Qwen2.5-VL backbone.</li><li>Built CoDeLayout (~20K samples) through an automated pipeline for compositional layout mining and QA generation.</li></ul>	
<b>Xidian University</b> Research Assistant   Supervisor: <a href="#">Prof. Xuefeng Liang</a>	Sep 2021 – Jun 2024 Xi'an, China
<ul style="list-style-type: none"><li>Published <b>1 ICASSP 2025 paper</b> on visual speech recognition in low-resource language settings.</li><li>Developed a dynamic LoRA framework to learn meta lip representations shared across languages, improving accuracy by 15% on low-resource languages (e.g., Portuguese and Italian) using pretrained 3DCNN backbones and the multilingual LLM BLOOMZ.</li><li>Published <b>1 ACM MM 2021 paper</b> on contrastive and attribute learning for visual speech recognition.</li><li>Proposed CALLip, integrating attribute learning with audio-visual contrastive learning to normalize cross-speaker lip shape variations, reducing word error rate by 50% on multilingual benchmarks.</li></ul>	
<b>Kyoto University</b> Research Student   Mentor: <a href="#">Prof. Takatsune Kumada</a>	Sep 2023 – Mar 2024 Kyoto, Japan
<ul style="list-style-type: none"><li>Developed a social-scene MLLM framework integrating body language cues via multiple visual branches, and applied CoT reasoning guided by cognitive theory to reduce modal bias and hallucinations in social interaction understanding.</li></ul>	

## PUBLICATIONS

### Accepted

- Yiyang Huang**, Liang Shi, Yitian Zhang, Yi Xu, Yun Fu. “SHIELD: Suppressing Hallucinations in LVLM Encoders via Bias and Vulnerability Defense.” ICLR 2026. [\[Paper\]](#) [\[Code\]](#)
- Yiyang Huang**, Yizhou Wang, Yun Fu. “D-CoDe: Scaling Image-Pretrained VLMs to Video via Dynamic Compression and Question Decomposition.” EMNLP 2025 (Main). [\[Paper\]](#) [\[Code\]](#)
- Shuai Zou, Xuefeng Liang, **Yiyang Huang**. “LipReading for Low-resource Languages by Language Dynamic LoRA.” ICASSP 2025. [\[Paper\]](#)
- Yiyang Huang**, Xuefeng Liang, Chaowei Fang. “CALLip: Lipreading using Contrastive and Attribute Learning.” ACM MM

2021. [Paper]

#### Submitted / Under Review

- **Yiyang Huang**, Yitian Zhang, Yizhou Wang, Mingyuan Zhang, Liang Shi, Huimin Zeng, Yun Fu. “Distorted or Fabricated? A Survey on Hallucination in Video LLMs.” ARR under-review.
- **Yiyang Huang**, Zhaowen Wang, Simon Jenni, Jing Shi, Yun Fu. “MASON: Compositional Design Layout Understanding in VLMs through Multimodal Alignment and Structural Perception.” CVPR under-review.
- Mingyuan Zhang, Yue Bai, Yifan Wang, **Yiyang Huang**, Yun Fu. “Rethinking Fine-Tuning: Unlocking Hidden Capabilities in Vision-Language Models.” CVPR under-review. [Paper]
- Liang Shi, **Yiyang Huang**, Yun Fu. “Capturing Individual Differences of Facial Expression for Authentic Expression Generation.” FG under-review.

#### ACADEMIC SERVICE

---

**Conference Reviewing:** ACL Rolling Review (ARR); IEEE Face & Gesture (FG)

**Journal Reviewing:** ACM TKDD

#### TEACHING

---

##### Teaching Assistant

DS 5110: Essentials of Data Science

Fall 2025

DS 5020: Fundamentals of Linear Algebra and Probability

Spring 2026

#### SKILLS

---

**Languages:** Python, L<sup>A</sup>T<sub>E</sub>X

**Frameworks/Tools:** PyTorch, HF Transformers, HF Datasets, ffmpeg, PyAV, decord, Git, Slurm

**Training/Finetuning:** Accelerate, PEFT, LoRA/QLoRA, bitsandbytes

**Inference:** vLLM, FlashAttention

**Evaluation:** lmms-eval, POPE, CHAIR, MME

#### AWARDS

---

- National Scholarship, China 2021
- Outstanding Student, Xidian University 2022
- 1st Prize, Undergraduate Computer Design Competition (National Level), China 2021
- 2nd Prize, RoboMaster National Robotics Competition, China 2019
- 3rd Prize, ICRA AI Challenge 2019