Qidong Huang

Ph.D, University of Science and Technology of China

Building No.7, USTC West Campus
Hefei, Anhui, China
☎ *(+86) 13085060686*
✉ *hqd0037@mail.ustc.edu.cn*
https://shikiw.github.io/

Short Biography

I am currently a final-year PhD student at University of Science and Technology of China. I have published more than 10 papers at top-tier conferences and journals, such as CVPR/ICCV/TIP. My research interests focus on multi-modal LLMs and trustworthy/efficient AI, including scalable MLLMs, efficient training/inference, AI privacy and robustness. I serve as the reviewer of conferences (e.g., CVPR, ICML, NeurIPS) and journals (e.g., TPAMI). I am working closely with Dongdong Chen, Xiaoyi Dong, Jiaqi Wang, and Gang Hua.

Education

09/2020–present **PhD of Cyberspace Technology**, *University of Science and Technology of China*, Hefei, China, CAS Key Laboratory of Electromagnetic Space Information. Supervised by Prof. Weiming Zhang, Prof. Nenghai Yu.

09/2016–06/2020 **Bachelor of Electronic Information**, *University of Science and Technology of China*, Hefei, China, Supervised by Prof. Weiming Zhang.

Experience

08/2023–present **Research Intern**, *Shanghai AI Laboratory*, Shanghai, China.
Member of InternLM-XComposer Group, supervised by Xiaoyi Dong, Jiaqi Wang. Research in multi-modal LLMs, especially in inference-time scaling for image/video caption scaling, cross-modal alignment, efficient training/inference, and multi-modal hallucination.

05/2022–07/2022 **Research Intern**, *iFlyTek Research*, Hefei, China.
Member of Avatar strip, supervised by Shan He. Research in Chinese text-to-image model based on conditional diffusion, focusing on large-scale image-text datasets such as Wukong.

Skills

⋆ **Expertise in multi-modal LLMs :** My recent researches mainly focus on multi-modal LLMs, including scalable image/video captioning, cross-modal alignment, efficient training/inference, and multi-modal hallucination. On these topics, I have published four papers.

— **1) Image/Video Caption Scaling :** I am currently leading a project based on multi-modal inference-time scaling to build more complete and detailed image/video captions.

— **2) MIR&MoCa :** We propose an effective and reliable metric named MIR for quantifying MLLM pre-training, and a light-weight modality calibration module MoCa to facilitate cross-modal alignment.

— **3) MMRC :** We propose a multi-modal conversation benchmark MMRC for evaluating open-ended abilities of MLLMs.

— **4) PyramidDrop :** We propose an efficient training/inference framework for MLLMs through vision redundancy reduction, especially working for high-resolution MLLMs and video LLMs, achieving ∼50% acceleration for models like LLaVA series and Video-LLMs.

— **5) OPERA :** We delve into the underlying causes of multi-modal hallucinations and give an explanation based on information attenuation. Based on this, we propose a training-free decoding algorithm to mitigate the hallucination issue. This work has earned over 50,000 reads and 4,000 shares on social media, with nearly 140 citations within a year.

- **Expertise in efficient AI :** Except for the aforementioned PyramidDrop for efficient MLLM training/inference, I have been researching the parameter-efficient fine-tuning for vision pre-trained models and published one paper on CVPR 2023. This paper proposes DAM-VP, a data diversity-aware method for efficient and adaptive vision prompt learning. This work addresses the mismatch issue between vision prompts and downstream data diversity.
- **Expertise in trustworthy AI :** I am currently curious about LLM safety/security and I have a work regarding jailbreak detection under review. Additionally, I have been researching the trustworthy issue for supervised/unsupervised vision models, where I published four paper (first author) on top-tier computer vision conferences. One is RobustMAE, which reveals the flaw of masked-autoencoder-style vision pre-training on adversarial robustness, and improve it with test-time frequency-domain prompting. Moreover, I have dedicated the early time of my PhD career to other topics, such as : 1) adversarial attack/defense methods for 3D models (e.g., SI-Adv and PointCAT, CVPR 2022 and TIP 2024) ; 2) backdoor attack for 2D models (e.g., Poison Ink, TIP 2022) ; 3) AIGC content safety for text-to-image diffusion models (e.g, SimAC, CVPR 2024) ; and 4) anti-DeepFake (e.g., AAAI 2021, where we are the first to propose the concept of "initiative defense" against DeepFakes by actively protecting users' facial privacy before manipulation, differing from previous post-hoc measures like DeepFake detection.)

Publications (First Author)

- **Qidong Huang**, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Yuhang Cao, Jiaqi Wang, Dahua Lin, Weiming Zhang, Nenghai Yu. Deciphering Cross-Modal Alignment in Large Vision-Language Models with Modality Integration Rate. Arxiv preprint 2410.07167 (**Under Review**), 2024.
- **Qidong Huang**, Xiaoyi Dong, Pan Zhang, Bin Wang, Conghui He, Jiaqi Wang, Dahua Lin, Weiming Zhang, Nenghai Yu. OPERA : Alleviating Hallucination in Multi-Modal Large Language Models via Over-Trust Penalty and Retrospection-Allocation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2024. (Highlight, 2.8% of submissions)*
- **Qidong Huang**, Xiaoyi Dong, Dongdong Chen, Hang Zhou, Weiming Zhang, Kui Zhang, Gang Hua, Nenghai Yu. PointCAT : Contrastive Adversarial Training for Robust Point Cloud Recognition. *IEEE Transactions on Image Processing (**TIP**), 2024.*
- **Qidong Huang**, Xiaoyi Dong, Dongdong Chen, Yinpeng Chen, Lu Yuan, Gang Hua, Weiming Zhang, Nenghai Yu. Improving Adversarial Robustness of Masked Autoencoders via Test-time Frequency-domain Prompting. *IEEE/CVF International Conference on Computer Vision (**ICCV**), 2023.*
- **Qidong Huang**, Xiaoyi Dong, Dongdong Chen, Weiming Zhang, Feifei Wang, Gang Hua, Nenghai Yu. Diversity-Aware Meta Visual Prompting. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2023.*
- **Qidong Huang**, Xiaoyi Dong, Dongdong Chen, Hang Zhou, Weiming Zhang, Nenghai Yu. Shape-invariant 3D Adversarial Point Clouds. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022.*
- **Qidong Huang***, Jie Zhang*, Wenbo Zhou, Weiming Zhang, Nenghai Yu. Initiative Defense against Facial Manipulation. *AAAI Conference on Artificial Intelligence (**AAAI**), 2021.* (*Qidong Huang and Jie Zhang contribute equally.)

Publications (Collaborate)

- Haochen Xue, Feilong Tang, Ming Hu, Yexin Liu, **Qidong Huang**, Yulong Li, Chengzhi Liu, Zhongxing Xu, Chong Zhang, Chun-Mei Feng, Yutong Xie, Imran Razzak, Zongyuan Ge, Jionglong Su, Junjun He, Yu Qiao. MMRC : A Large-Scale Benchmark for Understanding Multimodal Large Language Model in Real-World Conversation. Arxiv preprint 2502.11903 (**Under Review**), 2025.

- Yujie Zhou, Jiazi Bu, Pengyang Ling, Pan Zhang, Tong Wu, **Qidong Huang**, Jinsong Li, Xiaoyi Dong, Yuhang Zang, Yuhang Cao, Anyi Rao, Jiaqi Wang, Li Niu. Light-A-Video : Training-free Video Relighting via Progressive Light Fusion. Arxiv preprint 2502.08590, 2025. (**Project page : https ://bujiazi.github.io/light-a-video.github.io/**)
- Long Xing, **Qidong Huang**, Xiaoyi Dong, Jiajie Lu, Pan Zhang, Yuhang Zang, Yuhang Cao, Conghui He, Jiaqi Wang, Feng Wu, Dahua Lin. PyramidDrop : Accelerating Your Large Vision-Language Models via Pyramid Visual Redundancy Reduction. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2025.*
- Likai Liang, **Qidong Huang**, Weiming Zhang, Wenying Zhang. RDPI : Defending against Multi-Turn Jailbreak Attacks via Response-Based Dynamic Prompt Inference. (**Under Review**), 2024.
- Feifei Wang, Zhentao Tan, Tianyi Wei, Yue Wu, **Qidong Huang**†. SimAC : A Simple Anti-Customization Method against Text-to-Image Synthesis of Diffusion Models. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (**CVPR**), 2024. († Corresponding author)*
- Kui Zhang, Hang Zhou, Jie Zhang, **Qidong Huang**, Weiming Zhang, Nenghai Yu. Ada3Diff : Defending against 3D Adversarial Point Clouds via Adaptive Diffusion. *ACM International Conference on Multimedia (**MM**), 2023*
- Han Fang, Dongdong Chen, **Qidong Huang**, Jie Zhang, Zehua Ma, Weiming Zhang and Nenghai Yu. Deep Template-based Watermarking. *IEEE Transactions on Circuits and Systems for Video Technology (**TCSVT**), 2020.*
- Jie Zhang, Dongdong Chen, **Qidong Huang**, Jing Liao, Weiming Zhang, Huamin Feng, Gang Hua, Nenghai Yu. Poison ink : Robust and invisible backdoor attack. *IEEE Transactions on Image Processing (**TIP**), 2022.*

Services

- Reviewer for CVPR 2022-2025
- Reviewer for ICCV 2023-2025
- Reviewer for ECCV 2022-2024
- Reviewer for ACL 2025
- Reviewer for ICML 2025
- Reviewer for ICLR 2024
- Reviewer for NeurIPS 2024-2025
- Reviewer for IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- Reviewer for IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- Reviewer for IEEE Transactions on Image Processing (TIP)
- Reviewer for Pattern Recognition (PR)

Talk

2025 Toward Efficient & Effective Multi-Modal LLMs. Shanghai Innovation Institute.

2024 Exploring MLLM's Hallucination from A Causal Attention Perspective. AI SPOT, OpenMMLab.

Awards & Honors

2024 China National Scholarship

2021 China National Scholarship

2023 "Internet +" Innovation and Entrepreneurship Competition, Provincial Bronze Award

2023 Anheng Information Scholarship

2015 National High School Mathematics League Provincial First Prize

2014 National High School Mathematics League Provincial First Prize