WENBO (GORDON) HU

Los Angeles, CA   whu@cs.ucla.edu   [Homepage](#)   [Google Scholar](#)   [Github](#)

SUMMARY

My research primarily focused on multimodality, with several **first-authored** publications. These include advancements in large vision-language model techniques, such as MQT-LLaVA (**NeurIPS 2024**) and BLIVA (**AAAI 2024**), as well as evaluations of multimodal models' capabilities through MRAG-Bench (**ICLR 2025**) and VALOR-EVAL (**ACL 2024 Finding**). My current project works on 3D embodied long-term memory reasoning.

EDUCATION

University of California, Los Angeles   *Sept.2023 - Present*
Ph.D. Student in Computer Science   *Sept.2025 - Present*
- Advisor: Kai-Wei Chang and Nanyun Peng
Master of Science in Computer Science   *Sept.2023 - Dec.2024*
University of California, San Diego   *Sept.2019 - March.2023*
Bachelor of Science in Data Science
- Advisor: Zhuowen Tu

PUBLICATION

1. **Wenbo Hu**, Jia-Chen Gu, Zi-Yi Dou, Mohsen Fayyaz, Pan Lu, Kai-Wei Chang, Nanyun Peng, *"MRAG-Bench: Vision-Centric Evaluation for Retrieval-Augmented Multimodal Models"*, **ICLR 2025** [link](#)

2. **Wenbo Hu**, Zi-Yi Dou, Liunian Harold Li, Amita Kamath, Nanyun Peng, Kai-Wei Chang, *"Matryoshka Query Transformer for Large Vision-Language Models"*, **NeurIPS 2024** [link](#)

3. **Wenbo Hu***, Yifan Xu*, Yi Li, Weiyue Li, Zeyuan Chen, Zhuowen Tu, *"BLIVA: A Simple Multimodal LLM for Better Handling of Text-Rich Visual Questions"*, **AAAI 2024** [link](#) (100+ citations)

4. Haoyi Qiu*, **Wenbo Hu*** (Equal Contribution), Zi-Yi Dou, Nanyun Peng, *"VALOR-EVAL: Holistic Coverage and Faithfulness Evaluation of Large Vision-Language Models"*, **ACL 2024 Findings** [link](#)

5. Cheng-Fu Yang*, Da Yin*, **Wenbo Hu**, Heng Ji, Nanyun Peng, Bolei Zhou, Kai-Wei Chang. *"Verbalized Representation Learning for Interpretable Few-Shot Generalization."*, **Under Review of ICCV 2025** [link](#)

RESEARCH

Graduate Researcher at UCLA NLP mentored by Nanyun Peng and Kai-Wei Chang   *Sept.2023 - Present*
- Designed a novel model that enabled a flexible choice in number of visual tokens of representing an image to suit different tasks and various computational resources. (Paper accepted by NeurIPS 2024).
- This model matches LLaVA-1.5 performance across **11** benchmarks with only **half** number of LLaVA's visual tokens. Achieves **3x** TFLOPs speed-up on MMBench with even better performance, or **8x** speed-up sacrificing only 2.4 accuracy.

Research Intern at mlPC mentored by Zhuowen Tu at UCSD   *May.2023 - Sept.2023*
- Led research of a multimodal LLM for better handling of text-rich visual questions. (Paper accepted by AAAI 2024).
- Developed a model, BLIVA, which significantly enhances performance in **6** text-rich VQA benchmarks (up to **17.76%** in OCR-VQA task) and in undertaking **8** typical VQA benchmarks (up to **7.9%** in Visual Spatial Reasoning task).

Undergraduate Researcher at Hao Su Lab at UCSD   *Feb.2022 - December 2022*
- Conducted a follow-up research to minimize simulation to real robot transferability gap. Explored 2D and 3D computer vision models, motion planning and reinforcement learning algorithms. Worked on 6D Pose estimation problems.

WORK EXPERIENCE

Algorithm Engineer Intern at Synthesis Electronic Technology   *June.2021 - Aug.2021*
Software Engineer Intern at Inspur Groups   *July.2020 - Sept.2020*

SKILLS

Proficient in Python, Java, R, SQL, Shell, Javascript, and MATLAB; Utilize Pytorch, Numpy, Sklearn, Scipy, Pandas, AWS, Docker, Kubernetes, PySpark, Dask, OpeanAI Gym and PostgreSQL.