Yidong Huang

+1 (734)-882-7799 | owenhuangji@gmail.com | https://github.com/h6kplus | owenhuang.github.io

Education

University of Michigan Ann Arbor, MI
MS in Computer Science *Sept. 2023 – Apr.2025 (Expected)*

University of Michigan Ann Arbor, MI
B.S in Computer Science and Engineering *Sept. 2021 – Apr.2023*
GPA: 3.92

Shanghai Jiao Tong University Joint Institute (SJTU) Shanghai, China
B.S in Electronic and Computer Engineering *Sept. 2019 – Aug.2023*
GPA: 3.7

Experience

Situated Language and Embodied Dialogue Lab, University of Michigan Feb 2022 – now
Research assistant | Advised by Joyce Chai

- Assisted PhD students with several projects encompassing various aspects of artificial intelligence research, including:
  - Autonomous driving with situated dialogue.
  - Real-robot visual perception and task manipulation.
- Transitioned to a leadership role, currently leading or led several researching projects in collaboration with PhD students.:
  - The autonomous driving project
  - Input controlled diffusion model.
  - A finetuned vision language model that can understand and generate interleaved image and text data.

Boson AI May 2024 – now
Machine Learning Intern | Mentored by Yi Zhu

Projects

Consistent Image editing March 2023 – Nov 2023
Diffusion Model, Image Editing

- Conceived and developed a unique controlled diffusion model, leveraged for image manipulation, which conditions on both input image and text, and utilizes cycle consistency supervision.
- Proposed a new sampling pipeline that omit the inversion step in previous works, thus boosting the inference speed to less that 3 seconds.(10% time of the previous baseline)
- Proposed a new attention control technique that can edit a image consistently, becoming the new SOTA on PIE-bench dataset.
- Co-authored a paper which was accepted by NeurIPS 2023. Co-first authored another paper accepted by CVPR 2024.

DriVLMe: Exploring foundation models as autonomous driving agents that perceive, communicate, and navigate Feb. 2022 – now
Autonomous driving, Large Vision-language Model fintuning, Carla simularot

- Developed a client-server-based simulator platform for embodied AI tasks like navigation, complete with APIs for interaction and tools for data cleaning and annotation.
- Conducted human studies, organized data, and designed an episodic transformer-based model, enhancing the functionality and efficiency of the platform.
- Co-authored a paper on the project that was accepted to findings of EMNLP 2022.
- Led a follow up work on helping autonomous driving agent learn to presieve the envitonment, interact with human user and make plans according to the map topology.
- First authored a paper admitted to VLADR @ CVPR 2024 and also under review at another conference

A-ESRGAN Nov 2021 – Dec. 2021
Super Resolution, Pytorch, GAN, U-net

- Incorporated the multi-scale discriminator with RRDB generator and outperforms SOTA models by 15%.
- Project Github: https://github.com/aesrgan/A-ESRGAN
- Co-authored a paper detailing the project, which was accepted by PRICAI 2023

Publications

[1] Ziqiao Ma, Benjamin VanDerPloeg*, Cristian-Paul Bara*, **Yidong Huang***, Eui-In Kim, Felix Gervits, Matthew Marge, and Joyce Chai. Dorothie: Spoken dialogue for handling unexpected situations in interactive autonomous driving agents. In *Findings of the Association for Computational Linguistics: EMNLP 2022*, pages 4800–4822, 2022.

[2] **Yidong Huang**, Jacob Sansom, Ziqiao Ma, Felix Gervits, and Joyce Chai. DriVLMe: Exploring foundation models as autonomous driving agents that perceive, communicate, and navigate. In *First Vision and Language for Autonomous Driving and Robotics Workshop*, 2024.

[3] Zihao Wei, **Huang, Yidong**, Yuang Chen, Chenhao Zheng, and Jingnan Gao. A-esrgan: Training real-world blind super-resolution with attention u-net discriminators. In *Pacific Rim International Conference on Artificial Intelligence*, pages 16–27. Springer, 2023.

[4] Sihan Xu, Ziqiao Ma, **Yidong Huang**, Honglak Lee, and Joyce Chai. Cyclenet: Rethinking cycle consistency in text-guided diffusion for image manipulation. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023.

[5] Sihan Xu*, **Yidong Huang***, Jiayi Pan, Ziqiao Ma, and Joyce Chai. Inversion-free image editing with natural language. *arXiv preprint arXiv:2312.04965*, 2023.

Technical Skills

Programming Languages: Python, C/C++, Elm, Rust, Ocaml, JavaScript, HTML/CSS, Matlab
Developer Skills: PyTorch, TensorFlow, pygame, git, Linux, carla, React, Flask, Android Studio, Operating System, Web Development, Diffusion Models, LLM finetuning, LoRa finetuning