Weikai Huang

Email: weikaih@cs.washington.edu | Github: weikaih04 | Homepage: https://weikaih04.github.io/

Education

University of Washington, Seattle — Seattle, WA

Bachelor of Science (with Honors), Major in Computer Science, Minor in Applied Math and Stats (Overall GPA: 3.99) — *2023-2027*

Advisor: Prof. Ranjay Krishna (ranjaykrishna.com) and Jieyu Zhang (jieyuz2.github.io)

Related Course: CSE599J: Data-centric Machine Learning (grad), CSE599G1: Deep Learning (grad)

CSE546: Machine Learning (grad), CSE547: Natural Language Processing (grad)

CSE455: Computer Vision, CSE473: Artificial Intelligence

Research Interests

- **Multimodal Language Models, Text-to-Image/Video/3D Models, 2D/3D Segmentation and Detection, Synthetic Data Generation, Model Benchmarking, Embodied/Agentic AI**

Publications (* indicates equally contribution)

- **Generate Any Scene: Evaluating and Improving Text-to-Vision Generation via Programmatic Scene Graphs**
  With Jieyu Zhang, Ranjay Krishna, et al.
  In preparation for CVPR 2025

- **InstructVerse: Scaling Open-Source Visual Instruction Data by 10x Programmatically**
  With Jieyu Zhang, Ranjay Krishna, et al.
  In preparation for CVPR 2025

- **Task Me Anything**
  Jieyu Zhang, **Weikai Huang***, Zixian Ma*, Oscar Michel, Dong He, Tanmay Gupta, Wei-Chiu Ma, Ali Farhadi, Aniruddha Kembhavi, Ranjay Krishna
  NeurIPS 2024

- **m&m's: A Benchmark to Evaluate Tool-Use for Multi-Step Multi-Modal Tasks**
  Zixian Ma, **Weikai Huang**, Jieyu Zhang, Tanmay Gupta, Ranjay Krishna
  ECCV 2024

Research Experience

UW CSE RAIVN Lab & Allen Institute of AI (AI2) — Oct 2023 – Present

Research Assistant (Advised by: Prof. Ranjay Krishna, PhD students Jieyu Zhang and Zixian Ma) — *Seattle, WA*

- ***m&m's (ECCV 2024)***
  - *Creating a multi-step and executable tool-use benchmark for AI agents.*
  - *Implement over 20 tool interfaces or APIs, including image processing, segmentation, captioning tools, web search, location search APIs, and more for AI agents.*
  - *Implement the human annotation pipeline and interface for high-quality data annotation.*
- ***Task Me Anything (NeurIPS 2024)***
  - *Designed a programmatic benchmark generation engine capable of generating over 750 million Visual Question Answering (VQA) questions tailored to user needs for multimodal language models (MLMs).*
  - *Implement unified inference interfaces for over 20 MLMs and run large-scale experiments on over 100 GPUs at the AI2 and UW Hyak clusters.*
  - *Implement 3D images/videos generation and rendering pipeline with Blender.*
- ***Generate Any Scene (In preparation for CVPR 2025)***
  - *Designing programmatic scene graphs & prompts generation engine that can both evaluate and improve text-to-vision models.*
  - *Conducted experiments on over 30 text-to-image, text-to-video, and text-to-3D models across five metrics.*
  - *Finetuning Stable Diffusion 1.5 with LoRA under DreamSync and DreamBooth method*
- ***InstructVerse (In preparation for CVPR 2025)***

- *Building a programmatic pipeline that generates billion-scale visual instruction data to improve multimodal language models' visual reasoning capability from any images.*
- *Fine-tuning LLaVA 1.5 and achieve over 10% of performance increase in CVBench.*
- *Implement program-based instruction generation pipeline and process over millions of images with CV models like SAM and Depth Anything.*

UW CSE Interactive Intelligence (I2) Club Oct 2023 – Dec 2023

Research Lead *Seattle, WA*

- *Leading the Microsoft AutoGen MathGPT Project.*
- *Enhancing GPT-4's ability to solve math problems with Microsoft's multi-agent framework.*
- *Project selected for presentation on the UW CSE I2 Club's official website as a major research initiative.*

AWARD

- **2024 UW CSE John and JoAnne Wisniewski Endowed Scholarship 2024 (2 out of 2000 CS undergrads)** UW CSE
- **2023-2024 UW Annual Dean's List** University of Washington
- **First Prize, National Olympiad in Informatics in Provinces 2021** China Computer Federation (CCF)

SERVICES

- **Research Assistant:** UW RAIVN Lab & Allen Institute of AI (AI2) PI: Prof. Ranjay Krishna
- **Organizer:** Synthetic Data for Computer Vision Workshop @ CVPR 2024 CVPR2024
- **Hoster:** 2024 UW CSE Education Panel UW CSE

SKILLS

- **Languages and Tools**: Python, C++, Java, Docker, Bash, Git, LaTeX, Blender.
- **DL Libraries**: PyTorch, Transformers, Huggingface Trainer, Peft, Accelerator, DeepSpeed, Flash-attn, Bitsandbytes.
- **Techniques**: Distributed training and model evaluation on clusters, Large scale data processing, Data analysis.
- **Languages**: English, Chinese.