Weikai Huang

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

Short Biography

Weikai Huang is a junior undergraduate student at the Paul G. Allen School of Computer Science & Engineering, University of Washington, advised by Prof. Ranjay Krishna. His research lies in computer vision with an emphasis on 2D/3D grounding and tracking, controllable generative models, synthetic data generation for vision foundation models, and spatial/grounding-centric VLMs. His work has been published at top-tier venues including NeurIPS and ECCV, and received multiple oral awards.

Education

- **University of Washington, Seattle** Sep 2023 – March 2026 (Expected)
  B.S. (with Honors) in Computer Science; GPA: 3.93
  Advisor: Prof. Ranjay Krishna

Selected Research (* indicates equal contribution)

- **Synthetic Object Compositions for Scalable and Accurate Learning in Detection, Segmentation, and Grounding**
  Weikai Huang, Jieyu Zhang, Taoyang Jia, Chenhao Zheng, Ziqi Gao, Jae Sung Park, Ranjay Krishna
  In submission
  Contribution: Led the project. Trained 100+ model checkpoints with different configs using MMDetection and Detectron2. Processed 10M+ images using GCP and AI2 clusters.
- **Generate Any Scene: Evaluating and Improving Text-to-Vision Generation with Scene Graph Programming**
  Ziqi Gao*, **Weikai Huang***, Jieyu Zhang, Aniruddha Kembhavi, Ranjay Krishna
  In submission
  Contribution: Designed programmatic scene graph & prompt generation engine for evaluating and improving text-to-vision models. Conducted experiments on 30+ text-to-image, text-to-video, and text-to-3D models. Fine-tuned Stable Diffusion 1.5 with LoRA using DreamSync and DreamBooth methods.
- **Molmo2: Open Weights and Data for Vision-Language Models with Video Understanding and Grounding**
  AI2 Prior Team
  In submission
  Contribution: Multi-image data curation and processing, model fine-tuning, benchmark setup and evaluation.
- **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
  Contribution: Designed programmatic benchmark generation engine capable of generating 750M+ VQA questions for multimodal language models. Implemented unified inference interfaces for 20+ MLMs and ran large-scale experiments on 100+ GPUs. Built 3D image/video generation and rendering pipeline with Blender.
- **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
  Contribution: Implemented 20+ tool interfaces/APIs including image processing, segmentation, captioning, web search, and location search. Built human annotation pipeline and interface for high-quality data annotation.

Experience

- **Research Assistant:** UW CSE RAIVN Lab & Allen Institute for AI (AI2) Oct 2023 – Present

Services

- **Reviewer:** CVPR 2025
- **Organizer:** Synthetic Data for Computer Vision Workshop @ CVPR 2024, 2025
- **Teaching Assistant:** UW CSE 455 Computer Vision (Autumn 2025), UW CSE 493G Deep Learning (Spring 2025, Winter 2026)
- **Host:** 2024 UW CSE Education Panel

AWARD

- **Fellowships**
  - UW CSE John and JoAnne Wisniewski Endowed Scholarship (2 out of 2000 CS undergrads) 2024
- **Academic Honors**
  - UW Annual Dean's List 2023, 2024, 2025
  - First Prize, National Olympiad in Informatics in Provinces 2021

PUBLICATIONS (* INDICATES EQUAL CONTRIBUTION)

- **Synthetic Object Compositions for Scalable and Accurate Learning in Detection, Segmentation, and Grounding**
  Weikai Huang, Jieyu Zhang, Taoyang Jia, Chenhao Zheng, Ziqi Gao, Jae Sung Park, Ranjay Krishna
  In submission
- **Molmo2: Open Weights and Data for Vision-Language Models with Video Understanding and Grounding**
  AI2 Prior Team
  In submission
- **TrajTok: Learning Trajectory Tokens enables better Video Understanding**
  Chenhao Zheng, Jieyu Zhang, Jianing Zhang, **Weikai Huang**, Ashutosh Kumar, Quan Kong, Oncel Tuzel, Chun-Liang Li, Ranjay Krishna
  In submission
- **Synthetic Visual Genome 2: Extracting Large-scale Spatio-Temporal Scene Graphs from Videos**
  Ziqi Gao, Jieyu Zhang, Wisdom Oluchi Ikezogwo, Jae Sung Park, Tario G You, Daniel Ogbu, Chenhao Zheng, **Weikai Huang**, Yinuo Yang, Quan Kong, Rajat Saini, Ranjay Krishna
  In submission
- **Generate Any Scene: Evaluating and Improving Text-to-Vision Generation with Scene Graph Programming**
  Ziqi Gao*, **Weikai Huang***, Jieyu Zhang, Aniruddha Kembhavi, Ranjay Krishna
  In submission
- **ProVision: Programmatically Scaling Vision-centric Instruction Data for Multimodal Language Models**
  Jieyu Zhang, Le Xue, Linxin Song, Jun Wang, **Weikai Huang**, Manli Shu, An Yan, Zixian Ma, Juan Carlos Niebles, Silvio Savarese, Caiming Xiong, Zeyuan Chen, Ranjay Krishna, Ran Xu
  In submission
- **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

SKILLS

- **Languages and Tools**: Python, C++, Java, Docker, Bash, Git, LaTeX
- **DL Libraries**: PyTorch, Transformers, Huggingface Trainer, Peft, Accelerator, DeepSpeed, Flash-attn, Bitsandbytes, vLLM, veRL
- **3D/Vision Libraries**: Blender, Open3D, MeshLab, Detectron2, MMDetection, SAM, Depth Anything
- **Techniques**: Distributed training and model evaluation on clusters, Large scale data processing, Crowd worker data collection
- **Languages**: English, Chinese