Weitai Kang

k13711752197@gmail.com | weitaikang.github.io

RESEARCH INTERESTS

I work on Large Multimodal Model across text, image, GUI, 3D, and video. I specialize in Visual Grounding, Generalist Model, AI Agents, RL and Image Editing.

Education

University of Illinois Chicago — Chicago, IL
PhD student, Computer Science — *Dec. 2024 – May. 2027(expected)*
- Research Assistant, advised by Prof. Yan Yan.

Illinois Institute of Technology — Chicago, IL
PhD student, Computer Science — *Aug. 2022 – Dec. 2024*
- Research Assistant, advised by Prof. Yan Yan.

Sun Yat-sen University — Guangzhou, China
B.S., Mathematics — *Sep. 2017 – July 2022*
- GPA: 3.7/4.0
- Honors: Outstanding student scholarship (for four years)

EMPLOYMENT EXPERIENCE

Netflix — May 2026 – Aug. 2026
Research Intern — *Los Gatos, U.S.*
- Research on Video Understanding.

Adobe — May 2025 – May 2026
Research Intern — *San Jose, U.S., Chicago, U.S.*
- Research on Visual Grounding and Image Editing.

Sony AI — Oct. 2024 – Dec. 2024
Research Intern — *Chicago, U.S.*
- Research on 2D Large Multimodal Model.

Tencent — Oct. 2021 – July 2022
Machine Learning Engineer Intern — *Shenzhen, China*
- Research on Human Pose Detection.

SenseTime — July 2021 – Sep. 2021
Research Intern — *Shenzhen, China*
- Research on Video Super-Resolution.

RESEARCH EXPERIENCE

[1] Weitai Kang, Jason Kuen, Mengwei Ren, Zijun Wei, Yan Yan, Kangning Liu, "VGent: Visual Grounding via Modular Design for Disentangling Reasoning and Prediction", **CVPR 2026**.

[2] Weitai Kang, Bin Lei, Gaowen Liu, Caiwen Ding, Yan Yan, "GuirlVG: Incentivize GUI Visual Grounding via Empirical Exploration on Reinforcement Learning", **ICLR 2026**.

[3] Weitai Kang, Haifeng Huang, Yuzhang Shang, Mubarak Shah, Yan Yan, "Robin3D: Improving 3D Large Language Model via Robust Instruction Tuning", **ICCV 2025**.

[4] Weitai Kang, Mengxue Qu, Jyoti Kini, Yunchao Wei, Mubarak Shah, Yan Yan, "Intent3D: 3D Object Detection in RGB-D Scans based on Human Intention Language", **ICLR 2025**.

[5] Bin Lei*, Weitai Kang*, Zijian Zhang, Winson Chen, Xi Xie, Shan Zuo, Mimi Xie, Ali Payani, Mingyi Hong, Yan Yan, Caiwen Ding, "InfantAgent-Next: A Multimodal Generalist Agent for Automated Computer Interaction". * equal contribution. **NeurIPS 2025**.

[6] Weitai Kang, Luowei Zhou, Junyi Wu, Changchang Sun, Yan Yan, "Visual Grounding with Attention-Driven Constraint Balancing", **ACM MM 2025**.

[7] Weitai Kang, Gaowen Liu, Mubarak Shah, Yan Yan, "SegVG: Transferring Object Bounding Box to Segmentation for Visual Grounding", **ECCV 2024**.

[8] Weitai Kang, Weiming Zhuang, Zhizhong Li, Yan Yan, Lingjuan Lyu, "ExpVG: Investigating the Design Space of Visual Grounding in Multimodal Large Language Model".

[9] Yuzhang Shang, Bingxin Xu, Weitai Kang, Mu Cai, Yuheng Li, Zehao Wen, Zhen Dong, Kurt Keutzer, Yong Jae Lee, Yan Yan, "Interpolating Video-LLMs: Toward Longer-sequence LMMs in a Training-free Manner".

[10] Junyi Wu, Weitai Kang, Hao Tang, Yuan Hong, Yan Yan, "On the Faithfulness of Vision Transformer Explanations", **CVPR 2024**.

[11] Junyi Wu, Bin Duan, Weitai Kang, Hao Tang, Yan Yan, "Token Transformation Matters: Towards Faithful Post-hoc Explanation for Vision Transformer", **CVPR 2024**.

[12] Weitai Kang, Mengxue Qu, Yunchao Wei, Yan Yan, "ACTRESS: Active Retraining for Semi-supervised Visual Grounding".

[13] Wenxin Chen, Mengxue Qu, Weitai Kang, Yan Yan, Yao Zhao, Yunchao Wei, "3DResT: A Strong Baseline for Semi-Supervised 3D Referring Expression Segmentation". **IEEE Transactions on Multimedia (TMM) 2025**.

[14] Bin Lei, Yuchen Li, Yiming Zeng, Tao Ren, Yi Luo, Tianyu Shi, Zitian Gao, Zeyu Hu, Weitai Kang, Qiuwu Chen , "Infant Agent: A Tool-Integrated, Logic-Driven Agent with Cost-Effective API Usage".

PROJECT EXPERIENCE

Mouse Behavior Analysis Aug. 2022 – Present
Computer Vision and Multimedia Laboratory @ UIC *Chicago, IL*

- Mouse detection, pose estimation, re-ID and behavior clustering.

Theory Analysis of Cosmic Gravity Based on Convolutional Neural Network June 2020 – Dec. 2020
Sun Yat-Sen University 2020 College Student Innovation and Entrepreneurship Training Program *Zhuhai, China*

- Based on the VGG model and the ResNet model, we classify the numerically simulated density fields generated by the three gravitational field models of the universe.

The application of machine learning in asset allocation May 2020 – Oct. 2020

- Use lightGBM to predict the rise and fall of ETFs for the next week, select the best 10 ETFs from 297 ETFs, and make weekly adjustments to maximize returns.

Research on default risk of P2P online loan based on machine learning Oct. 2018 – May 2019

- Build a Back-Propagation neural network to study the relationship between loan users' loan repayment status and the economic status of their city.

COMPETITION EXPERIENCE

Mathematics competition of Chinese College Student Oct. 2020

- Awarded Third Prize.

Mathematics competition of Chinese College Student Oct. 2019

- Awarded Third Prize.

China Undergraduate Mathematical Contest in Model Sep. 2019

- Awarded Third Prize.

Mathematical Contest In Modeling Jan. 2019

- Awarded Successful Participant.

PROFICIENCY AND SKILLS

Technical Skills: PyTorch/Torch, Python, C/C++, Linux, Git, LaTeX, Matlab, etc.
Languages: English (proficient), Mandarin (native), Cantonese (native)

REFERENCES

Yan Yan, Associate Professor, University of Illinois Chicago, yyan55@uic.edu