

Zhaokai Wang (王肇凯)

Ph.D. candidate at Shanghai Jiao Tong University

Email: "%s%s@sjtu.edu.cn" % (lastname,firstname)

Wechat: wzk_1015 Homepage: <https://www.wzk.plus> Google Scholar

EDUCATION

Shanghai Jiao Tong University

Ph.D. Student

University College London

Visiting Student

Peking University

Double Bachelor Degree in Economics

Beihang University

B.Eng in Computer Science

09/2022 - 06/2027 (Expected)

Advisor: Prof. Jifeng Dai

12/2025 - Present

Advisor: Prof. Jun Wang

09/2019 - 06/2022

09/2018 - 06/2022

Overall GPA: 3.80

EXPERIENCE

Centre for Artificial Intelligence, UCL

Research Intern, collaborated with Prof. Jun Wang

London

12/2025 - Present

OpenGVLab (InternVL Team), Shanghai AI Laboratory

Research Intern, collaborated with Prof. Jifeng Dai and Dr. Xizhou Zhu

Shanghai

TuTu. AI (Startup)

Co-founder & Research Scientist

Shanghai

06/2023 - 02/2024

Fundamental Vision Group, SenseTime

Research Intern, collaborated with Prof. Jifeng Dai and Dr. Xizhou Zhu

Beijing

Sea AI Lab

Research Intern, collaborated with Prof. Shuicheng Yan and Dr. Jibin Wu

02/2022 - 10/2022

CoLab, Institute of Artificial Intelligence at Beihang University

Research Intern, collaborated with Prof. Si Liu

Remotely at Beijing

08/2021 - 02/2022

Beijing

08/2019 - 06/2022

PUBLICATIONS

(*: Co-first authors)

1. GenExam: A Multidisciplinary Text-to-Image Exam

Zhaokai Wang, Penghao Yin, Xiangyu Zhao, Changyao Tian, Yu Qiao, Wenhai Wang, Jifeng Dai, Gen Luo
Preprint

2. InternVL3.5: Advancing Open-Source Multimodal Models in Versatility, Reasoning, and Efficiency

Weiyun Wang*, Zhangwei Gao*, Lixin Gu*, Hengjun Pu*, Long Cui*, Xingguang Wei*, Zhaoyang Liu*, Linglin Jing*, Shenglong Ye*, Jie Shao*, Zhaokai Wang*, Zhe Chen*, Hongjie Zhang, Ganlin Yang, Haomin Wang, Qi Wei, Jinhui Yin, Wenhao Li, Erfei Cui, Guanzhou Chen, Zichen Ding, Changyao Tian, Zhenyu Wu, Jingjing Xie, Zehao Li, Bowen Yang, Yuchen Duan, Xuehui Wang, Zhi Hou, Haoran Hao, Tianyi Zhang, Songze Li, Xiangyu Zhao, Haodong Duan, Nianchen Deng, Bin Fu, Yinan He, Yi Wang, Conghui He, Botian Shi, Junjun He, Yingtong Xiong, Han Lv, Lijun Wu, Wenqi Shao, Kaipeng Zhang, Huipeng Deng, Biqing Qi, Jiaye Ge, Qipeng Guo, Wenwei Zhang, Songyang Zhang, Maosong Cao, Junyao Lin, Kexian Tang, Jianfei Gao, Haian Huang, Yuzhe Gu, Chengqi Lyu, Huanze Tang, Rui Wang, Haijun Lv, Wanli Ouyang, Limin Wang, Min Dou, Xizhou Zhu, Tong Lu, Dahua Lin, Jifeng Dai, Weijie Su, Bowen Zhou, Kai Chen, Yu Qiao, Wenhai Wang, Gen Luo
Technical Report

3. Parameter-Inverted Image Pyramid Networks for Visual Perception and Multimodal Understanding

Zhaokai Wang, Xizhou Zhu, Xue Yang, Gen Luo, Hao Li, Changyao Tian, Wenhan Dou, Junqi Ge, Lewei Lu, Yu Qiao, Jifeng Dai
TPAMI 2025

4. Mono-InternVL-1.5: Towards Cheaper and Faster Monolithic Multimodal Large Language Models

Gen Luo, Wenhan Dou, Wenhao Li, Zhaokai Wang, Xue Yang, Changyao Tian, Hao Li, Weiyun Wang, Wenhai Wang, Xizhou Zhu, Yu Qiao, Jifeng Dai
Preprint

5. Does Reinforcement Learning Really Incentivize Reasoning Capacity in LLMs Beyond the Base Model?

Yang Yue, Zhiqi Chen, Rui Lu, Andrew Zhao, Zhaokai Wang, Yang Yue, Shiji Song, Gao Huang
NeurIPS 2025 Oral (Ranked Top 1 in NeurIPS 2025) & ICML 2025 AI4MATH Workshop Best Paper Award (2/172)

6. Multimodal Music Generation with Explicit Bridges and Retrieval Augmentation
Baisen Wang, Le Zhuo, Zhaokai Wang, Chenxi Bao, Chengjing Wu, Xuecheng Nie, Jiao Dai, Jizhong Han, Yue Liao, Si Liu
ISMIR 2025 LLM4Music Workshop
7. Vision-to-Music Generation: A Survey
Zhaokai Wang, Chenxi Bao, Le Zhuo, Jingrui Han, Yang Yue, Yihong Tang, Victor Shea-Jay Huang, Yue Liao
ISMIR 2025
8. Sparkle: Mastering Basic Spatial Capabilities in Vision Language Models Elicits Generalization to Composite Spatial Reasoning
Yihong Tang*, Ao Qu*, Zhaokai Wang*, Dingyi Zhuang*, Zhaofeng Wu, Wei Ma, Shenhai Wang, Yunhan Zheng, Zhan Zhao, Jinhua Zhao
EMNLP 2025 Findings & IJCAI 2025 MKLM Workshop Best Paper Award
9. Mono-InternVL: Pushing the Boundaries of Monolithic Multimodal Large Language Models with Endogenous Visual Pre-training
Gen Luo*, Xue Yang*, Wenhan Dou*, Zhaokai Wang*, Jiawen Liu, Jifeng Dai, Yu Qiao, Xizhou Zhu
CVPR 2025
10. SynerGen-VL: Towards Synergistic Image Understanding and Generation with Vision Experts and Token Folding
Hao Li, Changyao Tian, Jie Shao, Xizhou Zhu, Zhaokai Wang, Jinguo Zhu, Wenhan Dou, Xiaogang Wang, Hongsheng Li, Lewei Lu, Jifeng Dai
CVPR 2025
11. Parameter-Inverted Image Pyramid Networks
Xizhou Zhu*, Xue Yang*, Zhaokai Wang*, Hao Li, Wenhan Dou, Junqi Ge, Lewei Lu, Yu Qiao, Jifeng Dai
NeurIPS 2024 Spotlight - Ranked Top 10 in NeurIPS 2024 (among 15671 submissions), Top 2 in Computer Vision Area
12. ITINERA: Integrating Spatial Optimization with Large Language Models for Open-domain Urban Itinerary Planning
Yihong Tang*, Zhaokai Wang*, Ao Qu*, Yihao Yan*, Zhaofeng Wu, Dingyi Zhuang, Jushi Kai, Kebing Hou, Xiaotong Guo, Jinhua Zhao, Zhan Zhao, Wei Ma
EMNLP 2024 Industry Track & KDD 2024 UrbComp Workshop Best Paper Award
13. Auto MC-Reward: Automated Dense Reward Design with Large Language Models for Minecraft
Hao Li*, Xue Yang*, Zhaokai Wang*, Xizhou Zhu, Jie Zhou, Yu Qiao, Xiaogang Wang, Hongsheng Li, Lewei Lu, Jifeng Dai
CVPR 2024
14. Video Background Music Generation: Dataset, Method and Evaluation
Le Zhuo*, Zhaokai Wang*, Baisen Wang*, Yue Liao, Chenxi Bao, Stanley Peng, Songhao Han, Aixi Zhang, Fei Fang, Si Liu
ICCV 2023
15. Video Background Music Generation with Controllable Music Transformer
Shangzhe Di, Zeren Jiang, Si Liu, Zhaokai Wang, Leyan Zhu, Zexin He, Hongming Liu, Shuicheng Yan
ACM Multimedia 2021 Best Paper Award (1/1942)
16. Confidence-aware Non-repetitive Multimodal Transformers for TextCaps
Zhaokai Wang, Renda Bao, Qi Wu, Si Liu
AAAI 2021

SELECTED AWARDS AND HONORS

Best Paper Award, ICML AI4MATH Workshop 2025	2025
Best Paper Award, IJCAI MKLM Workshop 2025	2025
Best Paper Award, KDD Urban Computing Workshop (UrbComp) 2024	2024
Best Zero to One Award, Alibaba Creator@AI Entrepreneur Hackathon Finals	2024
Outstanding Graduate of Beihang University	2022
Best Paper Award, ACM Multimedia 2021	2021
Best Video Award, IJCAI 2021 Video Competition	2021
First Place, CVPR TextCaps Challenge	2020

ADDITIONAL INFORMATION

Talks:

- 2025.4: Talk on Mono-InternVL at Open Multimodal Gathering Workshop hosted by NUS ShowLab.

Activities

- **Conference Reviewer:** ICCV 2023 & 2025, ECCV 2024, CVPR 2024 & 2025, EMNLP 2024, NeurIPS 2024, ICLR 2025, ICML 2025, AAAI 2025.
- **Journal Reviewer:** ACM Computing Survey.
- **Teaching Assistant:** Fundamentals of Computers (2021), Software Engineering (2022).

Skills

- **Programming languages:** Python, C/C++, Java, JavaScript.
- **Scientific packages:** Numpy, Pytorch, Tensorflow.
- **English:** TOEFL 111 (S26) , GRE 327+4.0, CET-4 669, CET-6 612.