Zhaokai Wang (王肇凯)

Ph.D. candidate at Shanghai Jiao Tong University

wangzhaokai [AT] sjtu [dot] edu [dot] cn | wechat: wzk_1015 | `https://www.wzk.plus`

EDUCATION

Shanghai Jiao Tong University   09/2022 - Present
Joint Ph.D. program with Shanghai AI Laboratory   Advisor: Prof. Jifeng Dai
Peking University   09/2019 - 06/2022
Double bachelor degree in Economics
Beihang University   09/2018 - 06/2022
B.Eng in Computer Science   Overall GPA: 3.80

EXPERIENCE

OpenGVLab, Shanghai AI Laboratory   Shanghai
Research intern, collaborated with Prof. Jifeng Dai and Dr. Xizhou Zhu   11/2022 - Present
Fundamental Vision Group, SenseTime   Beijing
Research intern, collaborated with Prof. Jifeng Dai and Dr. Xizhou Zhu   02/2022 - 10/2022
Sea AI Lab   Remotely at Beijing
Research intern, collaborated with Prof. Shuicheng Yan and Dr. Jibin Wu   08/2021 - 02/2022
CoLab, Institute of Artificial Intelligence at Beihang University   Beijing
Research intern, collaborated with Prof. Si Liu   08/2019 - 06/2022

PUBLICATIONS

1. 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
   arXiv preprint

2. 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, Shenhao Wang, Yunhan Zheng, Zhan Zhao, Jinhua Zhao
   arXiv preprint

3. 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
   arXiv preprint

4. 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
   arXiv preprint

5. 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 (Top 2.5%)**

6. 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

7. Synergizing Spatial Optimization with Large Language Models for Open-Domain Urban Itinerary Planning
   Yihong Tang*, **Zhaokai Wang***, Ao Qu*, Yihao Yan*, Kebing Hou, Dingyi Zhuang, Xiaotong Guo, Jinhua Zhao, Zhan Zhao, Wei Ma
   KDD UrbComp 2024 **Best Paper Award**

8. 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

9. 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

10. 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 (1/542)**

11. 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, 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

Activities

- **Conference Reviewer:** ICCV 2023, CVPR 2024 & 2025, ECCV 2024, NeurIPS 2024, EMNLP 2024, AAAI 2025, ICLR 2025.

- **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.