Siteng Huang

siteng.huang@gmail.com · kyonhuang.top · Google Scholar (**600+** citations) · Github (**4.6k+** stars)

Education

Zhejiang University Hangzhou, China
Ph.D. - Computer Science Sep. 2019 – Jun. 2024

- Thesis: Model Transfer for Multimodal Understanding and Generation

Westlake University Hangzhou, China
Visiting Student & Joint Program Oct. 2018 – Jun. 2024

- Advisor: Prof. Donglin Wang
- Affiliated with Machine Intelligence Laboratory (MiLAB)

Wuhan University Wuhan, China
Bachelor of Engineering - Software Engineering Sep. 2015 – Jun. 2019

- GPA: 3.7/4.0, Rank: 3/244

Research Areas

Focusing on **multi-modal large models**, especially **vision-language models (VLMs)**, including

- **Generation/AIGC**: text-to-image/video (T2I/V) generation, customized & controllable generation, test-time diffusion intervention, multi-modal large language models (MLLMs)
- **Understanding**: text-video retrieval, compositional zero-shot learning, few-shot learning, visual grounding
- **Transfer**: parameter-efficient fine-tuning (PEFT/PETL), meta-learning, domain adaptation
- **Embodied AI**: vision-language-action models (VLAs), foundation models for robotics

Work Experience

DAMO Academy, Alibaba Group Hangzhou, China
Algorithm Expert Sep. 2024 – Present

Internship Experience

DAMO Academy & TongYi Lab, Alibaba Group Hangzhou, China
Research Intern Mar. 2022 – July. 2024

- Directors: Gong Biao, Yu Liu, and Deli Zhao.

Peer-reviewed Conference Publications (∗ denotes equal contribution)

[C1] **ProFD: Prompt-Guided Feature Disentangling for Occluded Person Re-Identification**
Can Cui*, **Siteng Huang***, Wenxuan Song, Pengxiang Ding, Zhang Min, Donglin Wang
ACM Multimedia 2024 (ACMMM 2024)

[C2] **PiTe: Pixel-Temporal Alignment for Large Video-Language Model**
Yang Liu, Pengxiang Ding, **Siteng Huang**, Min Zhang, Han Zhao, Donglin Wang
European Conference on Computer Vision 2024 (ECCV 2024)

[C3] **QUAR-VLA: Vision-Language-Action Model for Quadruped Robots**
Pengxiang Ding, Han Zhao, Wenxuan Song, Wenjie Zhang, Min Zhang, **Siteng Huang**, Ningxi Yang, *et al.*
European Conference on Computer Vision 2024 (ECCV 2024)

[C4] **Learning Disentangled Identifiers for Action-Customized Text-to-Image Generation**
Siteng Huang, Biao Gong, Yutong Feng, Xi Chen, Yuqian Fu, Yu Liu, Donglin Wang
IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024 (CVPR 2024)

[C5] **Troika: Multi-Path Cross-Modal Traction for Compositional Zero-Shot Learning**
Siteng Huang, Biao Gong, Yutong Feng, Yiliang Lv, Donglin Wang
IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024 (CVPR 2024)

[C6] **Check, Locate, Rectify: A Training-Free Layout Calibration System for Text-to-Image Generation**
Biao Gong*, **Siteng Huang***, Yutong Feng, Shiwei Zhang, Yuyuan Li, Yu Liu
IEEE/CVF Conference on Computer Vision and Pattern Recognition 2024 (CVPR 2024)

[C7] **DARA: Domain- and Relation-aware Adapters Make Parameter-efficient Tuning for Visual Grounding**
Ting Liu, Xuyang Liu, **Siteng Huang**, Honggang Chen, Quanjun Yin, Long Qin, Donglin Wang, Yue Hu
IEEE Conference on Multimedia Expo 2024 (ICME 2024)

[C8] **VGDiffZero: Text-to-image Diffusion Models Can Be Zero-shot Visual Grounders**
Xuyang Liu*, **Siteng Huang***, Yachen Kang, Honggang Chen, Donglin Wang
IEEE International Conference on Acoustics, Speech and Signal Processing 2024 (ICASSP 2024)

[C9] **Prompt-based Distribution Alignment for Unsupervised Domain Adaptation**
Shuanghao Bai, Min Zhang, Wanqi Zhou, **Siteng Huang**, Zhirong Luan, Donglin Wang, Badong Chen
The 38th AAAI Conference on Artificial Intelligence (AAAI 2024)

[C10] **VoP: Text-Video Co-operative Prompt Tuning for Cross-Modal Retrieval**
Siteng Huang, Biao Gong, Yulin Pan, Jianwen Jiang, Yiliang Lv, Yuyuan Li, Donglin Wang
IEEE/CVF Conference on Computer Vision and Pattern Recognition 2023 (CVPR 2023)

[C11] **Reference-Limited Compositional Zero-Shot Learning**
Siteng Huang, Qiyao Wei, Donglin Wang
ACM International Conference on Multimedia Retrieval 2023 (ICMR 2023)

[C12] **Tree Structure-Aware Few-Shot Image Classification via Hierarchical Aggregation**
Min Zhang, **Siteng Huang**, Wenbin Li, Donglin Wang
European Conference on Computer Vision 2022 (ECCV 2022)

[C13] **Domain Generalized Few-shot Image Classification via Meta Regularization Network**
Min Zhang, **Siteng Huang**, Donglin Wang
IEEE International Conference on Acoustics, Speech and Signal Processing 2022 (ICASSP 2022)

[C14] **HINFShot: A Challenge Dataset for Few-Shot Node Classification in Heterogeneous Information Network**
Zifeng Zhuang, Xintao Xiang, **Siteng Huang**, Donglin Wang
ACM International Conference on Multimedia Retrieval 2021 (ICMR 2021)

[C15] **Pareto Self-Supervised Training for Few-Shot Learning**
Zhengyu Chen, Jixie Ge, Heshen Zhan, **Siteng Huang**, Donglin Wang
IEEE/CVF Conference on Computer Vision and Pattern Recognition 2021 (CVPR 2021)

[C16] **Attributes-Guided and Pure-Visual Attention Alignment for Few-Shot Recognition**
Siteng Huang, Min Zhang, Yachen Kang, Donglin Wang
The 35th AAAI Conference on Artificial Intelligence (AAAI 2021)

[C17] **DSANet: Dual Self-Attention Network for Multivariate Time Series Forecasting**
Siteng Huang, Donglin Wang, Xuehan Wu, Ao Tang
The 28th ACM International Conference on Information and Knowledge Management (CIKM 2019)

Preprints & Under Review (Only including publicly available work)

[P1] **Cobra: Extending Mamba to Multi-modal Large Language Model for Efficient Inference**
Han Zhao, Min Zhang, Pengxiang Ding, Wei Zhao, **Siteng Huang**, Donglin Wang

[P2] **Accelerating Diffusion Transformers with Token-wise Feature Caching**
Chang Zou, Xuyang Liu, Ting Liu, **Siteng Huang**, Linfeng Zhang

[P3] **Sparse-Tuning: Adapting Vision Transformers with Efficient Fine-tuning and Inference**
Ting Liu, Xuyang Liu, **Siteng Huang**, Liangtao Shi, Zunnan Xu, Yi Xin, Quanjun Yin, Xiaohong Liu

[P4] **M^{2}IST: Multi-Modal Interactive Side-Tuning for Memory-efficient Referring Expression Comprehension**
Xuyang Liu, Ting Liu, **Siteng Huang**, Yue Hu, Quanjun Yin, Donglin Wang, Honggang Chen

[P5] **Focus-Consistent Multi-Level Aggregation for Compositional Zero-Shot Learning**
Fengyuan Dai, **Siteng Huang**, Min Zhang, Biao Gong, Donglin Wang

Professional Services

- **Journal reviewer** for TNNLS, ACM TIST, JVCI, CPE
- **Conference reviewer** for CVPR, ICCV, ECCV, AAAI, IJCAI, ICME, ICMR, ACCV, ICPR

Honors & Awards

- Outstanding Graduates, Zhejiang University 2024
- **National Scholarship (Top 1%, highest scholarship from Ministry of Education of China)** 2020
- Postgraduate Academic Fellowship, Zhejiang University 2019–2024
- Excellent Student Scholarship, Wuhan University (Top 5%) 2016–2019