

# Siteng Huang

siteng.huang@gmail.com · kyonhuang.top · Google Scholar (700+ citations) · Github (4.8k+ stars)

## Education

### Zhejiang University

Ph.D. - Computer Science

Hangzhou, China

Sep. 2019 – Jun. 2024

- Thesis: Model Transfer for Multimodal Understanding and Generation

### Westlake University

Visiting Student & Joint Program

Hangzhou, China

Oct. 2018 – Jun. 2024

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

### Wuhan University

Bachelor of Engineering - Software Engineering

Wuhan, China

Sep. 2015 – Jun. 2019

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

## Work Experience

### DAMO Academy, Alibaba Group

Algorithm Expert

Hangzhou, China

Sep. 2024 – Present

- Responsible for the development of multimodal large models

## Internship Experience

### DAMO Academy & TongYi Lab, Alibaba Group

Research Intern

Hangzhou, China

Mar. 2022 – July. 2024

- Fundamental Visual Intelligence Team for Tongyi Wanxiang
- Directors: Gong Biao, Yu Liu, and Deli Zhao

## Peer-reviewed Conference Publications (\*: equal contribution, ✉: corresponding author.)

- [C1] **Accelerating Diffusion Transformers with Token-wise Feature Caching**  
Chang Zou\*, Xuyang Liu\*, Ting Liu, **Siteng Huang**, Linfeng Zhang  
*The 13th International Conference on Learning Representations (ICLR 2025)*
- [C2] **Cobra: Extending Mamba to Multi-Modal Large Language Model for Efficient Inference**  
Han Zhao, Min Zhang, Wei Zhao, Pengxiang Ding, **Siteng Huang**, Donglin Wang  
*The 39th AAAI Conference on Artificial Intelligence (AAAI 2025)*
- [C3] **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)*
- [C4] **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)*
- [C5] **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)*
- [C6] **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)*
- [C7] **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)*
- [C8] **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)*

- [C9] **DARA: Domain- and Relation-aware Adapters Make Parameter-efficient Tuning for Visual Grounding**  
Ting Liu, Xuyang Liu, **Siteng Huang**, Honggang Chen, Qunjun Yin, Long Qin, Donglin Wang, Yue Hu  
*IEEE Conference on Multimedia Expo 2024 (ICME 2024)*
- [C10] **VGDifZero: 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)*
- [C11] **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)*
- [C12] **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)*
- [C13] **Reference-Limited Compositional Zero-Shot Learning**  
**Siteng Huang**, Qiyao Wei, Donglin Wang  
*ACM International Conference on Multimedia Retrieval 2023 (ICMR 2023)*
- [C14] **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)*
- [C15] **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)*
- [C16] **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)*
- [C17] **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)*
- [C18] **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)*
- [C19] **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] **Compression with Global Guidance: Towards Training-free High-Resolution MLLMs Acceleration**  
Xuyang Liu, Ziming Wang, Yuhang Han, Yingyao Wang, Jiale Yuan, Jun Song, Bo Zheng, Linfeng Zhang, **Siteng Huang**, Honggang Chen
- [P2] **QUART-Online: Latency-Free Large Multimodal Language Model for Quadruped Robot Learning**  
Xinyang Tong, Pengxiang Ding, Donglin Wang, Wenjie Zhang, Can Cui, Mingyang Sun, Yiguo Fan, Han Zhao, Hongyin Zhang, Yonghao Dang, **Siteng Huang**, Shangke Lyu
- [P3] **Score and Distribution Matching Policy: Advanced Accelerated Visuomotor Policies via Matched Distillation**  
Bofang Jia, Pengxiang Ding, Can Cui, Mingyang Sun, Pengfang Qian, **Siteng Huang**, Zhaoxin Fan, Donglin Wang
- [P4] **CARP: Visuomotor Policy Learning via Coarse-to-Fine Autoregressive Prediction**  
Zhefei Gong, Pengxiang Ding, Shangke Lyu, **Siteng Huang**, Mingyang Sun, Wei Zhao, Zhaoxin Fan, Donglin Wang
- [P5] **Rethinking Token Reduction in MLLMs: Towards a Unified Paradigm for Training-Free Acceleration**  
Yuhang Han\*, Xuyang Liu\*, Pengxiang Ding, Donglin Wang, Honggang Chen, Qingsen Yan, **Siteng Huang**<sup>✉</sup>

- [P6] **Sparse-Tuning: Adapting Vision Transformers with Efficient Fine-tuning and Inference**  
Ting Liu, Xuyang Liu, **Siteng Huang**, Liangtao Shi, Zunnan Xu, Yi Xin, Qunjun Yin, Xiaohong Liu
- [P7] **M<sup>2</sup>IST: Multi-Modal Interactive Side-Tuning for Memory-efficient Referring Expression Comprehension**  
Xuyang Liu, Ting Liu, **Siteng Huang**<sup>✉</sup>, Yue Hu, Qunjun Yin, Donglin Wang, Honggang Chen<sup>✉</sup>
- [P8] **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, TCSVT, ACM TIST, JVCJ, 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