

# Siteng Huang

siteng.huang@gmail.com · kyonhuang.top · Google Scholar (1500+ citations) · Github (5k+ stars)

## Work Experience

### DAMO Academy, Alibaba Group

Algorithm Expert

Hangzhou, China  
Sep. 2024 – Present

- Embodied Intelligence Lab for RynnBot
- Projects: RynnVLA-001, WorldVLA, AffordDex

### TongYi Lab & DAMO Academy, Alibaba Group

Research Intern

Hangzhou, China  
Mar. 2022 – July. 2024

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

## 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

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

### Wuhan University

Bachelor of Engineering - Software Engineering

Wuhan, China  
Sep. 2015 – Jun. 2019

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

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

### [C1] VLA-Adapter: An Effective Paradigm for Tiny-Scale Vision-Language-Action Model

Yihao Wang\*, Pengxiang Ding\*, Lingxiao Li, Can Cui, Zirui Ge, Xinyang Tong, Wenxuan Song, Han Zhao, Wei Zhao, Pengxu Hou, **Siteng Huang**, Yifan Tang, Wenhui Wang, Ru Zhang, Jianyi Liu, Donglin Wang  
*The 40th AAAI Conference on Artificial Intelligence (AAAI 2026)*

### [C2] Towards Affordance-Aware Robotic Dexterous Grasping with Human-like Priors

Haoyu Zhao\*, Linghao Zhuang\*, Xingyue Zhao\*, Cheng Zeng, Haoran Xu, Yuming Jiang, Jun Cen, Kexiang Wang, Jiayan Guo, **Siteng Huang**✉, Xin Li, Deli Zhao, Hua Zou✉  
*The 40th AAAI Conference on Artificial Intelligence (AAAI 2026)*

### [C3] Filter, Correlate, Compress: Training-Free Token Reduction for MLLM Acceleration

Yuhang Han\*, Xuyang Liu\*, Zihan Zhang, Pengxiang Ding, Junjie Chen, Honggang Chen, Donglin Wang, Qingsen Yan, **Siteng Huang**✉

*The 40th AAAI Conference on Artificial Intelligence (AAAI 2026)*

### [C4] Global Compression Commander: Plug-and-Play Inference Acceleration for High-Resolution Large Vision-Language Models

Xuyang Liu, Ziming Wang, Junjie Chen, Yuhang Han, Yingyao Wang, Jiale Yuan, Jun Song, **Siteng Huang**, Honggang Chen

*The 40th AAAI Conference on Artificial Intelligence (AAAI 2026)*

### [C5] SSR: Enhancing Depth Perception in Vision-Language Models via Rationale-Guided Spatial Reasoning

Yang Liu, Ming Ma, Xiaomin Yu, Pengxiang Ding, Han Zhao, Mingyang Sun, **Siteng Huang**, Donglin Wang  
*The 39th Annual Conference on Neural Information Processing Systems (NeurIPS 2025)*

### [C6] Long-VLA: Unleashing Long-Horizon Capability of Vision Language Action Model for Robot Manipulation

Yiguo Fan, Shuanghao Bai, Xinyang Tong, Pengxiang Ding, Yuyang Zhu, Hongchao Lu, Fengqi Dai, Wei Zhao, Yang Liu, **Siteng Huang**, Zhaoxin Fan, Badong Chen, Donglin Wang

*The 9th Annual Conference on Robot Learning (CoRL 2025)*

- [C7] **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  
*International Conference on Computer Vision 2025 (ICCV 2025)*
- [C8] **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  
*The 2025 IEEE International Conference on Robotics and Automation (ICRA 2025)*
- [C9] **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)*
- [C10] **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)*
- [C11] **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)*
- [C12] **PiT: 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)*
- [C13] **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)*
- [C14] **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)*
- [C15] **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)*
- [C16] **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)*
- [C17] **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)*
- [C18] **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)*
- [C19] **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)*
- [C20] **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)*
- [C21] **Reference-Limited Compositional Zero-Shot Learning**  
**Siteng Huang**, Qiya Wei, Donglin Wang  
*ACM International Conference on Multimedia Retrieval 2023 (ICMR 2023)*
- [C22] **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)*

- [C23] **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)*
- [C24] **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)*
- [C25] **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)*
- [C26] **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)*
- [C27] **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)*

## Peer-reviewed Journal Publications (\*: equal contribution, $\square$ : corresponding author.)

---

- [J1] **M2IST: Multi-Modal Interactive Side-Tuning for Memory-efficient Referring Expression Comprehension**  
Xuyang Liu, Ting Liu, **Siteng Huang** $\square$ , Yue Hu, Quanjun Yin, Donglin Wang, Honggang Chen $\square$   
*IEEE Transactions on Circuits and Systems for Video Technology (TCSVT 2025)*

## Preprints & Under Review (Only including publicly available work.)

---

- [P1] **HiF-VLA: Hindsight, Insight and Foresight for Vision-Language-Action models**  
Minghui Lin, Pengxiang Ding, Shu Wang, Zifeng Zhuang, Yang Liu, Xinyang Tong, Wenxuan Song, Shangke Lyu, **Siteng Huang** $\square$ , Donglin Wang $\square$
- [P2] **RynnVLA-002: A Unified Vision-Language-Action and World Model**  
Jun Cen\*, **Siteng Huang**\*, Yuqian Yuan\*, Kehan Li\*, Hangjie Yuan, Chaohui Yu, Yuming Jiang, Jiayan Guo, Hao Luo, Fan Wang, Xin Li, Deli Zhao, Hao Chen
- [P3] **High-Fidelity Simulated Data Generation for Real-World Zero-Shot Robotic Manipulation Learning with Gaussian Splatting**  
Haoyu Zhao\*, Cheng Zeng\*, Linghao Zhuang\*, Yaxi Zhao, Shengke Xue, Hao Wang, Xingyue Zhao, Zhongyu Li, Kehan Li, **Siteng Huang** $\square$ , Mingxiu Chen, Xin Li, Deli Zhao, Hua Zou $\square$
- [P4] **RynnVLA-001: Using Human Demonstrations to Improve Robot Manipulation**  
Yuming Jiang, **Siteng Huang**, Shengke Xue, Yaxi Zhao, Jun Cen, Sicong Leng, Kehan Li, Jiayan Guo, Kexiang Wang, Mingxiu Chen, Fan Wang, Deli Zhao, Xin Li
- [P5] **WorldVLA: Towards Autoregressive Action World Model**  
Jun Cen, Chaohui Yu, Hangjie Yuan, Yuming Jiang, **Siteng Huang**, Jiayan Guo, Xin Li, Yibing Song, Hao Luo, Fan Wang, Deli Zhao, Hao Chen
- [P6] **VARD: Efficient and Dense Fine-Tuning for Diffusion Models with Value-based RL**  
Fengyuan Dai\*, Zifeng Zhuang\*, Yufei Huang, **Siteng Huang**, Bangyan Liao, Donglin Wang, Fajie Yuan
- [P7] **SSR: Enhancing Depth Perception in Vision-Language Models via Rationale-Guided Spatial Reasoning**  
Yang Liu, Ming Ma, Xiaomin Yu, Pengxiang Ding, Han Zhao, Mingyang Sun, **Siteng Huang**, Donglin Wang
- [P8] **OpenHelix: A Short Survey, Empirical Analysis, and Open-Source Dual-System VLA Model for Robotic Manipulation**  
Can Cui, Pengxiang Ding, Wenxuan Song, Shuanghao Bai, Xinyang Tong, Zirui Ge, Runze Suo, Wanqi Zhou, Yang Liu, Bofang Jia, Han Zhao, **Siteng Huang**, Donglin Wang
- [P9] **Unicorn: Text-Only Data Synthesis for Vision Language Model Training**  
Xiaomin Yu, Pengxiang Ding, Wenjie Zhang, **Siteng Huang**, Songyang Gao, Chengwei Qin, Kejian Wu, Zhaoxin Fan, Ziyue Qiao, Donglin Wang

[P10] **Exploring the Evolution of Physics Cognition in Video Generation: A Survey**

Minghui Lin, Xiang Wang, Yishan Wang, Shu Wang, Fengqi Dai, Pengxiang Ding, Cunxiang Wang, Zhen-grong Zuo, Nong Sang, **Siteng Huang<sup>✉</sup>**, Donglin Wang<sup>✉</sup>

[P11] **Humanoid-VLA: Towards Universal Humanoid Control with Visual Integration**

Pengxiang Ding, Jianfei Ma, Xinyang Tong, Binghong Zou, Xinxin Luo, Yiguo Fan, Ting Wang, Hongchao Lu, Panzhong Mo, Jinxin Liu, Yuefan Wang, Huaicheng Zhou, Wenshuo Feng, Jiacheng Liu, **Siteng Huang**, Donglin Wang

[P12] **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

[P13] **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

[P14] **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 RA-L, TNNLS, TCSV, TMM, ACM TIST, JVCI, CPE
- **Conference reviewer** for CVPR, ICCV, ECCV, AAAI, IJCAI, ACMMM, ICME, ICMR, ACCV, ICPR

## Honors & Awards

---

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