

Heng Su

Tel: +86 18172181265 | Email: hengsu@stu.cqu.edu.cn

EDUCATION

Chongqing University

M.S. in Computer Science | Advisor: Prof. Chao Chen

Chongqing, China
Sep. 2024 – Jun. 2027 (Expected)

Average Score: 88.16/100.

Chongqing University of Posts and Telecommunications

B.E. in Software Engineering

Chongqing, China
Sep. 2020 – Jun. 2024

Average Score: 84.06/100, ranked 6/271.

RESEARCH INTERESTS

World Action Model; Generalizable Robot Manipulation; 3D Affordance Grounding.

SELECTED PUBLICATIONS

- [1] **H. Su**, M. Xie, N. Cao, Y. Ding, B. Shao, X. Long, ... C. Chen. OVA-Fields: Weakly Supervised Open-Vocabulary Affordance Fields for Robot Operational Part Detection. ICCV 2025, 6385–6395. **CCF-A**.
- [2] B. Shao, M. Xie, **H. Su**, W. Zhang, M. Li, Y. Ding, F. Giunchiglia, C. Chen. GSAM: A Generalizable and Safe Robotic Framework for Articulated Object Manipulation. PPSN 2026. **CCF-B**.
- [3] T. Xu, **H. Su**, B. He, A. Tian, J. Guo. Influence of Multiple Spatiotemporal Resolutions on the Performance of Urban Growth Simulation Models. iScience, 27(1), 2024. **JCR Q1, IF 4.4**.
- [4] T. Xu, **H. Su**, X. Xiong, W. Wang, A. Dewan. Optimization of Land Subsidence Prediction Features Using Machine Learning and SHAP Analysis with Sentinel-1 InSAR Data in Chengdu and Chongqing, China. Geocarto International, 41(1), 2617682, 2026. **JCR Q1, IF 3.8**.

MANUSCRIPTS UNDER REVIEW

- [1] **H. Su**, Y. Xie, H. Zheng, Z. Yan, X. Dai, M. Li, M. Xie, H. Yang, C. Chen. ArcVLA: Action-Grounded Representation Calibration for Robust Vision-Language-Action Models. Manuscript under review.
- [2] **H. Su**, C. Chen, M. Xie, ... F. Gu. MLLM-Afford: Grounding 3D Affordances for Embodied Robots via Multimodal Large Language Models. Manuscript under review.

RESEARCH EXPERIENCE

ArcVLA: Action-Grounded Representation Calibration for Robust Vision-Language-Action Models

Apr. 2026 – Jul. 2026

First Author | Manuscript under review

- Developed ArcVLA to bridge the representation gap between the vision and action modalities, enabling VLA models to better understand tasks and perform robust manipulation.
- Proposed an action-anchored calibration pathway that pulls visual-language representations closer to action representations. The framework also uses an error-routed curriculum and temporal consistency, without additional parameters or training data.
- Across seven LIBERO-plus perturbations, six RoboTwin 2.0 tasks, and 30 dual-arm trials, improved $\pi_{0.5}$ from 86.7% to 93.7% on LIBERO-plus (58.7% to 97.7% under sensor noise), from 81.7% to 88.3% on randomized RoboTwin, and from 13/30 to 19/30 in real-world tasks.

OVA-Fields: Weakly Supervised Open-Vocabulary Affordance Fields for Robot Operational Part Detection

May 2024 – Nov. 2024

First Author | Accepted at ICCV 2025

- Proposed OVA-Fields, a queryable 3D affordance field that grounds open-vocabulary instructions to fine-grained operational parts, moving beyond object-centric and closed-set methods without dense point-level annotations.
- Combined multilevel geometric encoding with enhanced semantic-affordance embeddings, aligning RGB-D and vision-language features through contrastive supervision derived from 2D heatmaps and language descriptions.
- Evaluated on two in-house environments, three ScanNet scenes, three SceneFun3D scenarios, and 120 real-robot language queries; achieved 52.4% mIoU (3.4× the strongest baseline) and 90% refrigerator-opening success, leading to a patent filing and a first-author ICCV 2025 paper.

MLLM-Afford: Grounding 3D Affordances for Embodied Robots via Multimodal Large Language Models

Jul. 2025 – Jan. 2026

First Author | Manuscript under review

- Proposed MLLM-Afford to bring multimodal large language models into scene-level 3D affordance understanding, jointly reasoning over natural-language instructions, dense interaction regions, and motion directions for executable robot manipulation.
- Adapted LLaVA-Video-7B-Qwen2 with LoRA and integrated dynamic point sampling, instruction-guided multimodal fusion, spectral-enhanced coarse-to-fine decoding, and a motion expert.

- Trained and evaluated on 230 SceneFun3D scenes (200 training, 30 validation), achieving 23.28 mIoU versus 14.00 for the strongest evaluated baseline and 64.73% Succ@r; validated the pipeline on a mobile manipulator opening doors, pulling drawers, and operating appliance handles.

PROFESSIONAL EXPERIENCE

Astribot

Jul. 2026 – Present

Robot Learning Algorithm Intern

- Investigating how generative models can learn stronger physical-action representations for robot learning.
- Building on LingBot-Video and LingBot-VA to improve the physical grounding of generated motion and action.
- Evaluating and optimizing the models on WorldArena 2.0, with ongoing efforts to improve benchmark performance.

vivo, Central Research Institute – Robotics Lab

Mar. 2026 – Jul. 2026

Algorithm Engineer Intern

- Deployed Pi-series and ACoT-VLA policies on a Franka robotic arm, producing a hardware-side inference pipeline that connects multimodal policy outputs to the robot control stack.
- Integrated Sonic- and VR-based teleoperation for Unitree G1, establishing a workflow for real-time control, data collection, and policy validation.
- Contributed these inference and teleoperation modules to the AgiBot Challenge simulation-to-real pipeline, supporting end-to-end validation in both simulation and real-robot environments.

SELECTED COMPETITIONS

ICRA 2026 AgiBot World Challenge: Reasoning to Action Track

Mar. 2026 – Jun. 2026

PrismBot Team | **Global Champion**

- Developed long-horizon reasoning-to-action manipulation policies for AgiBot Genie Sim 2.0 and the Genie G2 real-robot platform.
- Improved ACoT-VLA and Pi0.5 through task-aware trajectory optimization, keyframe-sensitive training objectives, and oversampling of failure-prone subtasks.
- Validated the policies in simulation and on the Genie G2 platform, contributing to PrismBot's Global Champion result.

ICRA 2026 AgiBot World Challenge: RoboChallenge WBC Track

May. 2026 – Jun. 2026

PrismBot Team | **Global 3rd Place**

- Developed VLA-based solutions for supermarket whole-body control tasks requiring instruction understanding, navigation, and precise manipulation under shelf-layout shifts.
- Extended the ACoT-VLA weighted-loss framework with CLIP-style contrastive objectives and frame-level keyframe weighting to strengthen visual-language grounding under mismatched object placements.
- Applied the enhanced policy in the RoboChallenge WBC track, contributing to PrismBot's Global 3rd Place result.

HONORS AND AWARDS

National Scholarship	2025
Advanced Individual in Scientific and Technological Academic Innovation, Chongqing University	2025
Anshuo Scholarship (Top 1%)	2022
Third Prize, APMCM Asia-Pacific Mathematical Contest in Modeling	2021
First-Class Academic Scholarship (7 awards)	2020–2026
Merit Student	2020–2026

TECHNICAL SKILLS

Programming & Machine Learning: Python, PyTorch; large-model fine-tuning, inference, and deployment. **Robotics & Systems:** ROS, SLAM, motion planning, teleoperation; real-robot manipulation with Franka and Realman platforms. **Language & Tools:** English (CET-6: 539; IELTS: 6.5); Linux, Docker, Git.