

Hua Wang

Ph.D. Researcher | World Models & Embodied AI | 3D Computer Vision
China | huawang@stu.pku.edu.cn

Education

Peking University

2023 – Expected Jun 2028

School of Computer Science | Ph.D. in Computer Science and Technology | GPA: 3.80
Research: 6D Object Pose Estimation & Tracking; World-Action Models for Robotic Manipulation

Hebei University of Technology

2019 – 2023

School of Artificial Intelligence and Data Science
B.Eng. in Intelligent Science and Technology (AI Pilot Program)

Research Experience

SkeleWAM: Skeleton World-Action Modeling for Efficient Robotic Manipulation

Second Author | *arXiv Preprint*, 2026

[Paper](#) | [Project](#)

- Set up and debugged **LIBERO-Plus simulation environments**, trained and tuned SkeleWAM, and evaluated **zero-shot robustness** across **10,030 task variants** spanning 7 perturbation categories.
- Conducted **ablation studies and quantitative analysis**; SkeleWAM achieved **85.9% task success** (+3.7 pp vs. Cosmos-Policy) with 57.1M parameters. Co-authored the paper and produced technical figures.

CLIP-6D: Vision-Language Learning for Zero-Shot 6D Pose Estimation

First Author | *ACM Multimedia 2025 (Oral)*

[Paper](#)

- Independently designed and implemented **CLIP-6D**, bridging **vision-language category semantics and 3D geometric representations** via geometric self-alignment, interactive representation learning, and knowledge distillation.
- Developed the end-to-end pipeline for **zero-shot 6D object pose estimation**, a key perception capability for downstream **robotic grasping and manipulation**; achieved **10.8% and 16.4% performance gains** on 5°/2 cm and 10°/5 cm metrics, with **5.4 FPS faster inference**.

Selected Engineering Projects

Text2Sim: Language-Guided 3D Simulation & Synthetic Data Generation

Guangdong Provincial Key-Area R&D Program – New Generation AI Flagship Project

- Independently **architected and implemented a text-to-simulation pipeline** integrating LLM-based planning, Blender, and Unreal Engine 5 (Chaos Physics) to generate physics-validated 3D scenes, RGB-D images, instance masks, calibrated cameras, and object 6D pose annotations.
- Validated **50/50 scenes** (6 categories, 195 objects), generating **1,000 annotated frames** in under 3 hours of Unreal Engine batch processing, with a maximum reprojection consistency error of **1.02×10^{-4} px**.

Unitree G1: Detection-Guided Bottle Grasping

National Key R&D Program of China

Integrated object detection with scripted motions and demonstrated bottle grasping on a Unitree G1 with an Inspire Robots dexterous hand.

Additional Selected Publications

HypoPose — Training-Free Unseen 6D Pose Estimation. *ACM Multimedia 2026 (Accepted)*, Third Author.

RCSPose — Category-Level 6D Object Pose Estimation. *IEEE/RSJ IROS 2026 (Accepted)*, Third Author.

Honors & Awards

Finalist (Top 2%), COMAP MCM/ICM Mathematical Modeling Contest | **National Scholarship (China)**

National Second Prize, Contemporary Undergraduate Mathematical Contest in Modeling

Technical Skills

Programming & ML: Python, PyTorch, OpenCV

3D Vision & Multimodal: RGB / RGB-D Perception, 6D Pose Estimation & Tracking, Vision-Language Models (CLIP)

Simulation & Robot Hardware: Blender, Unreal Engine 5, MuJoCo, Unitree G1, Inspire Robots Dexterous Hand