

Yurun Jin

yurunjin@mail.ustc.edu.cn +86 190 5549 9020

Interning at LinS Lab and LightOrigins, focusing on Agentic for robotics.

I study how world models understand the physical world, and how robots learn to act within it.

Education

University of Science and Technology of China

Sep. 2023 to present

School of Computer Science and Technology · B.S. in Computer Science

Research interests

World models and robot learning

- Action Condition Injection and Embodied World Models: I study how to inject action conditions, including robot kinematics, end-effector geometry, historical observations, and interaction events, into generative world models through image-domain representations, latent conditioning, or attention modulation, so that future-state prediction, action generation, and closed-loop control become more consistent.
- Mutual Reinforcement of WAM and AC-WM: I explore complementary roles between WAM and AC-WM world-model paradigms in action understanding, environment prediction, and policy generation. Action-centric representations can improve controllable prediction in world models, while world models can feed spatiotemporal prediction and state-transition capabilities back into action generation, action evaluation, and policy optimization.
- Physics-Informed Embodied Intelligence: I am interested in enabling world models to better understand contact, occlusion, geometric constraints, object dynamics, and causal state transitions, and in transferring this physical knowledge to long-horizon planning, dexterous manipulation, cross-scene generalization, and stable real-world deployment.

Research experience

STARRY: Spatial-Temporal Action-Centric World Modeling for Robotic Manipulation

Feb. 2026 - Present

Second Author, NeurIPS under review Embodied Intelligence Group, Zhongguancun Academy, Advisor: Kai Chen

- Tech: Python, PyTorch, Wan 2.2, Motus, RoboTwin 2.0, Depth Anything V3, ArkInfinity R5, SAPIEN
- Research Contribution and Results: STARRY is a spatial-temporal action-centric world model that improves action-generation policies. It jointly denoises future spatiotemporal latents and action sequences, and introduces Geometry-Aware Selective Attention Modulation (GASAM), which converts predicted depth and end-effector geometry into token-aligned weights for selective action-attention modulation. On RoboTwin 2.0, STARRY achieves 93.82% / 93.30% Clean / Randomized average success rates, and improves real-world average success from 42.5% to 70.8% compared with $\pi_{0.5}$.
- My Contribution: Developed the Motus + Wan 2.2 implementation for STARRY's joint future-state and action-generation setting, and built the training/evaluation pipeline for RoboTwin 2.0 benchmark experiments. Conducted real-world validation on the ArkInfinity R5 dual-arm robot, including data collection, processing, model deployment, and evaluation under real-robot conditions.

EA-WM: Event-Aware Generative World Model with KVAFs

Feb. 2026 - Present

Second Author, NeurIPS under review Embodied Intelligence Group, Zhongguancun Academy, Advisor: Kai Chen

- Tech: Python, PyTorch, Wan 2.2, DiT, WorldArena
- Research Contribution and Results: EA-WM is an event-aware generative world model built on the Wan 2.2 backbone. It proposes Structured Kinematic-to-Visual Action Fields (KVAFs), projecting robot-arm kinematic actions into the target camera view and feeding them to the world model in the same RGB domain, preserving kinematic topology without modifying the backbone. It also introduces an event-aware bidirectional fusion module to capture object-state changes and interaction dynamics. On the official WorldArena evaluation, EA-WM reaches SOTA on the Overall metric and ranks No. 2 on the public leaderboard of the Tsinghua FIB-Lab world-model benchmark.
- My Contribution: Built the initial training and evaluation pipeline for EA-WM, covering data processing, training scripts, and WorldArena evaluation. Conducted 50+ architecture-combination experiments around image-domain action-condition representations and injection mechanisms. Performed hyperparameter tuning and ablation on the WorldArena Benchmark, contributing to a first-place result.

Embodied Intelligence Group, Zhongguancun Academy, Advisor: Kai Chen

- Tech: Python, PyTorch, VLM/VLA, World Models, SimplierEnv, StarVLA
- Research Contribution and Results: PhysBrain is a multimodal foundation model for embodied control. It addresses the egocentric to allocentric viewpoint gap and the lack of embodied supervision when transferring large models to end-to-end robot control. The framework unifies hand-object pose estimation, world models, and VLA training, enabling egocentric data to provide supervision for robot end-to-end control. It shows stronger physical priors and state-transition reasoning on multiple long-horizon planning and contact-rich interaction benchmarks.
- My Contribution: Conducted an in-depth literature review on hand-object pose estimation, multimodal large models, and world models, and produced a technical-route survey. Used an HPC cluster to evaluate fine-tuned PhysBrain models on mainstream benchmarks for long-horizon planning, state-transition reasoning, and contact-rich interaction.

IntentVLA: Short Horizon Intent Modeling for Aliased Robot Manipulation

May 2026 - Present

Embodied Intelligence Group, Zhongguancun Academy, Advisor: Kai Chen

- Tech: Python, PyTorch, VLA, RoboTwin2, AliasBench, MemoryVLA, SimplierEnv, LIBERO, RoboCasa
- Research Contribution and Results: IntentVLA is a history-conditioned VLA framework for aliased observation settings. It encodes recent visual-observation sequences into short-horizon intent representations and uses them as conditions for action-chunk generation, reducing intent ambiguity, chunk conflicts, and unstable execution in frame-conditioned policies under partial observability. It also introduces AliasBench, a benchmark of 12 aliased perception tasks based on RoboTwin2, with matching training and evaluation environments for isolating and measuring short-horizon observation aliasing. Across AliasBench, SimplierEnv, LIBERO, and RoboCasa, IntentVLA improves rollout stability and task success rates over multiple VLA baselines.
- My Contribution: Fine-tuned and evaluated MemoryVLA on the 12 AliasBench tasks, providing comparison data for a memory-centric baseline in the paper.

Projects

Multi-Spectral Wide-Field Satellite Camera for TLE Monitoring · USTC Undergraduate Innovation Program

Sep. 2025 - Jun. 2026

GCINE3243, IMX455, Vitis HLS, FPGA, ISP, 8K Imaging

- Worked on a prototype payload for multi-spectral wide-field imaging on the MERLIS micro/nano-satellite platform and TLE wide-area remote sensing. Validated image acquisition, data readout, image processing, and video compression pipelines on GCINE3243 and Sony IMX455 platforms.
- Designed an 8K RAW-data ISP pipeline using Vitis HLS and Vitis Vision. The prototype has captured 8K images and provides a basis for onboard intelligent processing, including ISP calibration, event detection, target recognition, and ROI extraction.

Auto-Aiming System Development and Optimization · RoboMaster Competition

Nov. 2024 - Aug. 2025

C++, ROS 2, OpenCV, ML

- Developed a high-performance automated target-recognition and tracking-control system based on ROS 2.
- Recognized armor plates using prior light-bar geometry and classified armor-plate digits with a CNN.
- Extracted light-bar keypoints and applied PnP with optimized Kalman filtering to estimate target 3D pose.
- Implemented an MPC controller for predictive target tracking and ballistic compensation.
- Achieved robust recognition of targets beyond 5 m with angular velocities above 10 rad/s, and reached a physical-projectile hit rate above 50%.

Autonomous Ore Exchange System Development · RoboMaster Competition

Nov. 2024 - Aug. 2025

C++, Python, ROS 2, OpenCV, DL

- Developed an automated end-effector targeting, alignment, and operation system based on ROS 2.
- Implemented inverse kinematics and spatial trajectory planning for a 7-DoF robotic arm.
- Built binocular/RGB-D vision modules for target detection and 6D pose estimation, with robust target-feature extraction and spatial localization in complex environments.
- Achieved a success rate above 80% on the most difficult ore-exchange task in the competition.

Python, LLM, ReAct, Android World, Vector DB

- Built an LLM reasoning and UI interaction loop based on the ReAct paradigm and the open-source Android World framework.
- Designed and implemented a reflection mechanism that extracts successful action sequences from task execution traces and converts them into structured standard operating procedures stored in a skill library.
- Introduced semantic-similarity retrieval augmentation, enabling the agent to recall relevant past experience for similar tasks, avoid long reasoning from scratch, and improve task efficiency and success rate.
- Used the Android World framework to help the LLM accurately interpret Android accessibility view trees and output deterministic interaction commands, including tap, swipe, and text input, improving success rates on long-horizon mobile tasks.

Python, Golang, LLM, Docker, Kafka, MySQL

- Built an interactive agent-capability evaluation system with Kafka-based asynchronous session management and distributed Docker container environments.
- Designed multi-dimensional dynamic game test cases and configured Mock Agents to simulate complex user intents and edge cases, comprehensively testing target agents on context understanding, tool use, and error-recovery robustness.

Awards

RoboMaster 2025 League Champion / Regional Champion / National Runner-up

Aug. 2025

RoboGame 2024 USTC First Prize

Oct. 2024

Campus Activities

Hosts Club Vice President

Sep. 2025 - Present

Photography Association Vice President

Sep. 2024 - Present