

Jun Sun

Tel: +86 19962174897 | Email: luciojunsun@gmail.com | WeiXin: qw2667180048

EDUCATION BACKGROUND

Xi'an Jiaotong-Liverpool University

BSc Information and Computing Science

Suzhou, China

2022.9-2026.7

- **GPA:** 3.94/4.00, **Major GPA:** 4.00/4.00 (**Rank 1**)
- **Scholarship:** 25/26 University Academic Excellence Award (Top 5%)
24/25 University Academic Excellence Award (Top 5%)
23/24 University Academic Excellence Award (Top 5%)
- **IELTS:** 7.5 (Listening: 8.0, Reading: 9.0)

PUBLICATION & PRESENTATION

- **J. Sun**, B. Yang, J. Zhang, N. Ma, C. Wu, S. Zhang, Y. Huang, Q. Wang, S. Liang, Y. Chen. *TempoFit: Plug-and-Play Layer-Wise Temporal KV Memory for Long-Horizon Vision-Language-Action Manipulation*. **Accepted by IROS 2026** ([Paper](#))
- Y. Huang, N. Ma, W. Zhao, Z. Liu, **J. Sun**, Q. Wang, Y. Chen. *Force-Aware Residual DAgger via Trajectory Editing for Precision Insertion with Impedance Control*. **Accepted by IROS 2026** ([Paper](#))
- Y. Wu, Q. Gao, Y. Liu, **J. Sun**, Z. Li, Y. Jin, Y. Yue, and X. Zhu. *MiTpose: Multi-Granularity Guided Vision Transformer for Human Pose Estimation*. **Accepted by International Conference on Industrial Informatics (INDIN) 2025**

PROFESSIONAL EXPERIENCES

Westlake Robotics

Algorithm R&D Department, Intern

Hangzhou, China

2025.6-2025.10

Deformable-object manipulation is challenging because object state changes continuously, observations are partial/occluded, and small perception errors compound into unstable long-horizon actions.

- **Task Design & Evaluation:** Designed real-robot experimental protocols for folding, unfolding, and stacking tasks, with quantitative metrics to evaluate perception-action decision quality
- **Real-Robot Data & Training:** Led real-robot data collection and processing; fine-tuned models with physical robot data and conducted large-scale pretraining across tasks and robot embodiments to improve transferability
- **End-to-End Pipeline:** Built an end-to-end data and training pipeline spanning annotation, data conversion, model training, and online inference; improved deployability via distributed training and inference optimization
- **System Integration & Demos:** Achieved a dexterous-hand towel-folding demo and led dual-arm long-horizon garment folding, becoming the first team in China (after Figure) to demonstrate real-robot dexterous towel folding on a physical robotic platform

RESEARCH EXPERIENCES

TempoFit: Training-Free Temporal KV Memory for Long-Horizon Vision-Language-Action Manipulation

Research Assistant, Supervisor: Prof. Yaran Chen

2025.11-2026.3

Long-horizon robotic manipulation challenges memoryless VLA policies under occlusion, state aliasing, and subtle post-action changes, motivating efficient temporal reasoning without additional retraining or architectural modifications.

- **Temporal Memory Design:** Proposed TempoFit, a training-free retrofit equipping frozen VLA policies with history awareness by caching intermediate-layer prefix K/V states without changing parameters or input length.
- **Retrieval & Fusion:** Developed K-to-K retrieval with Frame-Gap Temporal Bias and norm-preserving residual loading, suppressing stale context and minimizing distribution shift under frozen weights during inference.
- **Benchmark Performance:** Improved LIBERO-Long success from 92.6% to 96.6% on and from 90.8% to 94.4% on QwenGR00T, with consistent gains on CALVIN benchmarks.
- **Real-Robot Validation:** Validated TempoFit on a Realman RM-65B across three multi-stage tasks, increasing full-task success by up to 14.2% with minor latency and memory overhead.

Multimodal Embodied Intelligence Model Training with Octo Framework and 3D Depth Perception

Research Assistant, Supervisor: Prof. Yaran Chen

2025.1-2025.6

Language-guided manipulation is limited by RGB-only geometric ambiguity, occlusion, and sim-to-real shift; robust VLA requires explicit 3D grounding for stable grasping decisions.

- **Depth-Aware Fusion:** Added a dedicated monocular-depth stream to extract multi-scale geometric features and align them with Octo embeddings; injected depth via gated cross-attention with FiLM modulation, improving 3D spatial-semantic alignment without modifying the core Octo VLA architecture
- **Dataset Construction:** Expanded Bridge by generating monocular depth from RGB sequences to form an RGB–Depth–Language corpus; collected teleop RGB-D real-robot data for fine-tuning and sim-to-real alignment
- **Training & Evaluation:** Trained with behavior cloning plus RL fine-tuning, improving unseen grasping success from ~77% to ~89%; validated robustness via adversarial occlusion and depth-noise tests
- **Research Contribution:** Benchmarked against OpenVLA and other baselines to isolate depth gains; co-authored technical reports and led weekly reviews improving architecture decisions and dataset/pipeline reliability

Indoor 3D Reconstruction and Novel View Synthesis with UDF-Guided 3DGS

Research Assistant, Supervisor: Prof. Yong Yue

2024.5-2024.12

In low-texture indoor scenes, appearance-driven 3D reconstruction often suffers from geometric ambiguity and unstable optimization, limiting surface fidelity and novel view synthesis quality.

- **Modeling:** Built a NeRF-style UDF (unsigned distance + normals) coupled with a 3DGS renderer, optimized via feature–geometry joint losses with normal/depth-consistency regularization for low-texture stability
- **Optimization & Pruning:** Devised a multi-stage UDF-guided optimization—coarse geometry initialization followed by refining Gaussian positions, scales, and opacities—and introduced voxel-based partitioning to efficiently apply UDF guidance for Gaussian clone/prune, concentrating primitives near the zero-level set, removing redundancies, and reducing instability in low-texture scenes.
- **Evaluation & Benchmarking:** Processed ScanNet++ with COLMAP calibration and point-cloud initialization; evaluated reconstruction and novel-view quality using Chamfer Distance and PSNR, outperforming strong baselines

SKILLS & HOBBIES

Languages: Chinese (Native), English (Fluent), Spanish (Fluent)

Skills:

- 1) Programming & Machine Learning: Java, Python, PyTorch, JAX
- 2) Embodied AI & Robotics: Vision-Language-Action (Octo, OpenVLA, Pi0), World-Action-Models
- 3) Mathematical Foundations: Calculus, Linear Algebra, Probability Theory