

Research Interests

Spatial intelligence, long-form video understanding, world models, and embodied AI, with an emphasis on structured spatial reasoning, memory, and reliable decision-making in open-world environments.

Education

East China Normal University, Shanghai, China

M.S.; GPA: 91/100 (3.8/4.0), rank 1 in cohort; advisors: Lizhuang Ma and Xin Tan

Sep. 2024 – Jun. 2027 (expected)

First-Class Scholarship

Shanghai Normal University, Shanghai, China

B.S.; GPA: 92/100 (3.9/4.0)

Oct. 2020 – Jun. 2024

First-Class Scholarship

Selected Publications and Manuscripts

Towards Spatial Supersensing in the Wild

Tianjun Gu, Tianyu Xin, Kuan Zhang, Bowen Yang, Kok-Chung Chua, Peize Li, Xinran Zhang, Yupeng Chen, Qiyue Zhao, Qinlei Xie, Jianhang Liu, Yucheng Lu, Yinan Han, Marco Pavone, Yiming Li[†]

[ECCV 2026 \(Poster\)](#) | [Project](#) | [†] Corresponding author

Self-Evolving Spatial Belief for Video Spatial Reasoning

Tianjun Gu, Shao Zeng, Ruixin Zhang, Haoran Li, Chenghua Gong, Haonan Wang, Yuan Xie, Jingyu Gong, Jingyun Zhang, Jun Wang, Rizen Guo, Lizhuang Ma, Shouhong Ding, Xin Tan

[Under Review](#) | Manuscript available upon request

Vision-language models lag human performance on physical dynamics and intent reasoning

Tianjun Gu, Jingyu Gong, Zhizhong Zhang, Yuan Xie, Lizhuang Ma, Xin Tan, Athanasios V

[Nature Communications \(2nd Minor Revision\)](#) | [arXiv:2601.01547](#)

DORAEMON: Decentralized Ontology-aware Reliable Agent with Enhanced Memory Oriented Navigation

Tianjun Gu, Linfeng Li, Xuhong Wang, Chenghua Gong, Jingyu Gong, Zhizhong Zhang, Yuan Xie, Lizhuang Ma, Xin Tan

[Under Review](#) | [arXiv:2505.21969](#) | [Project](#)

Efficient Multimodal Large Language Models: A Survey

Yizhang Jin*, Jian Li*, **Tianjun Gu***, Yexin Liu, Bo Zhao, Jinxiang Lai, Zhenye Gan, Yabiao Wang, Chengjie Wang, Xin Tan, Lizhuang Ma

[Visual Intelligence, 3:27, 2025](#) | [Paper](#) | * Equal contribution

From Enhancement to Understanding: Build a Generalized Bridge for Low-light Vision via Semantically Consistent Unsupervised Fine-tuning

Sen Wang, Shao Zeng, **Tianjun Gu**, Zhizhong Zhang, Ruixin Zhang, Shouhong Ding, Jingyun Zhang, Jun Wang, Xin Tan, Yuan Xie, Lizhuang Ma

[ICCV 2025 \(Poster\)](#) | [Paper](#) | pp. 13804–13814

Diversified Generation of Theatrical Masks Based on SASGAN

Tianjun Gu, Suyu Xiong, Xiao Lin

[Journal of Graphics, 45\(1\):102, 2024](#) | [DOI](#)

Honors and Awards

- National-level Excellent Completion Award, Undergraduate Innovation and Entrepreneurship Program (PI).
- Provincial/Municipal-level Excellent Completion Award, Undergraduate Innovation and Entrepreneurship Program (PI).
- Meritorious Winner, Mathematical Contest in Modeling / Interdisciplinary Contest in Modeling.
- Third Prize, National College Student Mathematics Competition.

Research Experience

NYU Center for Robotics and Embodied Intelligence (CREO)

Jul. 2026 – Present

Visiting Research Intern | Advisor: Chen Feng; AI4CE Lab, NYU Tandon School of Engineering

- Investigating **active spatial supersensing** at AI4CE: how agents use spatial memory to guide observation, integrate multi-view and panoramic evidence, and build reusable knowledge for future tasks.

Tsinghua University Spatial Intelligence Lab

Nov. 2025 – Present

Visiting Research Intern | Advisor: Yiming Li; Beijing, China

- First-authored *Towards Spatial Supersensing in the Wild* (ECCV 2026), introducing VSI-Super-Wild, a four-task benchmark for continual agent-object-environment world-state reasoning with **442 videos (284.52 hours)** and **6,980 human-verified QA pairs**; evaluated **13 MLLMs** and found average performance drop from **35.0** on 0–10-minute videos to **26.3** beyond 120 minutes.
- Contributed to additional collaborative projects on generalist agents and foundation-model evaluation, including *Towards Generalist Game Players: An Investigation of Foundation Models in the Game Multiverse* (arXiv:2605.09965).

Digital Media & Computer Vision Lab (DMCV)

Sep. 2024 – Present

Graduate Researcher | SJTU-ECNU Joint Lab; Advisors: Lizhuang Ma and Xin Tan

- Developed DORAEMON, a zero-shot navigation framework with hierarchical semantic-spatial and topological memory and safety-aware VLM planning; achieved state-of-the-art SR and SPL on HM3D, MP3D, and GOAT.
- Contributed to national-level, municipal, and industry-sponsored research programs on embodied intelligence, autonomous driving, agent evaluation and safety, and AI-enabled social simulation.

Industry Experience

Meituan, Infrastructure R&D Platform / M17

Mar. 2026 – Present

Foundation Model Research Intern | Shanghai, China

- Improved LongCat-Flash-3B's omni-modal capabilities, focusing on multilingual instruction following through targeted data construction, supervised post-training, and systematic evaluation.
- Built an end-to-end GUI-agent data engine for LongCat-Flash, converting web-sourced videos into interaction trajectories and training data across generation, quality control, and model post-training.
- Developed reinforcement-learning schemes for interactive multimodal models spanning audio-visual and spoken interaction, covering task construction, reward design, and multi-turn policy optimization; also explored interactive world models through benchmark and memory-oriented evaluation design.

Tencent Youtu Lab

Jun. 2025 – Mar. 2026

Research Intern | Shanghai, China

- Developed a **self-evolving spatial agent** (under review) with query-relevant belief access, targeted video retrieval, and local updates; achieved **87.1** on VSI-Bench, versus **50.8** with static belief and **59.9** under the same frame budget.
- Proposed **Teleo-Spatial Intelligence** and built **EscherVerse** from **11,328 videos**, revealing physical-intent reasoning gaps across **27 VLMs** (Nature Communications, 2nd Minor Revision), with an **8,000-example benchmark** and a **35,963-example instruction-tuning set**.
- Explored autonomous research agents and automated research workflows, and supported post-training of Youtu foundation models through data construction, training-pipeline development, and evaluation.

Baidu, MEG / Library Strategy Group

Jan. 2024 – Jul. 2025

AIGC Multi-Model Algorithm Intern | Shanghai, China

- Built scenario- and character-specific LoRA datasets and fine-tuned image-generation models, improving animated-character consistency by 38 percentage points.
- Developed automated multi-character LoRA training pipelines and supported the production launch of Baidu Wenku's one-sentence comic feature.

SAIC Anji Intelligent IoT Technology Co., Ltd.

Sep. 2022 – Jun. 2023

Machine Vision Algorithm Engineer | Shanghai, China

- Built cross-site warehouse vision datasets using CenterNet-based candidate mining, similarity filtering, and end-to-end preprocessing pipelines.
- Trained and deployed workwear-detection models across multiple warehouse sites, iterating through on-site evaluation, error analysis, and operational feedback.

Academic Service

Reviewer: NeurIPS, ECCV. | **Conference Volunteer:** CCF CAD/CG.