

Yimo Cai

Email: caiym23@mails.tsinghua.edu.cn Github Weixin: Cai_YM_THU



Education

Tsinghua University, 3rd-year undergraduate majoring in Electronic Engineering. Sept 2023 – Now

- GPA: 3.96/4.0 (6/265, **top 3%**)
- **Related Courses (Graded A- or above):** Advanced Calculus, Linear Algebra, Discrete Mathematics, Probability and Stochastic Process, **Electromagnetic Fields and Waves**, **Electronic Circuits and Systems**, **Digital Logic and Processor**, **Signals and Systems**, **Communication and Networks**, Programming Basics (C/C++), Data Structures and Algorithms, Media and Recognition, Visual Auditory Information System.

Research Experience

Undergraduate Research Assistant, Lab of Prof. Duan, Tsinghua University Oct 2024 – Now

- Research focus on 3D vision, video generation, and multimodal model finetuning.
- Contributed to **Video-T1**, a test-time scaling framework for video generation.
- Conducting SFT and GRPO finetuning on MLLMs to improve: (1) 2D perception (e.g., Visual-RFT), (2) 3D understanding (e.g., Depth-LM, Spatial-MLLM).
- Experience with training pipelines and evaluation protocols for MLLM.
- Contributed to **Spatial-TTT**, a test-time training strategy for Spatial Understanding for MLLM.

Technical or Engineering Experience

Web Backend Development, EE Student Association of Science and Backend Technology July 2023 – Now

- Developed backend systems for student application management (TypeScript, GraphQL).
- Maintained the online platform for the THUAI programming competition. [Repo Link](#)

Publications

Video-T1: Test-Time Scaling for Video Generation Mar 2025
Fangfu Liu, Hanyang Wang, **Yimo Cai**, Kaiyan Zhang, Xiaohang Zhan, Yueqi Duan

- Responsibilities: code contributions, experiment implementation, visualization and analysis.
- Accepted to **ICCV 2025**.

arXiv:2503.18942

[Preprint]Spatial-TTT: Streaming Visual-based Spatial Intelligence with Test-Time Training Feb 2026

- Responsibilities: Training Dataset Construction, part of the paper and visualization.

arXiv:2603.12255

Awards

Excellence Award, THUEE Hardware Design Competition Sept 2024

- Built an intelligent bicycle system with real-time GPS data upload and remote control via WeChat program.
- Responsible for hardware control FSM via Arduino.

Second Prize, THUAI7 Multi-Agent Competition May 2024

- Led a team to design an automated multi-agent cooperation strategy for competitive tasks.
- Implemented in C++.

Studying Excellence Scholarship, Tsinghua University Oct 2024

Selected Projects

MIPS pipeline Processor

April 2025

- Implemented a five-stage MIPS pipeline processor in vivado using verilog and FPGA, being able to calculate and show results in 7-Seg.
- Tools: Vivado, Verilog and FPGA.

CoLearner-MCP: Automated Web-Learning Workflow

May 2025

- Built an MCP extension (based on Playwright-MCP) enabling LLMs to autonomously operate online learning platforms.
- Supports automated resource downloading and structured study workflows for THU platforms and Overleaf.
- Tools: TypeScript, Copilot.

Visual Agent RL Training

Aug 2025 – Oct 2025

- Attempted RL training on GUI tasks using AgentCPM-GUI; identified dataset limitations as the key bottleneck.
- Current work: SFT + RL finetuning of Qwen2.5-VL with VERL for 3D reasoning tasks such as coordinate estimation.

Other Activities

Teaching Assistant, University Physics A

June 2025

Database Instructor, Summer Seminar, EESAST.

Aug 2025

- Database Definition and Common Databases
- Database Security (e.g. injection)

Technical Skills

Core: Python, PyTorch, C++ **Additional:** JavaScript, TypeScript, GraphQL

Language

Native Chinese; English working proficiency (CET-6: 609).