Zicong Zhang

Shanghai, China | zhangzico@sjtu.edu.cn | +(86) 188 9181 2036

https://zhangzicong518.github.io

Education

Shanghai Jiao Tong University, B.S. in Computer Science — Sept 2023 – Present

- John Hopcroft Class, Zhiyuan College (Honor Program)
- GPA: 89.2/100
- **Coursework:**
  Math: Mathematical Logic(97), Information Theory(97), Linear Algebra(94)
  CS: Machine Learning Theory(98), Computer System Design(98), Program Design and Data Structure(96), Machine Learning(95), Algorithm(94), Programming Language Design and Implementation(94)

Research Experience

Multimodal Reasoning and Generic Agent, Microsoft Research Asia, intern — July 2025 – Jan 2026

- Proposed and implemented a tool-integrated RL paradigm that unifies long-chain-of-thought reasoning and tool-calling into a single hybrid thinking mode, enabling vision–language models to "think with images" more effectively
- Built and modified a large-scale, high-fidelity simulation environment that generates diverse, realistic agent–tool interaction trajectories, providing the massive training data required for robust tool-calling agents.

Continual Learning with Model Fusion, SJTU MIFA Lab, intern — July 2024 – April 2025

- Co-led development of a novel continual learning framework for VLMs by introducing model fusion
- Proposed aggregating the results of multiple decoupled task-specific models for prediction in zero-shot scenarios.
- Conducted extensive experiments on multiple benchmarks, demonstrate that outperforming the original pre-trained VLM and other state-of-the-art continual learning methods.

Awards

- Zhiyuan Honor Awards (Top 10% in SJTU) 2023,2024, 2025
- The Third Prize of Academic Scholarship (Top 30% in major) 2024, 2025

Publications

Enhanced Continual Learning of Vision-Language Models with Model Fusion

Haoyuan Gao*, Zicong Zhang*, Yuqi Wei, Linglan Zhao, Guilin Li, Yexin Li, Linghe Kong, Weiran Huang

Fisrt Workshop SCOPE The Thirteenth International Conference on Learning Representations

Tracing the Dance of Embeddings: Visualizing High-Dimensional Sample-wise Trajectory for Training Analysis

Yiming Liu*, Zicong Zhang*, Yun Lin, Ruofan Liu, Yuhuan Huang, Weiyu Kong, Jinsong Dong

[Under Review]

Projects

Environment Scaling for General Agentic Model – *Microsoft Research Asia* — Jul 2025 – Jan 2026

- Collected 150k real-world APIs from RapidAPI, MCP-Server and ToolBench; embedded api info to build a similarity-weighted dependency tool graph.
- Applied Louvain community detection to partition the graph into hundreds functional domains, and automatically derived a unified database schema per domain following read-write paradigm.
- Materialised each API as deterministic read/write operators on the domain-specific schema, enabling fully verifiable environment states
- Sampled long-horizon, logically coherent tool sequences on dependency graph, and generate diverse, high-fidelity agent–environment interaction trajectories for training

Scaling Active Perception for Multimodal Reasoning – *Microsoft Research Asia* Jul 2025 – Jan 2026

- Unified long-chain reasoning and tool-calling into a single hybrid thinking mode through SFT and RL, enabling vision-language models to autonomously switch thinking modes while "thinking with images"
- Curated high-quality long-CoT trajectories via teacher distillation and iterative rejection sampling and relieve tool-use hallucinations and inefficiencies
- Designed mode-aware process rewards that encourage the model to leverage both internal reasoning and external tools in a complementary manner
- Evaluated on MathVision, MathVista, HRBench, V* and other math or perception-oriented benchmarks, surpassing the original VLM baseline across various datasets

Continual Learning Framework for VLMs July 2024 - Jan 2025

Results in papers Enhanced Continual Learning of Vision-Language Models with Model Fusion

- Designed a continual learning framework for VLMs by introducing model fusion
- Deployed a pipeline and conducted extensive experiments on multiple benchmarks, achieving up to 2% improvement over other state-of-the-art continual learning methods.

Technologies

Language: Mandarin (native), English (CET-6 600, TOEFL under preparation)

Programming: C/C++, Python, Rust

Technologies: Git, Pytorch, Latex, vLLM, VeRL, LLaMA-Factory, Deepspeed

张子聪

上海市, 中国 | zhangzico@sjtu.edu.cn | +(86) 188 9181 2036

https://zhangzicong518.github.io

教育背景

上海交通大学, 计算机科学本科 2023 年 9 月–至今

- 约翰 · 霍普克罗夫特班，致远荣誉计划
- 平均绩点：89.2/100
- **核心课程：**
  数学：数理逻辑（97）、信息论（97）、线性代数（94）
  计算机：机器学习理论（98）、计算机系统设计（98）、程序设计与数据结构（96）、机器学习（95）、算法（94）、程序设计语言设计与实现（90）

科研经历

多模态推理与通用智能体, 微软亚洲研究院，实习生 2025 年 7 月–2026 年 1 月

- 提出并实现了一种融合工具的强化学习范式，将长链思维推理与工具调用统一为单一混合思维模式，使视觉–语言模型能够更有效地“借助图像思考”
- 构建并改进了大规模、高保真的仿真环境，可生成多样化且真实的智能体–工具交互轨迹，为鲁棒的工具调用智能体提供海量训练数据。

基于模型融合的持续学习, 上海交通大学 MIFA 实验室，实习生 2024 年 7 月–2025 年 4 月

- 共同主导开发了一种面向视觉–语言模型的全新持续学习框架，引入模型融合机制
- 提出在零样本场景下，通过聚合多个解耦的任务专用模型结果进行预测。
- 在多个基准上开展大量实验，表现超越原始预训练 VLM 及其他主流持续学习方法。

获奖情况

- 致远荣誉奖学金（上海交通大学前 10%）2023 年, 2024, 2025 年
- C 等学业奖学金（专业前 30%）2024, 2025 年

论文发表

Enhanced Continual Learning of Vision-Language Models with Model Fusion

Haoyuan Gao*, Zicong Zhang*, Yuqi Wei, Linglan Zhao, Guilin Li, Yexin Li, Linghe Kong, Weiran Huang

Fisrt Workshop SCOPE The Thirteenth International Conference on Learning Representations

Tracing the Dance of Embeddings: Visualizing Training Dynamics for Debugging Model Training Process

Yiming Liu*, Zicong Zhang*, Yun Lin, Ruofan Liu, Yuhuan Huang, Weiyu Kong

[Under Review]

项目经历

面向通用智能模型的环境扩展 –微软亚洲研究院 2025 年 7 月–2026 年 1 月

- 从 RapidAPI、MCP-Server 与 ToolBench 收集 15 万真实 API；将 API 信息嵌入，构建基于相似性的加权依赖工具图。
- 应用 Louvain 社区检测将图划分为数百个功能域，并按读写范式自动为每域生成统一的数据库模式。
- 将每个 API 实例化为域内确定性的读/写算子，使环境状态完全可验证
- 在依赖图上采样长程、逻辑一致的工具序列，生成多样化、高保真的智能体–环境交互轨迹，用于训练。

多模态推理的主动感知扩展 –微软亚洲研究院 2025 年 7 月–2026 年 1 月

- 通过监督微调与强化学习，将长链推理与工具调用统一为单一混合思维模式，使视觉–语言模型能在“think-with-image”时自主切换思维状态
- 通过教师蒸馏与迭代拒绝采样整理高质量长链思维轨迹，缓解工具使用的幻觉与低效问题
- 设计模式感知的过程奖励，鼓励模型以互补方式同时利用内部推理与外部工具
- 在 MathVision、MathVista、HRBench、V* 等数学或感知导向基准上评估，全面超越原始 VLM 基线。

视觉–语言模型持续学习框架 –上海交通大学 MIFA 实验室 2024 年 7 月–2025 年 1 月

成果见论文《基于模型融合的增强视觉–语言模型连续学习》

- 引入模型融合机制，为视觉–语言模型设计连续学习框架
- 搭建完整实验管线，在多个基准上开展大量实验，相比其他主流连续学习方法最高提升 2%。

技术技能

语言： 普通话（母语），英语（CET-6 600 分）

编程语言： C/C++、Python、Rust

相关技能： Git、PyTorch、LaTeX、vLLM、VeRL、LLaMA-Factory、Deepspeed