

孙博远

电话: 15253758970 邮箱: sbysbysby123@gmail.com

个人简介

南开大学媒体计算实验室在读四年级博士生，导师为侯淇彬教授和程明明教授。研究聚焦多模态基础模型，视频理解与长程 Agent，关注视觉指令跟随与高效表示、多模态理解、Agent 设计与轨迹数据构建，以及支持长程科学任务的 Agent Harness 构建与上下文管理。目前 Google Scholar 引用量 700+，GitHub 项目累计获 2000+ stars。

个人主页: <https://bbbbchan.github.io>

教育背景

- 南开大学，计算机科学与技术，硕博连读 2021.9 – 至今
- 西安电子科技大学，软件工程，学士 2017.9 – 2021.7

研究与实习经历

- 九坤投资-至知创新研究院-IQuest Team** 中国, 北京
Research Intern – AI for Science, Science Agent 2026.8 – 至今
 - AI for Science 基础模型**: 面向长程科学问题开展数据构建，将多步求解过程组织为 Agent 轨迹，实现长程问题的轨迹构建与模型训练；关注如何将科学问题求解过程转化为可用于能力学习的训练数据。
 - Science Agent Harness 构建**: 从 Agent 设计角度出发，关注面向科学问题的 Harness 组织和构建。包括 End-to-End 问题的后台执行，原子工具集合，异步管理，基于文件的可拓展上下文系统等内容
- 字节跳动** 中国, 北京
Research Intern – Multimodal Model Application, Agentic Streaming Video Understanding 2026.4 – 2026.8
 - 增量感知与记忆**: 针对持续视频输入导致上下文不断增长的问题，探索逐帧观测与结构化记忆结合的推理流程；以跨帧变化驱动状态更新，将在线信息积累与查询时的推理解耦，研究受控上下文预算下的长时视频理解。
 - 问题驱动的回访**: 结合记忆检索与历史片段重看，围绕用户问题回访相关视觉证据；研究如何以按需回看补充记忆压缩丢失的信息，使历史信息管理服务于当前问题，并分析信息保留与回访成本之间的取舍。
 - 多模态模型应用**: 参与流视频理解模型在国产 GPU 平台上的推理适配与部署，并面向部门多项下游业务开展多模态模型训练与适配，支持模型在不同任务场景中的应用。
- 上海人工智能实验室 - Agent A1 核心团队** 中国, 上海
Research Intern – MLE Agent, Science Discovery 2025.12 – 2026.4
 - 职责分解与上下文**: 针对算法优化 Agent 在编码迭代中混合承担数据分析、诊断与规划的问题，设计独立上下文的数据分析子 Agent；通过委托接口接收任务，以结构化报告返回诊断结论，隔离探索过程的中间输出，支持主 Agent 的后续决策。
 - 执行证据与诊断**: 以文件浏览、Shell 和代码执行为基础工具，主动探查数据及实验结果，要求报告中的数值关联实际计算；将反馈从单一指标扩展到数据与结果分析，为模型选择、实验规划及算法改进提供可追溯的依据。
 - 轨迹采集与训练**: 将分析轨迹与主流程的编码轨迹分开记录，通过模块接口解耦诊断与任务执行，为两类能力的独立数据采集及后续训练提供基础；参与长时机器学习算法优化系统的轨迹数据分析。
- 阿里巴巴 - 通义实验室** 中国, 杭州
Research Intern – Multimodal Video Understanding, Omni Model 2024.9 – 2025.12
 - 指令跟随与对齐**: 从对象名词与属性词的注意力差异出发，研究自然语言指代与具体视觉实例之间的对齐缺口；将对象 mask 转化为训练期的多层注意力监督，推理时保持自然语言交互，无需额外视觉提示或结构。进一步分析现有模型难以按自然语言关注目标的原因，包括语义偏执或过度的用户意图识别训练，并通过数据规模实验验证该监督策略的可扩展性。
 - 语义覆盖与压缩**: 针对基于注意力机制的显著性筛选可能遗漏非显著信息、重复保留相似内容的问题，在特征相似图上提取语义连通分量，分空间与时间两步合并 token；以语义覆盖与跨帧冗余组织免训练压缩，不要求语义相近信息在图像中物理相邻，关注有限预算下的信息保留，在保证语义覆盖的基础上实现极低 token 保留。
 - 条件化视觉表示**: 利用静态细节、时序变化与压缩表示的互补性，研究由用户指令控制的多 projector 自适应融合，使视觉信息组织适应不同问题；结合面部、身体及交互等不同粒度信息，参与人物中心表示与视听上下文理解研究。
 - 训练数据与评测**: 将区域标注转化为自然语言指代-视觉实例配对数据；结合模型初标与人工修正构建动态表情描述，通过事件语义匹配与顺序一致性评测；针对 Omni 模型设计人物建档的数据构造和多样化数据类型采集流程。

技术报告与开源项目

- [1] **Agents-A1**: Scaling the Horizon, Not the Parameters: Reaching Trillion-Parameter Performance with a 35B Agent. **技术报告** (HF Trending 上榜 3 周, 总下载量超 800K) .
参与长时科学发现 Agent 系统, 工作涉及通用 Science Agent 的数据分析设计与轨迹研究。
- [2] **SAIL**: Scientific Agentic Intelligence via a Science-Aware Loop. **技术报告**
面向科学发现 Agent, 负责科研文献调研与科学搜索相关研究, 参与长程任务轨迹构建、评测及执行框架设计。
- [3] **IQuest-Q1**. **IQuest 开源 320B 基座科学模型**
面向长程任务的智能体基础模型, 融合代码生成、推理与多步工具调用能力, 支持复杂任务的持续执行。
- [4] **MLEvolve**: A Self-Evolving Framework for Automated Machine Learning Algorithm Discovery. **技术报告**.
面向机器学习算法发现与迭代优化, 参与轨迹数据分析及数据分析 Agent 相关探索。
- [5] **HumanOmniV2**: From Understanding to Omni-Modal Reasoning with Context. **技术报告**.
面向基于强化学习的视听上下文的多模态推理, 参与视觉侧理解研究和 RL 策略设计。
- [6] **HumanOmni**: A Large Vision-Speech Language Model for Human-Centric Video Understanding. **技术报告**.
面向人物中心的视频理解, 参与面部、身体及交互信息的视觉表示与融合研究。
- [7] **cc-mini**: Ultra-light Harness Scaffolding for AI Agents. **开源项目**.
轻量级 AI Agent 执行框架, 极简 Claude Code 开源实现。

已发表工作 (第一/共一作者)

- [1] **Boyuan Sun**, Bowen Yin, Yuanming Li, Xihan Wei, Qibin Hou. “**See What I Mean**: Aligning Vision and Language Representations for Video Fine-grained Object Understanding.” **CVPR 2026**.
- [2] Jiaxing Zhao*, **Boyuan Sun***, Xiang Chen, Xihan Wei. “**Facial Dynamics in Video**: Instruction Tuning for Improved Facial Expression Perception and Contextual Awareness.” **AAAI 2026**.
- [3] **Boyuan Sun**, Yuqi Yang, Le Zhang, Ming-Ming Cheng, Qibin Hou. “**CorrMatch**: Label Propagation via Correlation Matching for Semi-Supervised Semantic Segmentation.” **CVPR 2024**.

在投工作与预印本 (第一/共一作者)

- [1] **Boyuan Sun**, Jiaxing Zhao*, Xihan Wei, Qibin Hou. “**LLaVA-Scissor**: Token Compression with Semantic Connected Components for Video LLMs.” **CVMJ 在投**
- [2] **Boyuan Sun**, Modi Jin*, Bowen Yin, Qibin Hou. “**Depth Anything** at Any Condition.” **TPAMI 在投**
- [3] **Boyuan Sun**, Jiaxing Zhao*, Xiang Chen, Xihan Wei, Qibin Hou. “**LLaVA-Octopus**: Unlocking Instruction-Driven Adaptive Projector Fusion for Video Understanding.”
- [4] **Boyuan Sun**, Shaoyong Jia, Shaohui Jiao, Qibin Hou, et. al. “**DeltaStreamer**: Agentic Streaming Video Understanding with Change-Aware Memory and Query-Driven Planning.” **ICLR 2027 在投**

系列工作

- [1] Modi Jin, Yiming Zhang, **Boyuan Sun**, Dingwen Zhang, Ming-Ming Cheng, Qibin Hou. “**GeoAgent**: Learning to Geolocate Everywhere with Reinforced Geographic Characteristics.” **CVPR 2026 Highlight**.
- [2] Zhong-Yu Li, Xin Jin, **Boyuan Sun**, Chun-Le Guo, Ming-Ming Cheng. “Towards **raw object detection** in diverse conditions” **CVPR 2025 Highlight**.
- [3] Bo-Wen Yin, Qize Yang, **Boyuan Sun**, Xihan Wei, Qibin Hou. “**Discriminative Visual Process Rewards** for Scaling Thinking at Test-Time with Images” **ICML 2026**.

荣誉奖项

- 国家奖学金 (2019); 南开大学一等奖学金 (2024); 西安电子科技大学一等奖学金 (2018)。
 - ACM/ICPC 全国邀请赛**铜奖**、四川省赛**银奖**、陕西省赛**铜奖**。
-