

Boyuan Sun

Phone: +86 15253758970 Email: sbysbysby123@gmail.com

Profile

Fourth-year Ph.D. student at the Media Computing Laboratory, Nankai University, advised by [Prof. Qibin Hou](#) and [Prof. Ming-Ming Cheng](#). Research focuses on **multimodal foundation models**, **video understanding**, and **long-horizon agents**, particularly visual grounding, efficient representations, trajectory data, and agent context management. **700+ citations** on [Google Scholar](#); **2,000+ stars** across [GitHub](#) projects.

Homepage: <https://bbbbchan.github.io>

Education

- **Nankai University**, Computer Science and Technology 2021.9 – Present
Integrated Master's–Ph.D. Program
- **Xidian University**, B.Eng. in Software Engineering 2017.9 – 2021.7

Research and Internship Experience

- **Ubiquant – IQuest Research, IQuest Team** Beijing, China
Research Intern – AI for Science, Science Agent 2026.8 – Present
 - **AI for Science foundation models**: Build long-horizon scientific problem-solving trajectories for model training, with a focus on literature review, scientific search, and deep research.
 - **Science agent harnesses**: Develop persistent background execution, atomic tools, and asynchronous task management. Explore file-based memory and retrieval to support scientific workflows beyond a single context window.
 - **ByteDance** Beijing, China
Research Intern – Multimodal Model Application, Agentic Streaming Video Understanding 2026.4 – 2026.8
 - **Incremental perception and memory**: Use frame-by-frame observations and change-driven structured memory to separate continuous perception from query-time reasoning, keeping context bounded as video streams grow.
 - **Query-driven revisiting**: Retrieve memory and revisit relevant video segments on demand to recover evidence lost during compression, balancing information retention against revisiting cost.
 - **Multimodal model applications**: Adapt and deploy streaming video models on domestic GPU platforms; train and adapt multimodal models for downstream business applications.
 - **Shanghai AI Laboratory – Agents-A1 Core Team** Shanghai, China
Research Intern – MLE Agent, Science Discovery 2025.12 – 2026.4
 - **Role decomposition and context isolation**: Design a data-analysis agent for MLE and extend it into an Agents-A1 subagent. Return structured findings while isolating exploratory outputs, preserving the main agent's context for long-horizon planning.
 - **Execution-grounded diagnosis**: Ground analysis in file inspection, shell commands, and executed code. Replace metric-only feedback with verifiable diagnoses that inform model selection, experiment planning, and algorithm improvement.
 - **Trajectory collection and training**: Separate analysis trajectories from coding trajectories through a delegated interface, enabling independent data collection and training for diagnosis and execution.
 - **Alibaba – Tongyi Laboratory** Hangzhou, China
Research Intern – Multimodal Video Understanding, Omni Model 2024.9 – 2025.12
 - **Instruction following and alignment**: Investigate why natural-language references fail to ground visual targets through noun-attribute attention differences. Use mask-supervised attention without extra inference inputs or architecture changes; data-scaling experiments support the approach's scalability.
 - **Semantic coverage and compression**: Replace saliency-only token selection with semantic connected components on feature-similarity graphs. Merge tokens across space and time to retain broad semantic coverage at low token budgets, without requiring spatial adjacency.
 - **Instruction-conditioned visual representations**: Adaptively fuse projectors for static details, temporal changes, and compressed representations according to user instructions; extend complementary representations to face, body, and interaction cues.
 - **Training data and evaluation**: Build reference–instance pairs, dynamic facial-expression annotations, and human-profile data for omni-modal models. Evaluate event semantics and temporal ordering to capture errors beyond overall caption similarity.
-

Technical Reports and Open-Source Projects

- [1] **Agents-A1**: Scaling the Horizon, Not the Parameters: Reaching Trillion-Parameter Performance with a 35B Agent. **Technical Report** (3 weeks on HF Trending; 800K+ downloads).
Designed data-analysis agents and studied trajectories for long-horizon scientific discovery.
- [2] **SAIL**: Scientific Agentic Intelligence via a Science-Aware Loop. **Technical Report**.
Responsible for literature review and scientific search; contributed trajectories, evaluation, and execution infrastructure.
- [3] **IQuest-Q1**. **IQuest 320B Open-Weight Foundation Model for Science**.
Combines code generation, reasoning, and multi-step tool use for long-horizon tasks.
- [4] **MLEvolve**: A Self-Evolving Framework for Automated Machine Learning Algorithm Discovery. **Technical Report**.
Contributed trajectory analysis and data-analysis agents for automated ML algorithm discovery.
- [5] **HumanOmniV2**: From Understanding to Omni-Modal Reasoning with Context. **Technical Report**.
Contributed visual understanding and RL strategies for context-grounded audiovisual reasoning.
- [6] **HumanOmni**: A Large Vision-Speech Language Model for Human-Centric Video Understanding. **Technical Report**.
Contributed face, body, and interaction representations for human-centric video understanding.
- [7] **cc-mini**: Ultra-light Harness Scaffolding for AI Agents. **Open Source**.
A minimal open-source runtime for Claude Code-style coding agents.

Published Papers (First / Co-first Author)

- [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**.

Submissions and Preprints (First / Co-first Author)

- [1] **Boyuan Sun***, Jiaxing Zhao*, Xihan Wei, Qibin Hou. “**LLaVA-Scissor**: Token Compression with Semantic Connected Components for Video LLMs.” **CVMJ, under review**
- [2] **Boyuan Sun***, Modi Jin*, Bowen Yin, Qibin Hou. “**Depth Anything** at Any Condition.” **TPAMI, under review**
- [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, under review**

Collaborative Publications

- [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**.

Honors and Awards

- **National Scholarship** (2019); First-Class Scholarships, Nankai University (2024) and Xidian University (2018).
 - ACM/ICPC: **Bronze**, National Invitational; **Silver**, Sichuan Provincial Contest; **Bronze**, Shaanxi Provincial Contest.
-