Shuangrui Ding

Ph.D. candidate, CUHK MMLab

Research Scientist Intern, Meta Superintelligence Labs

On the job market for Research Scientist roles

(+1) 925-953-4752 (US)
(+86) 13777842202 (China)
mark12ding@gmail.com
Homepage
GitHub
Google Scholar

Short Biography

Shuangrui Ding is a final-year Ph.D. candidate at CUHK MMLab, advised by Prof. Dahua Lin. His research centers on vision-language models and interactive multimodal agents. He has authored 10 papers as a first or co-first author, with publications appearing in top-tier conferences including ICLR, ICCV, CVPR, ECCV, NeurIPS, ICML, and ACL.

Google Scholar: 2,797 citations, h-index 20, i10-index 26 (snapshot: August 12, 2026).

Education

The Chinese University of Hong Kong, Department of Information Engineering — Hong Kong, China
Ph.D. in Information Engineering; Advisor: Prof. Dahua Lin — Sept 2023 – Now

Shanghai Jiao Tong University, Department of Electrical Engineering — Shanghai, China
M.S. in Information and Communication Engineering; National Scholarship Awardee — Sept 2021 – June 2023

University of Michigan, College of Engineering — Ann Arbor, Michigan
B.S.E. in Computer Science; Summa Cum Laude; GPA: 3.9/4.0 — Sept 2019 – May 2021

Shanghai Jiao Tong University, UM-SJTU Joint Institute — Shanghai, China
B.S.E. in Electrical and Computer Engineering; National Scholarship Awardee; GPA: 3.8/4.0 — Sept 2017 – Aug 2021

Industry Internship Experience

Meta Superintelligence Labs — Bellevue, Washington
Developed Vision-Language Models tailored for video grounding; supervised by Jie Lei — May 2026 – Oct 2026

Meta Superintelligence Labs — London, UK
Worked on multimodal interactivity and video grounding; supervised by Nicolas Carion — May 2025 – Oct 2025

Shanghai AI Laboratory — Shanghai, China
Focused on LLMs and advanced video understanding; supervised by Jiaqi Wang — May 2023 – May 2025

Huawei Inc. — Shanghai, China
Designed efficient and streamlined video foundation model; supervised by Xiaopeng Zhang — Aug 2022 – May 2023

Tencent AI Lab — Shenzhen, China
Designed self-supervised video representation learning; supervised by Maomao Li — Feb 2021 – Aug 2021

Selected Research Impact

SAM 3: Segment Anything with Concepts — 2025

- Co-developed multimodal interactivity and video grounding for SAM 3, a unified model for concept-level detection, segmentation, and tracking across images and videos; released at scale with 11.2K+ GitHub stars, 18M+ all-time Hugging Face downloads, and 880+ Google Scholar citations.

WildClawBench: A Benchmark for Real-World, Long-Horizon Agent Evaluation — 2026

- Project lead for a real-world, long-horizon AI agent benchmark with 60 human-authored bilingual multimodal tasks; used as an evaluation benchmark in the model release posts for Hunyuan Hy3, Seed2.1, and Muse Glimmer, and released with 500+ GitHub stars and 80K+ all-time Hugging Face downloads.

Interactive and Grounded Video Foundation Models — 2024 – 2026

- First-authored and led two widely adopted projects: SAM2Long for robust video memory (567 GitHub stars; 112 citations) and Dispider for proactive real-time Video-LLM interaction (124 citations), published at ICCV and CVPR.

Publications

* Equal contribution; † Project lead. Citation counts are organized from *GScholar* using the August 12, 2026 snapshot.

SAM, LLMs, and Agent Benchmarks

[1] Meta SAM 3 Team, incl. **Shuangrui Ding**. *SAM 3: Segment Anything with Concepts*. **ICLR 2026.** [Cited by 887]

[2] **Shuangrui Ding**[*†], Xuanlang Dai[*], Long Xing[*], et al. *WildClawBench: A Benchmark for Real-World, Long-Horizon Agent Evaluation*. **arXiv 2026.** [Cited by 32]

[3] **Shuangrui Ding**, Rui Qian, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Yuhang Cao, Yuwei Guo, Dahua Lin, Jiaqi Wang. *SAM2Long: Enhancing SAM 2 for Long Video Segmentation with a Training-Free Memory Tree*. **ICCV 2025.** [Cited by 112]

[4] **Shuangrui Ding**[*], Zihan Liu[*], Xiaoyi Dong, Pan Zhang, Rui Qian, Junhao Huang, Conghui He, Dahua Lin, Jiaqi Wang. *SongComposer: A Large Language Model for Lyric and Melody Generation in Song Composition*. **ACL 2025 Main.** [Cited by 98]

[5] Rui Qian[*], **Shuangrui Ding**[*], Xiaoyi Dong, Pan Zhang, Yuhang Zang, Yuhang Cao, Dahua Lin, Jiaqi Wang. *Dispider: Enabling Video LLMs with Active Real-Time Interaction via Disentangled Perception, Decision, and Reaction*. **CVPR 2025.** [Cited by 124]

[6] Zhixiong Zhang[*], **Shuangrui Ding**[*], Xiaoyi Dong, Songxin He, Jianfan Lin, Junsong Tang, Yuhang Zang, Yuhang Cao, Dahua Lin, Jiaqi Wang. *SeC: Advancing Complex Video Object Segmentation via Progressive Concept Construction*. **ICLR 2026.** [Cited by 20]

[7] Rui Qian, Xiaoyi Dong, Pan Zhang, Yuhang Zang, **Shuangrui Ding**, Dahua Lin, Jiaqi Wang. *Streaming Long Video Understanding with Large Language Models*. **NeurIPS 2024.** [Cited by 230]

[8] Long Xing et al., incl. **Shuangrui Ding**. *ScaleCap: Inference-Time Scalable Image Captioning via Dual-Modality Debiasing*. **ICLR 2026.** [Cited by 12]

[9] Junbo Niu et al., incl. **Shuangrui Ding**. *OVO-Bench: How Far is Your Video-LLMs from Real-World Online Video Understanding?*. **CVPR 2025.** [Cited by 118]

[10] Pan Zhang et al., incl. **Shuangrui Ding**. *InternLM-XComposer2.5-OmniLive: A Comprehensive Multimodal System for Long-Term Streaming Video and Audio Interactions*. **arXiv 2024.** [Cited by 51]

[11] Pan Zhang et al., incl. **Shuangrui Ding**. *InternLM-XComposer: A Vision-Language Large Model for Advanced Text-Image Comprehension and Composition*. **arXiv 2023.** [Cited by 375]

Lead / First-author Work

[12] **Shuangrui Ding**[*], Rui Qian[*], Haohang Xu, Dahua Lin, Hongkai Xiong. *Betrayed by Attention: A Simple yet Effective Approach for Self-supervised Video Object Segmentation*. **ECCV 2024.** [Cited by 20]

[13] **Shuangrui Ding**, Peisen Zhao, Xiaopeng Zhang, Rui Qian, Hongkai Xiong, Qi Tian. *Prune Spatio-temporal Tokens by Semantic-aware Temporal Accumulation*. **ICCV 2023.** [Cited by 44]

[14] **Shuangrui Ding**, Rui Qian, Hongkai Xiong. *Dual Contrastive Learning for Spatio-temporal Representation*. **ACM MM 2022.** [Cited by 31]

[15] **Shuangrui Ding**, Maomao Li, Tianyu Yang, Rui Qian, Haohang Xu, Qingyi Chen, Jue Wang, Hongkai Xiong. *Motion-aware Self-supervised Video Representation Learning via Foreground-background Merging*. **CVPR 2022.** [Cited by 95]

[16] Jiaqi Ma[*], **Shuangrui Ding**[*], Qiaozhu Mei. *Towards More Practical Adversarial Attacks on Graph Neural Networks*. **NeurIPS 2020.** [Cited by 210]

Selected Collaborations

[17] Zihan Liu, **Shuangrui Ding**, Zhixiong Zhang, Xiaoyi Dong, Pan Zhang, Yuhang Zang, Yuhang Cao, Dahua Lin, Jiaqi Wang. *SongGen: A Single Stage Auto-regressive Transformer for Text-to-Song Generation*. **ICML 2025.** [Cited by 50]

[18] Yuwei Guo, Ceyuan Yang, Anyi Rao, Chenlin Meng, Omer Bar-Tal, **Shuangrui Ding**, Maneesh Agrawala, Dahua Lin, Bo Dai. *Keyframe-guided Creative Video Inpainting*. **CVPR 2025.** [Cited by 14]

[19] Han Li, Shaohui Li, **Shuangrui Ding**, Wenrui Dai, Maida Cao, Chenglin Li, Junni Zou, Hongkai Xiong. *Image Compression for Machine and Human Vision with Spatial-Frequency Adaptation*. **ECCV 2024.** [Cited by 57]

[20] Rui Qian, **Shuangrui Ding**, Dahua Lin. *Rethinking Image-to-Video Adaptation: An Object-centric Perspective*. **ECCV 2024.** [Cited by 17]

[21] Li Ding, Wen Fei, Yuyang Huang, **Shuangrui Ding**, Wenrui Dai, Chenglin Li, Junni Zou, Hongkai Xiong. *AMPA: Adaptive Mixed Precision Allocation for Low-Bit Integer Training*. **ICML 2024.** [Cited by 12]

[22] Rui Qian, **Shuangrui Ding**, Xian Liu, Dahua Lin. *Semantics Meets Temporal Correspondence: Self-supervised Object-centric Learning in Videos*. **ICCV 2023.** [Cited by 34]

[23] Rui Qian, **Shuangrui Ding**, Xian Liu, Dahua Lin. *Static and Dynamic Concepts for Self-supervised Video Representation Learning*. **ECCV 2022.** [Cited by 41]

[24] Rui Qian, Yuxi Li, Huabin Liu, John See, **Shuangrui Ding**, Xian Liu, Dian Li, Weiyao Lin. *Enhancing Self-supervised Video Representation Learning via Multi-level Feature Optimization*. **ICCV 2021.** [Cited by 43]

Invited Talks

ICCV 2025 LSVOS Workshop | Keynote Speaker — Oct 2025
From Pixels to Meaning: Towards Reliable Video Object Segmentation across Frames — Honolulu, Hawaii

InternLM Community Open Mic | Invited Speaker — Mar 2024
When Songwriting Meets Large Models: SongComposer as an AI Musician — Online

Selected Honors

CUHK Vice-Chancellor's Ph.D. Scholarship (80,000 HKD) — Mar 2023
Graduate National Scholarship (Top 2%) — Sep 2022
Shanghai Outstanding Graduate (Top 5%) — May 2021
Finalist, Mathematical Contest in Modeling (Top 0.3%) — Apr 2019
Undergraduate National Scholarship (Top 2%) — Sep 2018

Skills & Academic Service

Research Areas: Vision-language models, interactive video grounding, multimodal agents, video object segmentation, self-supervised video representation learning

Programming: Python/PyTorch, C/C++, Java, HTML/CSS, SQL, MATLAB, LaTeX

Languages: Chinese (Native), English (Fluent; TOEFL 101)

Reviewer: CVPR, ICCV, ECCV, NeurIPS, ICLR, ICML, AAAI, ACL, ACM MM