

Shuangrui Ding

Ph.D. candidate, CUHK MMLab

Research Scientist Intern, Meta Superintelligence Labs

☎ (+1) 9259534752

✉ 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: 2,211 citations, h-index 19, i10-index 25 (snapshot: June 9, 2026).

Education

The Chinese University of Hong Kong, Department of Information Engineering

Ph.D. in Information Engineering; Advisor: Prof. Dahua Lin

Hong Kong, China

Sept 2023 – Now

Shanghai Jiao Tong University, Department of Electrical Engineering

M.S. in Information and Communication Engineering; National Scholarship Awardee

Shanghai, China

Sept 2021 – June 2023

University of Michigan, College of Engineering

B.S.E. in Computer Science; Summa Cum Laude; GPA: 3.9/4.0

Ann Arbor, Michigan

Sept 2019 – May 2021

Shanghai Jiao Tong University, UM-SJTU Joint Institute

B.S.E. in Electrical and Computer Engineering; National Scholarship Awardee; GPA: 3.8/4.0

Shanghai, China

Sept 2017 – Aug 2021

Industry Internship Experience

Meta Superintelligence Labs

Developed Vision-Language Models tailored for video grounding; supervised by Jie Lei

Bellevue, Washington

May 2026 – Oct 2026

Meta Superintelligence Labs

Worked on multimodal interactivity and video grounding; supervised by Nicolas Carion

London, UK

May 2025 – Oct 2025

Shanghai AI Laboratory

Focused on LLMs and advanced video understanding; supervised by Xiaoyi Dong

Shanghai, China

May 2023 – May 2025

Huawei Inc.

Designed efficient and streamlined video foundation model; supervised by Xiaopeng Zhang

Shanghai, China

Aug 2022 – May 2023

Tencent AI Lab

Designed self-supervised video representation learning; supervised by Maomao Li

Shenzhen, China

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 10K+ GitHub stars and 12M+ Hugging Face downloads.

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; adopted by Hunyuan Hy3-preview and released with 430+ GitHub stars and 31K+ Hugging Face downloads.

Interactive and Grounded Video Foundation Models

2024 – 2026

- First-authored and led a coherent line of VLM research spanning robust video memory (SAM2Long) and proactive real-time Video-LLM interaction (Dispider), with publications at top-tier venues including ICCV, CVPR, and ICLR.

Publications

* Equal contribution; † Project lead. Citation counts are organized from *GScholar* using the June 9, 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 500]
- [2] **Shuangrui Ding**^{*}†, Xuanlang Dai^{*}, Long Xing^{*}, et al. *WildClawBench: A Benchmark for Real-World, Long-Horizon Agent Evaluation*. **arXiv 2026**. [Cited by 12]
- [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 96]
- [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 91]
- [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 95]
- [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 16]
- [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 207]
- [8] Long Xing et al., incl. **Shuangrui Ding**. *ScaleCap: Inference-Time Scalable Image Captioning via Dual-Modality Debiasing*. **ICLR 2026**. [Cited by 9]
- [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 85]
- [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 46]
- [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 357]

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 19]
- [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 42]
- [14] **Shuangrui Ding**, Rui Qian, Hongkai Xiong. *Dual Contrastive Learning for Spatio-temporal Representation*. **ACM MM 2022**. [Cited by 29]
- [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 94]
- [16] Jiaqi Ma^{*}, **Shuangrui Ding**^{*}, Qiaozhu Mei. *Towards More Practical Adversarial Attacks on Graph Neural Networks*. **NeurIPS 2020**. [Cited by 203]

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 37]
- [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 52]
- [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 40]
- [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 40]

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, $\text{\LaTeX}$

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

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