Mingfei Han *Ph.D.*

https://scholar.google.com/citations?user=wJEoIXsAAAAJ&hl=en
Email: hmf282@gmail.com Homepage: https://mingfei.info/

Research Interests

- My research lies at the interface of computer vision and robotics. I currently investigate large vision–language models, summarising videos, analysing their hallucination behaviour and extending their use beyond embodied agents. Recent work spans video–language understanding, with a focus on long video understanding, video grounding tasks such as Referring Video Object Segmentation, alongside vision–language navigation and manipulation for robots. Prior projects addressed individual and group activity recognition and video object detection under both fully and weakly supervised regimes, while my master's thesis centred on moving-object detection and multi-object tracking.

Education

Ph.D. *2021-2025* University of Technology Sydney, Sydney, NSW, Australia
- Major: Computer Science
- Advisor: Prof. Xiaojun Chang, Dr. Mahesh Prakash
- Thesis: Evolving Video Analysis: From Object Perception to Holistic Understanding

M.Eng. *2016-2019* Chinese Academy of Sciences, Beijing, China
- Major: Computer Technology
- Thesis: Research on Object Detection and Tracking using Satellite Videos

B.Eng. *2012-2016* Nankai University, Tianjin, China
- Major: Computer Science
- Ranking: 11/79

Publications

[1]. RoomTour3D: Geometry-Aware Video-Instruction Tuning for Embodied Navigation, *CVPR 2025*

Mingfei Han, Liang Ma, Kamila Zhum., Ekaterina Radi., Jingyi Zhang, Xiaojun Chang, Xiaodan Liang, Ivan Laptev

[2]. Shot2Story: A New Benchmark for Comprehensive Understanding of Multi-shot Videos, *ICLR 2025*

Mingfei Han, Linjie Yang, Xiaojun Chang, Heng Wang

[3]. Think Before You Segment: An Object-aware Reasoning Agent for Referring Audio-Visual Segmentation, *AAAI 2026*

Jinxing Zhou, Yanghao Morty Zhou, **Mingfei Han**, Tong Wang, Xiaojun Chang, Hisham Cholakkal, Rao Muhammad Anwer

[4]. Token Painter: Training-Free Text-Guided Image Inpainting via Mask Autoregressive Models, *AAAI 2026*

Longtao Jiang, Jie Huang, **Mingfei Han**, Lei Chen, Yongqiang Yu, Feng Zhao, Xiaojun Chang, Zhihui Li

[5]. MALMM: Multi-Agent Large Language Models for Zero-Shot Robotics Manipulation, *IROS 2025*
Harsh Singh, Rocktim Jyoti Das, **Mingfei Han**, Preslav Nakov, Ivan Laptev

[6]. LongVLM: Efficient Long Video Understanding via Large Language Models, *ECCV 2024 (Oral)*
Yuetian Weng, **Mingfei Han**, Haoyu He, Xiaojun Chang, Bohan Zhuang

[7]. Video Recognition in Portrait Mode, *CVPR 2024*
Mingfei Han, Linjie Yang, Xiaojie Jin, Jiashi Feng, Xiaojun Chang, Heng Wang

[8]. HTML: Hybrid Temporal-scale Multimodal Learning Framework for Referring Video Object Segmentation, *ICCV 2023*
Mingfei Han, Yali Wang, Zhihui Li , Lina Yao, Xiaojun Chang and Yu Qiao

[9]. Mask Propagation for Efficient Video Semantic Segmentation, *NeurIPS 2023*
Yuetian Weng, **Mingfei Han**, Haoyu He, Mingjie Li, Lina Yao, Xiaojun Chang and Bohan Zhuang

[10]. Dual-AI: Dual-path Actor Interaction Learning for Group Activity Recognition, *CVPR 2022 (Oral)*
Mingfei Han, David Junhao Zhang, Yali Wang, Rui Yan, Lina Yao, Xiaojun Chang and Yu Qiao

[11]. An Efficient Spatio-Temporal Pyramid Transformer for Action Detection, *ECCV 2022*
Yuetian Weng, Zizheng Pan, **Mingfei Han**, Xiaojun Chang, Bohan Zhuang

[12]. Progressive Frame-Proposal Mining for Weakly Supervised Video Object Detection, *Trans. on Image Processing 2024*
Mingfei Han, Yali Wang, Mingjie Li, Xiaojun Chang, Yi Yang and Yu Qiao

[13]. Mining Inter-Video Proposal Relations for Video Object Detection, *ECCV 2020*
Mingfei Han, Yali Wang, Xiaojun Chang and Yu Qiao

[14]. Object Tracking in Satellite Videos by Improved Correlation Filters with Motion Estimations, *IEEE Transactions on Geoscience and Remote Sensing 2019*
Shiyu Xuan, Shengyang Li, **Mingfei Han**, et al.

Research Experience

2024.8– Ongoing   Department of Computer Vision, MBZUAI, Abu Dhabi, UAE
- Postdoctoral Associate I - Advised by Prof. Ivan Laptev
- Working on embodied perception and intelligence, including visual-and-language navigation and manipulation.
  (CVPR 2025, IROS 2025 and 2 in submission)

2024.1– 2024.8   Department of Computer Vision, MBZUAI, Abu Dhabi, UAE
- Research Associate - Advised by Prof. Xiaodan Liang
- Worked on web-video boosted vision-and-language navigation.
  (CVPR 2025)

2022.8– 2024.1   ByteDance Seed
- Research Intern - Mentored by Heng Wang, Linjie Yang, Xiaojie Jin
- Worked on vision–language understanding and generation, including large-scale multimodal dataset curation, and multimodal hallucination elimination.
  (CVPR 2024, ICLR 2025 and 2 in submission)

2021.2– 2024.8   ReLER Lab, AAII, University of Technology Sydney

- Ph.D. Candidate - Advised by Prof. Xiaojun Chang
- Worked on video-language models, individual and group activity recognition, and dense prediction in videos.
(CVPR 2022, ICCV 2023, NeurIPS 2023, CVPR 2024, ECCV 2024, ICLR 2025)

2019.10 – 2021.1   MMLab, SIAT, Chinese Academy of Sciences

- Advised by Prof. Yali Wang and Prof. Yu Qiao
- Worked on action recognition/object detection in videos, and its variants under limited supervision. (ECCV 2020, CVPR 2022 and TIP 2024)

2017.7 – 2019.7   CSU, Chinese Academy of Sciences

- Advised by Prof. Shengyang Li
- Worked on moving objects detection and tracking in satellite videos. (Trans. on GRS in 2019)

Patents

- 视频目标分割方法、装置、电子设备及存储介质. **韩鸣飞**; 王亚力; 乔宇. CN202310383303.5
- 一种视频目标检测方法、装置、设备以及存储介质. **韩鸣飞**; 王亚力; 乔宇. CN202010819564.3
- 一种基于互补时空信息建模的群体行为识别方法. **韩鸣飞**; 王亚力; 乔宇. CN202210342854.2
- 一种文本生成方法、装置、电子设备及存储介质. 韩鸣飞; 杨林杰; 王恒. CN202311695859.4
- TEXT GENERATION METHOD AND APPARATUS, ELECTRONIC DEVICE, AND STORAGE MEDIUM. **Mingfei Han**; Linjie Yang; Heng Wang. US2025191394A1

Honours & Awards

- **First Prize** in Social Navigation Track in RoboSense IROS Challenge (2025)
- **First Prize** in Challenge of Multimodal Robot Learning (IROS, Shanghai AI Lab) (2025)
- Data61 PhD Scholarship (**full** scholarship, 2021)
- **First Prize** in the TRECVID 2019 competition: Activities in Extended Video (2019)
- **First Prize** in the "Vehicle Tracking of Satellite Videos" (NSFC) (2018)
- CSU Excellent Master Scholarship (2018, top 3%)
- **Graduate with Honour** of Nankai University (2016, top 3%)
- National Encouragement Scholarship (2014/2015)