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 interests lie in computer vision and machine learning. Currently, I am focusing on large vision-language models and their application in robotics. I worked on video-language downstream tasks related to object prediction in videos, like Referring-VOS. Previously, I worked on individual and group activity recognition, and video object detection with full and limited supervision. During my Master's thesis, I worked on moving object detection and tracking.

Education

Ph.D.   *2023-2025*   University of Technology Sydney, Sydney, NSW, Australia
- Major: Computer Science
- Advisor: Prof. Xiaojun Chang, Dr. Mahesh Prakash
  2021-2023   Monash University, Melbourne, VIC, Australia
- Major: Computer Science
- Advisor: Prof. Xiaojun Chang, Dr. Mahesh Prakash

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]. MALMM: Multi-Agent Large Language Models for Zero-Shot Robotics Manipulation, *arxiv 2024*
Harsh Singh, Rocktim Jyoti Das, **Mingfei Han**, Preslav Nakov, Ivan Laptev

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

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

[6]. 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

[7]. Mask Propagation for Efficient Video Semantic Segmentation, *NeurIPS 2023*

Yuetian Weng, **Mingfei Han**, Haoyu He, Mingjie Li, Lina Yao, Xiaojun Chang and Bohan Zhuang

[8]. 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

[9]. An Efficient Spatio-Temporal Pyramid Transformer for Action Detection, *ECCV 2022*

Yuetian Weng, Zizheng Pan, **Mingfei Han**, Xiaojun Chang, Bohan Zhuang

[10]. Progressive Frame-Proposal Mining for Weakly Supervised Video Object Detection, *Trans. on Image Processing 2021*

Mingfei Han, Yali Wang, Mingjie Li, Xiaojun Chang, Yi Yang and Yu Qiao

[11]. Mining Inter-Video Proposal Relations for Video Object Detection, *ECCV 2020*

Mingfei Han, Yali Wang, Xiaojun Chang and Yu Qiao

[12]. 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– Onging   Department of Computer Vision, MBZUAI, Abu Dhabi, UAE
- Post-Doctoral Associate - Advised by Prof. Ivan Laptev
- Working on embodied perception and intelligence, including visual language navigation and manipulation.
(CVPR 2025 and 2 in submission)

2023.4– 2024.8   ReLER Lab, AAII, University of Technology Sydney
- Ph.D. Candidate - Advised by Prof. Xiaojun Chang
- Working on video-language models, individual and group activity recognition, and dense prediction in videos.
(CVPR 2022, ICCV 2023, NeurIPS 2023, CVPR 2024, ECCV2024, ICLR 2025)

2021.2 – 2023.4   Monash University
- Advised by Prof. Xiaojun Chang

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 2021)

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 and conference papers)

Honours & Awards

- 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" competition organized by NSFC (National Natural Science Foundation of China) (2018)
- CSU Excellent Master Scholarship (2018, top 3%)
- **Graduate with Honour** of Nankai University (2016, top 3%)
- National Encouragement Scholarship (2014/2015)