# Yunzhong Hou

B141, 115 North Rd, the Australian National University, Canberra, ACT 2601, Australia

Email: yunzhong.hou@anu.edu.au

Phone: +61 415310351

Webpage: [homepage] [scholar] [linkedin]

I am an early-career researcher specializing in deep learning and computer vision. Building on my previous work on 3D understanding, my current research focuses on AI videography and photography, embodied and LLM agents, and 3D generation.

# Education

Doctor of Philosophy, Engineering and Computer Science	March 2019 – December 2023	
Australian National University	Canberra, Australia	
Supervisors: Prof. Liang Zheng, Prof. Stephen Gould, Prof. Hongdong Li		
• Multi-view 3D understanding systems for object detection and tracking.		
• Explainable AI and visualization of learned knowledge in neural networks.		
Bachelor of Engineering, Electronic Engineering	August 2014 – June 2018	
Tsinghua University	Beijing, China	
• Efficient application of neural networks in distributed computing with edge	e-device cooperation.	
Experiences		
Australian National University	Canberra, Australia	
Postdoc Research Fellow, School of Computing	April 2023 – present	
Supervisors: Prof. Liang Zheng and Prof. Tom Gedeon		
· Controlling camera movements for AI videography and privacy-preserving	perceptions for robots.	
• Camera layout optimization for multi-view 3D understanding and camera of	control.	
• Supervising one Master-of-Philosophy as primary supervisor and two PhD students as co-supervisors.		

Amazon Web Services Applied Scientist Intern

*California, USA* July 2022 – October 2022

• Investigated multi-modal data augmentation for pre-training open-vocabulary object detectors.

# **Selected Publications**

### Topic: AI videography & Image Generation

**Yunzhong Hou**, Liang Zheng, and Philip Torr, "Learning Camera Movement Control from Real-World Drone Videos," *arXiv; 2024.* <u>Impact</u>: Controlling camera movements to capture the real world into videos. Xingjian Leng, Jaskirat Singh, **Yunzhong Hou**, Zhenchang Xing, Saining Xie, Liang Zheng, "Repa-e: Unlocking VAE for End-to-End Tuning with Latent Diffusion Transformers," *arXiv; 2025.* <u>Impact</u>: Unlocking VAEs for joint optimization with diffusion models using alignment loss with vision foundation models.

#### Topic: 3D multi-view detection and tracking

**Yunzhong Hou**, Liang Zheng, and Stephen Gould, "Multiview Pedestrian Detection with Feature Perspective Transformation," [CCF B] ECCV; 2020. **Impact**: The first end-to-end fully convolutional pipeline for multiview detection. Its codebase is widely adopted in the latest studies. *Cited 122 times*.

Yunzhong Hou and Liang Zheng, "Multiview Detection with Shadow Transformer (and View-Coherent Data Augmentation)," [CCF A] ACM MM; 2021. <u>Impact</u>: The first Transformer-based solution. Cited 53 times.
Yunzhong Hou, Zhongdao Wang, Shengjin Wang, and Liang Zheng, "Adaptive Affinity for Associations in Multi-Target Multi-Camera Tracking," [CCF A. Journal] IEEE Trans. on Image Processing; 2022. <u>Impact</u>: Pointed out the difference in task requirement between tracking and re-identification. Cited 39 times.
Yunzhong Hou, Stephen Gould, and Liang Zheng, "Learning to Select Camera Views: Efficient Multiview Understanding at Few Glances," [CCF A] CVPR; 2024. <u>Impact</u>: Early study on the impact of multi-camera layout for computer perception including classification and detection.

**Yunzhong Hou**, Xingjian Leng, Tom Gedeon, and Liang Zheng, "Optimizing Camera Configurations for Multi-View Pedestrian Detection," *arXiv*; 2023. <u>Impact</u>: Automatically search for the best multi-camera layout for multi-view detection and tracking.

#### Topic: Explainable AI

**Yunzhong Hou** and Liang Zheng, "Visualizing Adapted Knowledge in Domain Transfer," [CCF A] *CVPR*; 2021. **Impact**: The first attempt at explaining and visualizing the domain adaptation process. Also serves as a strong baseline for source-free domain adaptation. *Cited 115 times*.

**Yunzhong Hou**, Liang Zheng, and Stephen Gould, "Learning to Structure an Image with Few Colors," [CCF] <u>A]</u> *CVPR*; 2020. <u>Impact</u>: Studied how color and structure contribute to the perception of convolutional neural networks. The results can be used for color quantization and for creating pixel art. *Cited 29 times*.

**Yunzhong Hou**, Stephen Gould, and Liang Zheng, "Scalable Deep Color Quantization: a Cluster Imitation Approach," [CCF A, Journal] *IEEE Trans. on Image Processing; 2024.* <u>Impact</u>: A color quantization neural network that can cluster the colors and create pixel and knitting art from existing images.

#### Others

**Yunzhong Hou**, Stephen Gould, and Liang Zheng, "View-coherent correlation consistency for semisupervised semantic segmentation," [CCF B, Journal] *Pattern Recognition; 2024.* <u>Impact</u>: Applied findings from multiple camera (physical) views to multiple augmented (virtual) views to semi-supervised semantic segmentation.

Xiaoxiao Sun, **Yunzhong Hou**, Weijian Deng, Hongdong Li, and Liang Zheng, "Ranking Models in Unlabeled New Environments," [CCF A] ICCV; 2021. <u>Impact</u>: Model generalization study without labels.

# **Research Grants and Awards**

Research Grants		
Efficient Video Dynamics Distillation for Scalable Models ( <u>1M GPU hours, Co-Investigator</u> )	2025	
<ul> <li>Investigated motion knowledge distillation from specialist models to LLMs.</li> </ul>		
OpenAI Research Credits (5,000 USD, Personal)		
<ul> <li>Awarded for research on AI videography powered by LLMs.</li> </ul>		
Google Cloud Research Credits (5,000 USD, Personal)	2024	
<ul> <li>Awarded for research on AI videography with foundation models and LLMs.</li> </ul>		
Socioeconomic impact of water reforms in Murray-Darling Basin (79,497 AUD, Co-Investigator)	2024	
• Developed algorithms to investigate the social-economic impact of water-related policies.		
• Collaborated with industry partners at Marsden Jacob Associates and led a research assistant.		
Privacy-preserving perception for robotics (25,000 AUD, Lead-Investigator)	2024	
• Project lead and assembled a team with other Co-Investigators from different disciplines, two Lecturers		
from Computer Science and one post-doc from the School of Philosophy.		
<ul> <li>Collaborated with other investigators and jointly supervised two master students.</li> </ul>		
Towards building general-purpose multimodal foundation models ( <u>1M GPU hours, Co-Investigator</u> )	2023	
<ul> <li>Investigated the language modality for instructing the computer vision models.</li> </ul>		
Awards		
CVPR outstanding reviewer	2025	
Australian Al Awards (Al innovator, ANU TechLauncher)	2024	
iAwards 2024 ACT Merit Receipt (Student & Education category, ANU TechLauncher)	2024	
<i>CVPR outstanding reviewer</i> 2		

# Teaching

 TechLauncher (Class of 300 students, Examiner for 70 and 30 students over two semesters)
 2024

 COMP3500/ COMP3550/ COMP4500/ COMP8715
 Australian National University

- A project-based learning experience where groups of students work with their clients from industry, business and government organizations to develop, prototype and launch real solutions.
- Worked as the examiner for three audits and evaluated the students' contributions based on their codebase and outputs, decision-making process, teamwork, and collaboration with clients.

# **Academic Services**

Convenor: AI+ML+Friends seminar

- Organize weekly seminars on AI and ML related topics and oral presentations for PhD completion. *Session chair*: ANU Intelligent System HDR Retreat 2024
- Session chair: ANU Intelligent System HDR Retreat 2024
- Hosted the annual retreat and managed the research talks.

Session chair: ACM MM 2024 Oral session 13 - Machine Learning for Multimedia

• Hosted the research talks.

*Conference reviewer*: CVPR, ICCV, ECCV, NeurIPS, ICLR, ICML, ACM MM, AAAI, IJCAI, etc. *Journal reviewer*: Nature Communications, IEEE TPAMI, IEEE TIP, IEEE TMM, IEEE TCSVT, PR, etc. *Area chair*:

- ICASSP 2024 and 2025: Area chair on self-supervised and semi-supervised learning / explainable and interpretable machine learning. Handled 47 and 20 papers, respectively.
- ACM MM 2024 and 2025: Area chair on multi-modal fusion and user engagement study. Handled 13 and 15 papers, respectively. Hosted the Oral Session on Machine Learning for Multimedia in Melbourne 2024. *Organizing team*:
  - ACM Web Conference MORE'25 Multimedia Object Re-ID Workshop: Handled 12 papers as the Program Chair and coordinated 23 participant teams. Hosted the workshop session in Sydney.

# **Invited Talks**

22 January 2025, Embassy of the People's Republic of China in the Commonwealth of Australia, ACT and Tsinghua Student Forum, *Learning Camera Movement Control from Real-World Drone Videos*. 28 October 2024, ACM MM 2024, Area Chair Workshop, *Learning Camera Controls from Videos: a Pilot Study on Drone Videography*.

19 July 2024, University of Canberra, International Research Workshop Data Science and AI & Robotics (DSAIR24), *Learning Camera Controls from Videos: a Pilot Study on Drone Videography.* 29 March 2021, ZhiDongXi open seminar, *Multi-View 3D Object Detection And Tracking.*