Yun Wang

x

Yun Wang

Shanghai, China ✉ ywang3875-c@my.cityu.edu.hk G Google Schloar ⌂ Yun Wang cocowy1

Reasearch Topic

3D Vision, Multi-modal Vision Language Models

Education

PH.D **City University of Hong Kong (QS Ranking: 63)**, Computer Science *Kowloon, HK SAR*
Aug.2023 - 2027
- GPA: 3.5
- Supervisor: Prof. **Dapeng Wu** ↗, IEEE Fellow

M.S **Sun Yat-sen University (985)**, Information and Communication Engineering *Guangzhou, China*
Aug.2020 - Jun.2023
- GPA: 88.03 (top 10%)
- Supervisor: Prof. **Yulan Guo** ↗, the National Science Fund for Excellent Young Scholars

B.E **China University of Geosciences (211)**, Electronic Information Engineering *Wuhan, China*
Aug.2016 - Jun.2020
- GPA: 3.75/4.0 (1/116, top 1%)

Experience

Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS) *Shenzhen, China*
Research Internship for 3D Vision *Aug. 2024 – May. 2025*
- Develop a self-supervised depth estimation algorithm for autonomous driving, achieving high estimation accuracy comparable to supervised methods without relying on ground-truth labels. **AAAI 2025 Oral**.
- Develop a robust depth estimation algorithm by adapting Vision Foundation Models for 3D Reconstruction, achieving state-of-the-art generalization ability compared to counterparts. **ICCV 2025 Highlight**.
- Develop an efficient and highly accurate dynamic scene 3D reconstruction MVS method, achieving state-of-the-art performance compared to counterparts. **NeuralPS 2025, Pre-rebuttal: 554**.

The Institute of Artificial Intelligence, China Telecom (TeleAI) Technology *ShangHai, China*
Research Internship for Multi-modal Video Understanding *May. 2025 – Sep. 2025*
- We present a training-free, plug-and-play framework that boosts long-form video understanding in Video-LLMs by encoding events into episodic memory and using a chain-of-thought strategy to select an informative minimal subset for advanced spatio-temporal reasoning. **AAAI 2026, Under Review**.
- We present a comprehensive survey of training-free methods for long-form video understanding, examining existing approaches, benchmarking datasets, and key challenges in the field. **AI Flow, Under Review**.

Yun Wang

Shanghai, China   ✉ ywang3875-c@my.cityu.edu.hk   G Google Schloar   ⌂ Yun Wang   cocowy1

Reasearch Topic

3D Vision, Multi-modal Vision Language Models

Education

PH.D **City University of Hong Kong (QS Ranking: 63)**, Computer Science — *Kowloon, HK SAR, Aug.2023 - 2027*
- GPA: 3.5
- Supervisor: Prof. **Dapeng Wu** ↗, IEEE Fellow

M.S **Sun Yat-sen University (985)**, Information and Communication Engineering — *Guangzhou, China, Aug.2020 - Jun.2023*
- GPA: 88.03 (top 10%)
- Supervisor: Prof. **Yulan Guo** ↗, the National Science Fund for Excellent Young Scholars

B.E **China University of Geosciences (211)**, Electronic Information Engineering — *Wuhan, China, Aug.2016 - Jun.2020*
- GPA: 3.75/4.0 (1/116, top 1%)

Experience

Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS) — *Shenzhen, China, Aug. 2024 – May. 2025*
Research Internship for 3D Vision
- Develop a self-supervised depth estimation algorithm for autonomous driving, achieving high estimation accuracy comparable to supervised methods without relying on ground-truth labels. **AAAI 2025 Oral**.
- Develop a robust depth estimation algorithm by adapting Vision Foundation Models for 3D Reconstruction, achieving state-of-the-art generalization ability compared to counterparts. **ICCV 2025 Highlight**.
- Develop an efficient and highly accurate dynamic scene 3D reconstruction MVS method, achieving state-of-the-art performance compared to counterparts. **NeuralPS 2025, Pre-rebuttal: 554**.

The Institute of Artificial Intelligence, China Telecom (TeleAI) Technology — *ShangHai, China, May. 2025 – Sep. 2025*
Research Internship for Multi-modal Video Understanding
- We present a training-free, plug-and-play framework that boosts long-form video understanding in Video-LLMs by encoding events into episodic memory and using a chain-of-thought strategy to select an informative minimal subset for advanced spatio-temporal reasoning. **AAAI 2026, Under Review**.
- We present a comprehensive survey of training-free methods for long-form video understanding, examining existing approaches, benchmarking datasets, and key challenges in the field. **AI Flow, Under Review**.

Yun Wang

Shanghai, China   ✉ ywang3875-c@my.cityu.edu.hk   G Google Schloar   ⌂ Yun Wang   cocowy1

Reasearch Topic

3D Vision, Multi-modal Vision Language Models

Education

PH.D **City University of Hong Kong (QS Ranking: 63)**, Computer Science — *Kowloon, HK SAR, Aug.2023 - 2027*
- GPA: 3.5
- Supervisor: Prof. **Dapeng Wu** ↗, IEEE Fellow

M.S **Sun Yat-sen University (985)**, Information and Communication Engineering — *Guangzhou, China, Aug.2020 - Jun.2023*
- GPA: 88.03 (top 10%)
- Supervisor: Prof. **Yulan Guo** ↗, the National Science Fund for Excellent Young Scholars

B.E **China University of Geosciences (211)**, Electronic Information Engineering — *Wuhan, China, Aug.2016 - Jun.2020*
- GPA: 3.75/4.0 (1/116, top 1%)

Experience

Shenzhen Institute of Artificial Intelligence and Robotics for Society (AIRS) — *Shenzhen, China, Aug. 2024 – May. 2025*
Research Internship for 3D Vision
- Develop a self-supervised depth estimation algorithm for autonomous driving, achieving high estimation accuracy comparable to supervised methods without relying on ground-truth labels. **AAAI 2025 Oral**.
- Develop a robust depth estimation algorithm by adapting Vision Foundation Models for 3D Reconstruction, achieving state-of-the-art generalization ability compared to counterparts. **ICCV 2025 Highlight**.
- Develop an efficient and highly accurate dynamic scene 3D reconstruction MVS method, achieving state-of-the-art performance compared to counterparts. **NeuralPS 2025, Pre-rebuttal: 554**.

The Institute of Artificial Intelligence, China Telecom (TeleAI) Technology — *ShangHai, China, May. 2025 – Sep. 2025*
Research Internship for Multi-modal Video Understanding
- We present a training-free, plug-and-play framework that boosts long-form video understanding in Video-LLMs by encoding events into episodic memory and using a chain-of-thought strategy to select an informative minimal subset for advanced spatio-temporal reasoning. **AAAI 2026, Under Review**.
- We present a comprehensive survey of training-free methods for long-form video understanding, examining existing approaches, benchmarking datasets, and key challenges in the field. **AI Flow, Under Review**.

Selected Publications

3D Vision Models

- **Wang, Yun** and Wang, Longguang, and Kunhong, Li, and Wang, Zi, and Dapeng Oliver Wu, and Guo, Yulan*. Revisiting Cost Aggregation in Stereo Matching from Disparity Classification. **IEEE Trans on Image Processing (SCI Top Journal, CCF-A), 2024**.
- **Wang, Yun** and Li, Kunhong, and Wang, Longguang, and Hu, Junjie, and Dapeng Oliver Wu, and Guo, Yulan*. AD-Stereo: Learning Stereo Matching with Adaptive Downsampling and Disparity Alignment. **IEEE Trans on Image Processing (SCI Top Journal, CCF-A), 2025**.
- **Wang, Yun** and Zheng, Jiahao, and Zhang Chenghao, and Zhang Zhanjie, and Li Kunhong, and Zhang Yongjian, and Hu Junjie*. DualNet: Self-supervised Learning for Stereo Matching with Pseudo-label Supervision. **AAAI 2025 Oral (CCF-A, Acceptance Rate: 4.6%)**
- Zhang, Yongjian, and Wang, Longguang, and Li, Kunhong, and **Wang, Yun**, and Guo, Yulan*. Learning representations from foundation models for domain generalized stereo matching. **ECCV 2024 (CCF-B, Acceptance Rate: 27.4%)**.
- **Wang, Yun**, and Wang, Longguang, and Zhang, Chenghao, and Zhang, Yongjian, and Zheng, Jiahao, and Zhang, Zhanjie, and Fan, Chenyou, and Lam, Tin Lun, and Hu, Junjie*. SMoEStereo: Selective Mixture of Experts with Parameter-Efficient Fine-tuning for Robust Stereo Matching. **ICCV 2025 Highlight (CCF-A, Acceptance Rate:10%)**.

Multi-modal Vision Language Tasks

- Wang, Bimei, and Dang, Jishen, and **Wang, Yun**, and Chen, Zixuan, and Li, Jiyuan, and Wang, Teng, and Yang, Jun*. Instruction-aware Memory Network for Video Recognition. **ICME 2025 (CCF-B, Acceptance Rate: 30%)**.
- Ma, Ao, and Feng, Jiasong, and Cao, Ke, and **Wang, Yun**, and Zhang, Quanwei, and Zhang, zhanjie*. Lay2Story: Extending Diffusion Transformers for Layout-Togglable Story Generation **ICCV 2025 (CCF-A, Acceptance Rate: 24%)**.
- **Wang, Yun**, and Zhang, Long, and Liu, Jingren, and Yan, Jiaqi, and Zhang, Zhanjie, and Zheng, Jiahao, and Yang, Xun, and Wu, Dapeng, and Chen, Xiangyu*, and Li, Xuelong*. Episodic Memory Representation for Long Video Understanding **AAAI 2025, Under Review**.
- Liu, Jingren[†], and **Wang, Yun**[†], and Zhang, Long, and Wang, Yiheng, and Xu, Shuning, and Wang, Ling, and Yan, Jiaqi, and Zhang, Dell, and Chen, Xiangyu*. Towards Training-Free Long Video Understanding: Methods, Benchmarks, and Open Challenges. **AI Flow, Under Review**.

Skills & Language Abilities

Technologies:

- Deep Learning: Pytorch
- Basic Programing: Python, C/C++, javascript

Languages:

- Chinese: Native Speaker
- English: IELTS 6.5 (Reading: 9.0)

Summary

I am a 2nd Ph.D. student in the Computer Science Department of City University of Hong Kong, advised by Prof. **Dapeng Wu** ⧉, and my master's degree is obtained in Electronic Information Engineering from SYSU, which was advised by Prof. **Yulan Guo** ⧉. My research interests lie in 3D computer vision and Multi-modal Large Models, focusing on related fields such as 3D reconstruction and generation, and Multi-modal Vision Language Models.