Yun Wang

Shenzhen, China ywang3875-c@my.cityu.edu.hk 15827391198 Google Schloar cocowy

Reasearch Topic

3D Vision, Depth Estimation, Style Transfer, 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

- 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, Published**.
- 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, Published**.
- Develop an efficient and highly accurate dynamic scene 3D reconstruction MVS method, achieving state-of-the-art performance compared to counterparts. **NeuralIPS 2025, Under Review**.

Publications

Lightweight Depth Estimation Model Design

- **Wang, Yun** and Wang, Longguang, and Wang, Hanyun and Guo, Yulan*. SPNet: Learning Stereo Matching with Slanted Plane Aggregation. **IEEE Robotics and Automation Letters (JCR Q1), 2022**
- Guo, Yulan*, **Wang, Yun** and Wang, Longguang and Wang, Zi. CVCNet: Learning Cost Volume Compression for Efficient Stereo Matching. **IEEE Trans on Multimedia (SCI Top Journal, CCF-B), 2022**.
- **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*. ADStereo: Learning Stereo Matching with Adaptive Downsampling and Disparity Alignment. **IEEE Trans on Image Processing (SCI Top Journal, CCF-A), 2025**.

Self-supervised Model Desgin

- **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%)**

- **Wang, Yun** and Yang, Zhengjie, and Zhang, Chenghao, and Dapeng Oliver Wu*, and Guo, Yulan. SMFormer: Empowering Self-supervised Stereo via Foundation Models and Contrastive Learning. **IEEE Trans on Image Processing (SCI Top Journal, CCF-A, Under Review)**.
- **Wang, Yun** and Yang, Zhengjie, and Zhang, Chenghao, and Hou, Junhui, and Dapeng Oliver Wu*. RoSe: Robust Self-supervised Stereo Matching under Challenging Weather Conditions. **IEEE Trans on Circuits and Systems for Video Technology (SCI Top Journal, CCF-B, Under Review)**.

Robust Model Desgin

- 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 (CCF-A, Acceptance Rate:24%)**.

Dynamic 3D Construction

- **Wang, Yun**, and Dong, Qiaole, and Zheng, Jiahao, and Zhang, Zhanjie, and Dapeng Oliver Wu*, and Hu, Junjie. MemStereo: Consistent Dynamic Stereo Matching with Dynamic Memory Aggregation. **NeuralIPS 2025 (CCF-A, Preparation)**.

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%)**.

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 Self-supervised Learning, Lightweight design, Vision Foundation Models, and multi-modal vision language models.