Zhenlong Yuan

yuanzhenlong21b@ict.ac.cn | zhenlongyuan.github.io | github.com/ZhenlongYuan | Google Scholar

EDUCATION

Institute of Computing Technology, Chinese Academy of Sciences (ICT, UCAS) Sep 2021 - Jun 2026

- Degree: Ph D. | Advisor: Zhaoqi Wang | Major: Computer Application Technology | GPA: 3.77 / 4.0
- Specialization: Vision-Language Model (VLM) | Large Language Model (LLM) | Embodied Agents | 3DV

Beijing University of Posts and Telecommunications (BUPT) Sep 2017 - Jun 2021

- Degree: Bachelor | Major: Telecommunications Engineering and Management | Rank: 2 / 158 | GPA: 3.83 / 4.0

ACADEMIC PAPERS

1. **Yuan, Z.**, Qu, X., et al. Video-STAR: Reinforcing Zero-shot Video Understanding with Tools. *International Conference on Learning Representations (ICLR 2026)*. **(Under Review)**
2. **Yuan, Z.**, Tang, J., et al. AutoDrive-R^2: Incentivizing Reasoning and Self-Reflection Capacity for VLA Model in Autonomous Driving. *International Conference on Learning Representations (ICLR 2026)*. **(Under Review)**
3. Qu, X.*, **Yuan, Z.***, et al. From Scale to Speed: Adaptive Test-Time Scaling for Image Editing. *International Conference on Learning Representations (ICLR 2026)*. **(Under Review)**
4. Nan, S.*, **Yuan, Z.***, et al. Reflection from Retrieval: MLLM-Guided Iterative Reasoning for Zero-Shot Composed Image Retrieval. *International Conference on Learning Representations (ICLR 2026)*. **(Under Review)**
5. **Yuan, Z.**, Luo, J., et al. DVP-MVS: Synergize depth-edge and visibility prior for multi-view stereo. *In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2025)*. **(Accepted)**
6. **Yuan, Z.**, Liu, C., et al. MSP-MVS: Multi-Granularity Segmentation Prior Guided Multi-View Stereo. *In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2025)*. **(Accepted)**
7. **Yuan, Z.**, Cao, J., et al. SD-MVS: Segmentation-driven Deformation Multi-View Stereo with Spherical Refinement and EM optimization. *In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2024)*. **(Accepted)**
8. **Yuan, Z.**, Zhang, D., et al. DVP-MVS++: Synergize Depth-Normal-Edge and Harmonized Visibility Prior for Multi-View Stereo. *(IEEE TCSVT 2025)*. **(Accepted, IF=11.1)**
9. **Yuan, Z.**, Yang, Z., et al. SED-MVS: Segmentation-Driven and Edge-Aligned Deformation Multi-View Stereo with Depth Restoration and Occlusion Constraint. *(IEEE TCSVT 2025)*. **(Accepted, IF=11.1)**
10. **Yuan, Z.**, Cao, J., et al. TSAR-MVS: Textureless-aware Segmentation and Correlative Refinement Guided Multi-View Stereo. *Pattern Recognition (PR 2024)*. **(Accepted, IF=7.6)**
11. Zhang, D.,**Yuan, Z.**, et al. ADDI: A Simplified E2E Autonomous Driving Model with Distinct Experts and Implicit Interactions. *International Conference on Learning Representations (ICLR 2026)*. **(Under Review)**
12. Chen, K.,**Yuan, Z.**, et al. Dual-Level Precision Edges Guided Multi-View Stereo with Accurate Planarization. *In Proceedings of the AAAI Conference on Artificial Intelligence (AAAI 2025)*. **(Accepted)**
13. Liu, C.,**Yuan, Z.**, et al. MR-IntelliAssist: A World Cognition Agent Enabling Adaptive Human-AI Symbiosis in Industry 4.0. *International Conference on Human-Computer Interaction (HCII 2025)*. **(Accepted)**
14. Chen, K.,**Yuan, Z.**, et al. Learning Multi-View Stereo with Geometry-Aware Prior. *IEEE Transactions on Circuits and Systems for Video Technology (IEEE TCSVT 2025)*. **(Accepted, IF=11.1)**
15. Cao, J.,**Yuan, Z.**, et al. NPMVS: NeRF-based Polarimetric MVS. *Pattern Recognition (PR 2025)*. **(Accepted, IF=7.6)**
16. Luo J., Ren W., Zheng Q., Zhao Y., **Yuan, Z.**, et al. InstructHOI: Context-Aware Instruction for Multi-Modal Reasoning in Human-Object Interaction Detection. *(NeurIPS 2025)*. **(Accepted, Spotlight)**

INTERNSHIP EXPERIENCE

AMAP, Alibaba ML R&D Department Jun 2025 - Now

- Position: Research Intern | Advisor: Xiangxiang Chu, Lei Sun | Specialization: Tool-Augmented RL
- We propose **Video-STAR**, a novel reasoning framework which leveage **visual toolbox (Agentic RL)** to generate **multimodal CoT** for robust open-vocabulary action recognition.
- We propose **AutoDrive-R²**, a novel VLA framework that enhances both reasoning and self-reflection capabilities of autonomous driving systems through **chain-of-thought (CoT)** processing and **reinforcement learning (RL)**.

REVIWERS & AWARDS

- **Conference Reviewers**: NIPS, ICLR, ICML, CVPR, ICCV, ECCV, AAAI | **Journal Reviewers**: TIP, TCSVT, TMM, PR
- **Awards:** ICT Lenovo Enterprise Scholarship **(Top 3%)** | ICT National Scholarships **(Top 5%)**

SKILLS

Programming Python, C/C++, CUDA, Matlab | **Technologies** Linux, Tensorflow, Pytorch, OpenCV