Yanyuan Qiao

Personal Website · Google Scholar · Email

Bio

Yanyuan Qiao is a postdoctoral research fellow at Swiss Federal Technology Institute of Lausanne (EPFL). She has been working on multimodal learning tasks, such as Vision-and-Language Navigation, Text-to-Image Synthesis, and Visual Question Answering, and has published 30+ peer-reviewed papers in top journals and conferences such as TPAMI, TMLR, NeurIPS, ICLR, CVPR, ICCV, ECCV, ICRA, IROS, EMNLP and ACM MM, etc. She has been recognized as an ICRA 2025 Best Paper Award Finalist and was selected for the ICCV Doctoral Consortium 2023. Currently, she is focusing on the research of Embodied AI.

Research Interest

Vision and Language, Embodied AI, AI-Driven Robotics Design.

Work Experience

Swiss Federal Institute of Technology Lausanne, Lausanne, Switzerland   *Jul. 2025 – Present*
Postdoctoral Research Fellow

Australian Institute For Machine Learning, Adelaide, Australia   *Jul. 2023 – Jul. 2025*
Postdoctoral Research Fellow

Education

The University of Adelaide, Adelaide, Australia   *Sep. 2020 – Dec. 2023*
PhD in Computer Science

University of Chinese Academy of Sciences, Beijing, China   *Sep. 2016 – Jun. 2019*
MEng in Electronics and Communication Engineering

Southeast University, Nanjing, China   *Aug. 2012 – Jun. 2016*
BEng in Sensor Network Technology

Publications

Journals

[1] Siqi Zhang, **Yanyuan Qiao**, Qunbo Wang, Longteng Guo, Zhihua Wei, Jing Liu. FlexVLN: Flexible Adaptation for Diverse Vision-and-Language Navigation Tasks. *IEEE Transactions on Multimedia (TMM)*, 2025.

[2] Yue Zhang, Ziqiao Ma, Jialu Li, **Yanyuan Qiao**, Zun Wang, Joyce Chai, Qi Wu, Mohit Bansal, Parisa Kordjamshidi. Vision-and-Language Navigation Today and Tomorrow: A Survey in the Era of Foundation Models. *Transactions on Machine Learning Research (TMLR)*, 2024.

[3] **Yanyuan Qiao**, Yuankai Qi, Yicong Hong, Zheng Yu, Peng Wang, Qi Wu. HOP+: History-enhanced and Order-aware Pre-training for Vision-and-Language Navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, 2023.

[4] **Yanyuan Qiao**, Chaorui Deng, Qi Wu. Referring Expression Comprehension: A Survey of Methods and Datasets. *IEEE Transactions on Multimedia (TMM)*, 2020.

[5] Zhiwei Fang, Jing Liu, Yong Li, **Yanyuan Qiao**, Hanqing Lu. Improving Visual Question Answering Using Dropout and Enhanced Question Encoder. *Pattern Recognition (PR)*, 2019.

Conferences

[1] **Yanyuan Qiao**, Haodong Hong, Wenqi Lyu, Dong An, Siqi Zhang, Yutong Xie, Xinyu Wang, Qi Wu. NavBench: Probing Multimodal Large Language Models for Embodied Navigation. *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2025.

[2] Wenbo Zhang, Tianrun Hu, **Yanyuan Qiao**, Hanbo Zhang, Yuchu Qin, Yang Li, Jiajun Liu, Tao Kong, Lingqiao Liu, Xiao Ma. Chain-of-Action: Trajectory Autoregressive Modeling for Robotic Manipulation. *Annual Conference on Neural Information Processing Systems (NeurIPS)*, 2025.

[3] Xiao Wu, Ting-Zhu Huang, Liang-Jian Deng, Yanyuan Qiao, Imran Razzak, Yutong Xie. A Knowledge-driven Adaptive Collaboration of LLMs for Enhancing Medical Decision-making. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025.

[4] Yue Zhang, Tianyi Ma, Zun Wang, **Yanyuan Qiao**, Parisa Kordjamshidi. Vision-and-Language Navigation with Analogical Textual Descriptions in LLMs. *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2025.

[5] Siqi Zhang, **Yanyuan Qiao**, Qunbo Wang, Zike Yan, Qi Wu, Zhihua Wei, Jing Liu. COSMO: Combination of Selective Memorization for Low-cost Vision-and-Language Navigation. *International Conference on Computer Vision (ICCV)*, 2025.

[6] Xiangyu Shi, Zerui Li, Wenqi Lyu, Jiatong Xia, Feras Dayoub, **Yanyuan Qiao**, Qi Wu. SmartWay: Enhanced Waypoint Prediction and Backtracking for Zero-Shot Vision-and-Language Navigation. *International Conference on Intelligent Robots and Systems (IROS)*, 2025.

[7] Xiangyu Shi, **Yanyuan Qiao**, Lingqiao Liu, Feras Dayoub. Embodied Domain Adaptation for Object Detection. *International Conference on Intelligent Robots and Systems (IROS)*, 2025.

[8] **Yanyuan Qiao**, Wenqi Lyu, Hui Wang, Zixu Wang, Zerui Li, Yuan Zhang, Mingkui Tan, Qi Wu. Open-Nav: Exploring Zero-Shot Vision-and-Language Navigation in Continuous Environment with Open-Source LLMs. *International Conference on Robotics and Automation (ICRA)*, 2025.

[9] Junyou Zhu, **Yanyuan Qiao**, Siqi Zhang, Xingjian He, Qi Wu, Jing Liu. MiniVLN: Efficient Vision-and-Language Navigation by Progressive Knowledge Distillation. *International Conference on Robotics and Automation (ICRA)*, 2025. (**Best Paper Finalist**)

[10] Wenbo Zhang, Yang Li, **Yanyuan Qiao**, Siyuan Huang, Jiajun Liu, Feras Dayoub, Xiao Ma, Lingqiao Liu. Effective Tuning Strategies for Generalist Robot Manipulation Policies. *International Conference on Robotics and Automation (ICRA)*, 2025.

[11] Zerui Li, Gengze Zhou, Haodong Hong, Yanyan Shao, Wenqi Lyu, **Yanyuan Qiao**, Qi Wu. Ground-level Viewpoint Vision-and-Language Navigation in Continuous Environments. *International Conference on Robotics and Automation (ICRA)*, 2025.

[12] Haodong Hong, **Yanyuan Qiao**, Sen Wang, Jiajun Liu, Qi Wu. General Scene Adaptation for Vision-and-Language Navigation. *International Conference on Learning Representations (ICLR)*, 2025.

[13] Qianyi Liu, Siqi Zhang, **Yanyuan Qiao**, Junyou Zhu, Xiang Li, Longteng Guo, Qunbo Wang, Xingjian He, Qi Wu, Jing Liu. GroundingMate: Aiding Object Grounding for Goal-Oriented Vision-and-Language Navigation. *Winter Conference on Applications of Computer Vision (WACV)*, 2025

[14] **Yanyuan Qiao**, Qianyi Liu, Jiajun Liu, Jing Liu, Qi Wu. LLM as Copilot for Coarse-grained Vision-and-Language Navigation. *European Conference on Computer Vision (ECCV)*, 2024.

[15] Mingzhen Sun, Weining Wang, **Yanyuan Qiao**, Jiahui Sun, Zihan Qin, Longteng Guo, Xinxin Zhu, Jing Liu. MM-LDM: Multi-Modal Latent Diffusion Model for Sounding Video Generation. *ACM Multimedia Conference (ACMMM)*, 2024.

[16] **Yanyuan Qiao**, Yuankai Qi, Zheng Yu, Jing Liu, Qi Wu. March in Chat: Interactive Prompting for Remote Embodied Referring Expression. *International Conference on Computer Vision (ICCV)*, 2023.

[17] **Yanyuan Qiao**, Zheng Yu, Qi Wu. VLN-PETL: Parameter-Efficient Transfer Learning for Vision-and-Language Navigation. *International Conference on Computer Vision (ICCV)*, 2023.

[18] **Yanyuan Qiao**, Yuankai Qi, Yicong Hong, Zheng Yu, Peng Wang, Qi Wu. HOP: History-and-Order Aware Pre-training for Vision-and-Language Navigation. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022.

[19] **Yanyuan Qiao**, Qi Chen, Chaorui Deng, Ning Ding, Yuankai Qi, Mingkui Tan, Xincheng Ren, Qi Wu. R-GAN: Exploring Human-like Way for Reasonable Text-to-Image Synthesis via Generative Adversarial

Networks. *ACM International Conference on Multimedia (ACM MM)*, 2021.

[20] **Yanyuan Qiao**, Zheng Yu, Jing Liu. RANKVQA: Answer Re-ranking for Visual Question Answering. *IEEE International Conference on Multimedia and Expo (ICME)*, 2020. (**Oral**)

[21] **Yanyuan Qiao**, Zheng Yu, Jing Liu. VC-VQA: Visual calibration mechanism for Visual Question Answering. *IEEE International Conference on Image Processing (ICIP)*, 2020.

[22] Zhiwei Fang, Jing Liu, **Yanyuan Qiao**, Qu Tang, Yong Li, Hanqing Lu. Enhancing Visual Question Answering Using Dropout. *ACM International Conference on Multimedia (ACM MM)*, 2018.

Workshops

[1] **Yanyuan Qiao**, Zheng Yu, Zijia Zhao, Sihan Chen, Mingzhen Sun, Longteng Guo, Qi Wu, Jing Liu. VL-Mamba: Exploring State Space Models for Multimodal Learning. *NeurIPS Workshop on Efficient Natural Language and Speech Processing*, 2024.

[2] Xiangyu Shi, **Yanyuan Qiao**, Qi Wu, Lingqiao Liu, Feras Dayoub. Improving Online Source-free Domain Adaptation for Object Detection by Unsupervised Data Acquisition. *ECCV Workshop on ROAM*, 2024.

[3] Zheng Yu, **Yanyuan Qiao**, Yutong Xie, Qi Wu. Multi-Modal Adapter for Medical Vision-and-Language Learning. *International Workshop on Machine Learning in Medical Imaging (MLMI@MICCAI)*, 2023.

Preprints

[1] **Yanyuan Qiao**, Kieran Gilday, Yutong Xie, Josie Hughes. Lang2Morph: Language-Driven Morphological Design of Robotic Hands. *ArXiv*, 2025.

[2] Xiangyu Shi, Zerui Li, **Yanyuan Qiao**, Qi Wu. Fast-SmartWay: Panoramic-Free End-to-End Zero-Shot Vision-and-Language Navigation. *ArXiv*, 2025.

[3] Wenqi Lyu, Zerui Li, **Yanyuan Qiao**, Qi Wu. BadNAVer: Exploring Jailbreak Attacks On Vision-and-Language Navigation. *ArXiv*, 2025.

HONORS & AWARDS

- NeurIPS Scholar Award   2025
- ICRA Best Paper Award (Finalist)   2025
- ICCV Doctoral Consortium   2023
- Dean's Commendation for Doctoral Thesis Excellence   2023
- 3rd place in the Large-Scale Robot and Embodied AI Challenge at the 2023 CCF ChinaSoft Cup   2023
- School of Computer Science PhD Scholarship   2020
- Ecoles d'été France Excellence Scholarship   2018
- 4th place in the VQA Challenge at CVPR Workshop 2018   2018

GRANTS & FUNDS

- **Swiss AI Initiative Compute Grant** (35,000 GPU Hours)   2025
  Awarded compute resources to support research in LLMs for generative robotics design.

TEACHING

Guest Lecture   *2025 Autumn Semester*
Course: Human and Computer Vision
Department of Computer Vision, Mohamed bin Zayed University of Artificial Intelligence

Guest Lecture   *2025 Spring Semester*
Course: Advanced Techniques in Visual Object Recognition and Detection: Multimodal Pretraining
Department of Computer Vision, Mohamed bin Zayed University of Artificial Intelligence

Guest Lecture   *2024 Semester 2*

Course: Applied Machine Learning: AI in Robotics
School of Computer and Mathematical Sciences, University of Adelaide

Teaching Assistant   *2024 Trimester 2&3*
Course: Artificial Intelligence and Machine Learning Research Project
School of Computer and Mathematical Sciences, University of Adelaide

TALKS

"Language Model × Robotics – From Embodied Navigation to AI-Driven Robot Hand Design."
- Mohamed bin Zayed University of Artificial Intelligence   Oct. 2025

"Exploring Zero-shot Embodied Vision-and-Language Navigation in Physical World."
- TechBeat, Online   Apr. 2025

"Advancing Multimodal Foundation Models for Embodied AI."
- University of Zurich, Online   Mar. 2025

"Comprehension, Reasoning and Interaction: Exploring Multimodal Agent in The Era of Foundation Models."
- ETH Zurich, Online   Jan. 2025

"Exploring Vision-and-Language Navigation in the Era of Foundation Models."
- Shanghai Jiao Tong University, Online   Dec. 2024

"Advancing Embodied Vision-and-Language Navigation in the Era of Foundation Models."
- Monash University   Nov. 2024

"Vision-and-Language Navigation."
- Australia Institue for Machine Learning Research Showcase   Aug. 2023

"General Vision and Language Methods in Real Applications: A Focus on Vision-and-Language Navigation."
- QingYuan Talk, Online   Aug. 2023
- Australia Institue for Machine Learning   Jul. 2023

ADVISING & MENTORSHIP

Co-Advisor
- Jian Zhou, *Mphil* at University of Adelaide (2024 - 2025).
- Zerui Li, *Mphil* at University of Adelaide (2024 - 2025). Next: *PhD* at University of Adelaide

Research Mentor
- Haodong Hong, *PhD* at The University of Queensland (2024 - 2025).
- Wenqi Lyu, *MSc* at University of Adelaide (2024 - 2025).
- Siqi Zhang, *PhD* at Tongji University (2023 - 2025).
- Xiangyu Shi, *Mphil* at University of Adelaide (2023 - 2025). Next: *PhD* at University of Adelaide
- Mingzhen Sun, *PhD* at University of Chinese Academy of Sciences (2023 - 2024). Next: *Researcher* at ByteDance
- Junyou Zhu, *Mphil* at University of Chinese Academy of Sciences (2023 - 2024).
- Qianyi Liu, *Mphil* at University of Chinese Academy of Sciences (2023 - 2024). Next: *PhD* at Peking University

PROFESSIONAL ACTIVITIES

- Journal Reviewer: IJCV, TIP, TCSVT, PR, TOMM, RAL, Neurocomputing, ACM Computing Surveys
- Conference Reviewer: ICLR 2025/2026, ICML 2025, CVPR 2024/2025/2026, ICCV 2025, ECCV 2024, NeurIPS 2024/2025, CoRL 2025, ACM MM 2024/2025, AAAI 2024, ICRA 2024/2025/2026, IROS 2025, BMVC 2021/2023/2024, WACV 2024.