Rui Qian

Email : qr021@ie.cuhk.edu.hk
Github : https://github.com/shvdiwnkozbw

SHB 702, CUHK
Hong Kong, China

OBJECTIVE

Ph.D. candidate in Multi-Media Lab, Department of Information Engineering, The Chinese University of Hong Kong, supervised by Prof. Dahua Lin. I am interested in computer vision and machine learning, especially self-supervised learning, video understanding and multi-modal large language models.

EDUCATION

• The Chinese University of Hong Kong, Hong Kong, China
Ph.D. Candidate, Information Engineering, August 2021 - Present
• Shanghai Jiao Tong University, Shanghai, China
Undergraduate Student, Information Engineering, September 2017 - June 2021
GPA: 3.95/4.3, Score: 91.70/100, Rank: 2/147

TECHNICAL SKILLS

Languages : Python, Matlab, C++
Tools/Framework : PyTorch, OpenCV
Research Interests : Video Understanding, Self-supervised Representation Learning, Multi-modal LLM

EXPERIENCE

Shanghai AI Lab Research Intern **Dec. 2023 - Present**
• Research on Multi-modal Large Language Models, especially for effective long video understanding.
† Supervised by Dr. Jiaqi Wang

CUHK MMLab **Aug. 2021 - Present**
• Research on Self-supervised Video Representation Learning, with one paper accepted by ECCV 2022, one paper accepted by ACMMM 2022.
• Research on Unsupervised Object-centric Video Analysis, with one paper accepted by ICCV 2023.
† Supervised by Prof. Dahua Lin

SJTU MIN Lab **Dec. 2018 - Jun. 2021**
• Research on Joint Audiovisual Learning especially Sound Source Localization, with one paper accepted by ECCV 2020.
• Participation in the organization of Human-in-Events Challenge on ACM Multimedia 2020 for large-scale human-centric video analysis in complex events.
• Research on Self-supervised Video Representation Learning, with one paper accepted by ICCV 2021.
† Supervised by Prof. Weiyao Lin

Baidu Research Cooperation **Mar. 2020 - Jun. 2020**
• Research on discriminatively localizing sounding objects in a cocktail-party scenario in a self-supervised manner, with one paper accepted by NeurIPS 2020.
† Supervised by Prof. Di Hu

SenseTime Research Intern **Feb. 2021 - Jun. 2021**
• Work in OpenMMLab group on transformer and video understanding.
† Supervised by Dr. Kai Chen

AWARDS

- **National Scholarship at SJTU** Oct. 2018
- **Ji Hanbing Scholarship at SJTU** Nov. 2019
- **Rongchang Technology Innovation Scholarship at SJTU** Nov. 2020
- **SenseTime Scholarship** Dec. 2020
- **Hong Kong PhD Fellowship Scheme** Apr. 2021
- **Top 1% Bachelor Thesis Award of SJTU** Jun. 2021
- **Outstanding Graduate of Shanghai** Jun. 2021

PUBLICATIONS

- **R. Qian**, X. Dong, P. Zhang, Y. Zang, S. Ding, D. Lin, J. Wang. Streaming Long Video Understanding with Large Language Models. arXiv preprint, 2024.
- S. Ding, Z. Liu, X. Dong, P. Zhang, **R. Qian**, C. He, D. Lin, J. Wang. SongComposer: A Large Language Model for Lyric and Melody Composition in Song Generation. arXiv preprint, 2024.
- S. Ding*, **R. Qian***, H. Xu, D. Lin, H. Xiong. Betrayed by Attention: A Simple yet Effective Approach for Self-supervised Video Object Segmentation. arXiv preprint, 2023.
- **R. Qian**, S. Ding, X. Liu, D. Lin. Semantics Meets Temporal Correspondence: Self-supervised Object-centric Learning in Videos. The IEEE International Conference on Computer Vision (ICCV), 2023.
- S. Ding, P. Zhao, X. Zhang, **R. Qian**, H. Xiong, Q. Tian. Prune Spatio-temporal Tokens by Semantic-aware Temporal Accumulation. The IEEE International Conference on Computer Vision (ICCV), 2023.
- L. Zhu*, X. Liu*, X. Liu, **R. Qian**, Z. Liu, L. Yu. Taming Diffusion Models for Audio-Driven Co-Speech Gesture Generation. The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.
- **R. Qian**, S. Ding, X. Liu, D. Lin. Static and Dynamic Concepts for Selfsupervised Video Representation Learning. The European Conference on Computer Vision (ECCV), 2022.
- S. Ding, **R. Qian**, H. Xiong. Dual Contrastive Learning for Spatio-temporal Representation. The ACM International Conference on Multimedia (ACMMM), 2022.
- S. Ding, M. Li, T. Yang, **R. Qian**, H. Xu, Q. Chen, J. Wang, H. Xiong. The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- X. Liu, Q. Wu, H. Zhou, Y. Xu, **R. Qian**, X. Lin, X. Zhou, W. Wu, B. Dai, B. Zhou. Learning Hierarchical Cross-Modal Association for Co-Speech Gesture Generation. The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
- X. Liu*, **R. Qian***, H. Zhou*, D. Hu, W. Lin, Z. Liu, B. Zhou, X. Zhou. Visual Sound Localization in the Wild by Cross-modal Interference Erasing. The AAAI Conference on Artificial Intelligence (AAAI), 2022.
- S. Li*, H. Liu*, **R. Qian**, Y. Li, J. See, M. Fei, X. Yu, W. Lin. TA2N: TwoStage Action Alignment Network for Few-shot Action Recognition. The AAAI Conference on Artificial Intelligence (AAAI), 2022.
- **R. Qian**, Y. Li, H. Liu, J. See, S. Ding, X. Liu, D. Li, W. Lin. Enhancing Self-supervised Video Representation Learning via Multi-level Feature Optimization. The IEEE International Conference on Computer Vision (ICCV), 2021.
- D. Hu, **R. Qian**, M. Jiang, X. Tan, S. Wen, E. Ding, W. Lin, D. Dou. Discriminative Sounding Objects Localization via Self-supervised Audiovisual Matching. Advances in Neural Information Processing Systems (NeurIPS), 2020.
- **R. Qian**, D. Hu, H. Dinkel, M. Wu, N. Xu, W. Lin. Multiple Sound Sources Localization from Coarse to Fine. The European Conference on Computer Vison (ECCV), 2020.
- **R. Qian**, D. Hu, H. Dinkel, M. Wu, N. Xu, W. Lin. A Two-Stage Framework for Multiple Sound-Source Localization. The IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020.
- S. Li, J. Li, H. Tang, **R. Qian**, W. Lin. ATRW: A Benchmark for Amur Tiger Re-identification in the Wild. The ACM International Conference on Multimedia (ACMMM), 2020.

- Y. Li, W. Lin, T. Wang, J. See, **R. Qian**, N. Xu, L. Wang, S. Xu. Finding Action Tubes with a Sparse-to-Dense Framework. The AAAI Conference on Artificial Intelligence (AAAI), 2020.