Zhijian Hou

zjhou3-c@my.cityu.edu.hk • (+86) 13517984847 • Shenzhen, China
[LinkedIn](#) • [Github](#) • [Home Page](#) • [Google Scholar](#)

SUMMARY OF QUALIFICATION

I am a skilled AI researcher with a Ph.D. in computer science and six years of experience in cutting-edge AI research. My expertise lies in computer vision and natural language processing, especially video retrieval, with a proven track record of 4 peer-reviewed papers and 5 workshop papers in top conferences. I am proficient in developing novel and forefront deep learning models via PyTorch and Python and deeply understand the latest AI technologies. I am motivated to extend beyond mere research and apply the latest AI technologies (e.g., large language models and multi-modal systems) for practical and feasible solutions to complex business challenges through my excellent analytical thinking and problem-solving skills.

EDUCATION

City University of Hong Kong   **Sep 2018–Oct 2023**
Ph.D. in Computer Science   Hong Kong, China

Shandong University   **2014–2018**
B.Sc. in Computer Science and Technology   Jinan, China

WORK EXPERIENCE

Microsoft Research Asia   **Apr 2022–June 2023**
Research Intern   Beijing, China
Work on the task of temporal grounding in videos via natural language queries.

- Led the development of advanced deep learning algorithms to enhance model performance and efficiency.
- Presented the progress weekly with researchers inside the company and outside collaborators.
- Won the championship award as the first author in the CVPR 2023 Ego4D workshop.
- Published a first-author paper in ACL 2023 to convey technical insights.

SELECTED PROJECTS

- **CONE:** A coarse-to-fine multi-modal alignment framework using the vision-language pre-trained models, accepted in ACL 2023. It significantly shortens the inference speed by 15x on hour-long videos. See the source on [GitHub](#).
- **GroundNLQ:** A multi-scale multi-modal grounding model with two-stage vision-language pre-training learning. It won the championship award in the CVPR 2023 Ego4D workshop. It achieves a sizeable performance boost of 81%. See the source on [GitHub](#).
- **CONQUER:** A two-step vision-language fusion model for temporal grounding in multi-modal video content, accepted in ACM MM 2021 oral. It obtains a 50% performance increase on a popular dataset. See the source on [GitHub](#).

SELECTED PUBLICATIONS

- **Zhijian Hou**, Wanjun Zhong, Lei Ji, Difei Gao, Yan Kun, Wing-Kwong Chan, Chong-Wah Ngo, Mike Zheng Shou, Nan Duan, "**CONE: An Efficient Coarse-to-fine Alignment Framework for Long Video Temporal Groundin**g," In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (**ACL**)*, 2023.
- **Zhijian Hou**, Lei Ji, Difei Gao, Wanjun Zhong, Chao Li, Yan Kun, Wing-Kwong Chan, Chong-Wah Ngo, Nan Duan, Mike Zheng Shou, "**GroundNLQ @ Ego4D Natural Language Queries Challenge 2023**," In *the 3rd International Ego4D workshop at **CVPR***, 2023.
- **Zhijian Hou**, Chong-Wah Ngo, and Wing-Kwong Chan, "**CONQUER: Contextual Query-aware Ranking for Video Corpus Moment Retrieval**," In *Proceedings of the 29th ACM International Conference on Multimedia (**ACM MM**)*, 2021 (**Oral**).