Yueqian Wang（王越千）

wangyueqian@pku.edu.cn | linkedIn/wangyueqian | https://yellow-binary-tree.github.io

EDUCATION

School of Intelligence Science and Technology, Peking University 2021.09-2026.07 (est.)
Ph.D. in Computer Applied Technology
Supervised by: Prof. Dongyan Zhao & Ass. Prof. Huishuai Zhang

Department of Information Management, Peking University 2017.09-2021.07
BS in Information Management & Information Systems

School of Electronics Engineering and Computer Science, Peking University 2018.09-2021.07
Double Major BS in Computer Science

PUBLICATIONS

Research Interests: Video Dialogue, Video Question Answering (Video QA), Multimodal Large Language Models (MLLM)

VideoLLM Knows When to Speak: Enhancing Time-Sensitive Video Comprehension with Video-Text Duet Interaction Format - Under Review
Yueqian Wang, Xiaojun Meng, Yuxuan Wang, Jianxin Liang, Jiansheng Wei, Huishuai Zhang, Dongyan Zhao
Proposed MMDuet, a video-text MLLM for real-time interaction, which autonomously decides its response timing during video playback, and its training dataset MMDuetIT.

HawkEye: Training Video-Text LLMs for Grounding Text in Videos - Under Review
Yueqian Wang, Xiaojun Meng, Jianxin Liang, Yuxuan Wang, Qun Liu, Dongyan Zhao
Proposed HawkEye, one of the first video-text LLMs that can perform temporal video grounding in a fully text-to-text manner, and InternVid-G, a large-scale video-text dataset for temporal video grounding training.

Friends-MMC: A Dataset for Multi-modal Multi-party Conversation Understanding - AAAI 2025
Yueqian Wang, Xiaojun Meng, Yuxuan Wang, Jianxin Liang, Qun Liu, Dongyan Zhao
Defined tasks related to "multimodal multi-party conversation understanding", collected the Friends-MMC dataset from TV series, and introduced baseline models.

Understanding Multimodal Hallucination with Parameter-Free Representation Alignment - Under Review
Yueqian Wang, Jianxin Liang, Yuxuan Wang, Huishuai Zhang, Dongyan Zhao
Proposed a parameter-free and training-free method for analyzing the quantity of information within image representations, and its applications in multimodal hallucination ascription.

STAIR: Spatial-Temporal Reasoning with Auditable Intermediate Results for Video Question Answering - AAAI 2024
Yueqian Wang, Yuxuan Wang, Kai Chen, Dongyan Zhao
Proposed STAIR, A neural module network (NMN) based method for video qa with long videos and complicated questions.

Overview of the NLPCC 2023 Shared Task 10: Learn to Watch TV: Multimodal Dialogue Understanding and Response Generation - NLPCC 2023
Yueqian Wang, Yuxuan Wang, Dongyan Zhao
Hosted a shared task about video dialogue understanding and next utterance prediction at NLPCC 2023.

SMASH: Improving SMAll Language Models' Few-SHot Ability with Prompt-Based Distillation - EMNLP 2022 Findings
Yueqian Wang, Chang Liu, Kai Chen, Xi Wang, Dongyan Zhao
Proposed SMASH, a prompt-based model distillation method for enhancing few-shot learning ability of language models.

ReasVQA: Advancing VideoQA with Imperfect Reasoning Process - NAACL 2025
Jianxin Liang, Xiaojun Meng, Huishuai Zhang, Yueqian Wang, Jiansheng Wei, Dongyan Zhao
Proposed ReasVQA, a framework for extracting reasoning processes from general-purpose large MLLMs to fine-tune smaller models on specific downstream tasks.

VideoHallucer: Evaluating Intrinsic and Extrinsic Hallucinations in Large Video-Language Models - Under Review
Yuxuan Wang, Yueqian Wang, Dongyan Zhao, Cihang Xie, Zilong Zheng
Collected VideoHallucer, a video qa benchmark with 6 different tasks to evaluate hallucination of video MLLMs.

OmniMMI: A Comprehensive Multi-modal Interaction Benchmark in Streaming Video Contexts - CVPR 2025
Yuxuan Wang, Yueqian Wang, Bo Chen, Tong Wu, Dongyan Zhao, Zilong Zheng
Collected OmniMMI, a video qa benchmark with 6 different tasks to evaluate time-sensitive tasks such as multiple-round dialogue with video localization.

Efficient Temporal Extrapolation of Multimodal Large Language Models with Temporal Grounding Bridge - EMNLP 2024
Yuxuan Wang, Yueqian Wang, Pengfei Wu, Jianxin Liang, Dongyan Zhao, Zilong Zheng
Proposed TGB, a method uses optical flow for relevant frame selection, to improve video MLLMs' video localization and QA abilities.

End-to-End Video Question Answering with Frame Scoring Mechanisms and Adaptive Sampling - Under Review
Jianxin Liang, Xiaojun Meng, Yueqian Wang, Chang Liu, Qun Liu, Dongyan Zhao
Proposed VidF4, an efficient and effective video question and answer method through key frame selection.

VSTAR: A Video-grounded Dialogue Dataset for Situated Semantic Understanding with Scene and Topic Transitions - ACL 2023
Yuxuan Wang, Zilong Zheng, Xueliang Zhao, Jinpeng Li, Yueqian Wang, Dongyan Zhao
Collected VSTAR, a million-turn-scaled video video dialogue corpus collected from TV series with scene and segment transition annotation, and conducted studies on tasks like scene segmentation and topic segmentation.

PATENTS

Yueqian Wang, Xiaojun Meng, Qun Liu, Dongyan Zhao: Conversation processing methods, equipments, electronic devices and storage media（对话处理方法、装置、电子设备及存储介质）. Under Review.

Dongyan Zhao, Yueqian Wang, Yuxuan Wang, Zilong Zheng: A multimodal dialogue method and system based on semantic graphs（一种基于语义图的多模态对话方法和系统）. 2024100614011.

Dongyan Zhao, Yueqian Wang, Yuxuan Wang: A video-text multi-modal question answering method and system based on neural modular networks（一种基于模块神经网络的视频-文本多模态问答方法和系统）. 2023114332336.

INTERNSHIPS

MiraclePlus - Industry Research | Beijing (Online) 2024.07-

Investigate latest developments in AI, assist in writing industry research reports and other documents.

ByteDance - Douyin Search Algorithm | Beijing 2020.06-2020.09

Improve personal search strategies of Douyin search.

TEACHING EXPERIENCE

TA @ Introduction to Computation (B) 2022.02-2022.07

TA @ Data Analysis with Python 2021.02-2021.07