Shijie Wang

shijie_wang@brown.edu

Department of Computer Science, Brown University

Education

Brown University, Providence, RI, US *Department of Computer Science* 09/2021 - Now

Ph.D. student in Computer Science

- Research Areas: Computer Vision, Multimodal Learning
- Advisor: Prof. Chen Sun

Tsinghua University, Beijing, China *School of Software* 09/2016 - 07/2021

B.Eng. in Software Engineering

Outstanding Graduate

- Research Areas: Transfer Learning, Computer Vision

Research

Revisiting Concept Binding in Contrastive Language-Image Pretraining (Under Review)

Supervised by Prof. Chen Sun, Collaboration with Meta AI

- We investigate whether contrastive VLMs bind concepts and reason relations with entity-centric representations. Practically, we utilize bounding-box or masks as oracle localization knowledge to build entity-centric representations.
- Experiments in a controlled synthetic environment show that an explicit decomposition of scene-level features into entity-centric representations benefits both the entity-level binding task and the inter-entity relational reasoning task.
- However, the post-hoc entity-centric representations still struggle on fine-grained real-world datasets for part attribute binding, indicating a potential direction for future vl-pre-training methods: the integration of inductive biases that promote the emergence of entity-centric information.

Can Large Language Models Help Long-term Action Anticipation from Videos? (Under Review)

Supervised by Prof. Chen Sun, Collaboration with Honda Research

- We propose **AntGPT**, a framework to leverage LLM in long-term action anticipation tasks in both bottom-up methods to predict future actions directly and top-down methods guided by high-level goals using ICL/CoT or fine-tuned models.
- Experiments show LLMs encode rich prior knowledge for temporal dynamics, which substantially enhances bottom-up LTA predictions and LLMs' ability to infer reasonable long-term goals from observed actions. With LLM-generated goals, top-down predictions show further improvement compared with bottom-up ones.
- Achieve competitive SoTA performance on the Ego4D LTA v1/v2, EK-55, and EGTE benchmark.

Goal-Conditioned Predictive Coding as an Implicit Planner for Offline Reinforcement Learning (Under Review)

Supervised by Prof. Chen Sun, Brown University

- We investigate if trajectories can be condensed into powerful representations useful for policy learning.
- We design a two-stage framework that first summarizes trajectories using sequence modeling techniques, and then uses these representations to learn a policy along with a desired goal.
- We demonstrate that our proposed framework learns a goal-conditioned latent representation of the future, which serves as an "implicit planner", and enables it to achieve competitive performance on three benchmarks.

Prompt-based Object-centric Video Representation for Action Anticipation (Under Review)

Supervised by Prof. Chen Sun, Collaboration with Honda Research

- We propose to build object-centric video representations by leveraging visual-language pre-trained models by 'object prompts', an approach to extract task-specific object-centric representations from general-purpose pre-trained models without finetuning.
- Conduct evaluations on various action anticipation benchmarks. Both quantitative and qualitative results confirm the effectiveness of our proposed object prompts and the overall model.

Bottleneck Hallucination for Modality-missing Robust Video Understanding 06/2022 – 03/2023

Mentored by Dr. Yin Cui, Google Research

- Investigate the influence of missing modalities on multimodal video understanding.
- Proposed bottleneck hallucination and modality dropout to improve MBT's (multimodal bottleneck transformer) robustness against video and audio missing during evaluation without prior information about the missing modality.

Pose Recognition with Cascade Transformers 07/2020 - 11/2020

Supervised by Prof. Zhuowen Tu, University of California, San Diego

- Presented a regression-based 2D human pose recognition method using cascade Transformers consisting of a person detection Transformer and a keypoint detection Transformer named Pose Regression TRansformers (PRTR).

- PRTR achieves SOTA compared to other existing regression-based methods on the challenging COCO dataset.
- The work has been accepted by CVPR 2021.

INTERNSHIP

Google Research | Student Researcher 05/2022 – 03/2023

- Working on the research topic of multimodal models' robustness towards modality-missing data.
- Working on Video Understanding and got 3rd prize in Ego4D Object State Change Classification Challenge at ECCV 2022 Workshop.

Kwai Inc. | *Machine Learning Intern of MultiMedia Understanding Group* 07/2019 - 08/2020

- Kwai is one of the largest social media companies in China.
- Built a **multimodal** machine learning model with multi-frame features, text features, and audio features for video content review, resulting in great improvement in F-score; our model has been put into practical use.
- Accumulated machine learning life cycle and big data system development experience, including data wrangling, feature engineering, and model deployment.

PUBLICATION

Pose Recognition with Cascade Transformers (CVPR 2021)

Ke Li*, Shijie Wang*, Xiang Zhang*, Yifan Xu, Weijian Xu, Zhuowen Tu
(*equal contribution)

AWARDS & HONORS

3rd Prize of Ego4D Object State Change Classification Challenge, ECCV 2022	2022
Outstanding Graduate Awards, Tsinghua University	2021
Scholarship for Academic Excellence, Tsinghua University	2018&2019&2020
Member of Tsinghua University Initiative Scientific Research Program (funding: 30,000￥)	2019
1st Prize in Student Research Training Program, Tsinghua University	2019
2nd Prize in Software Design Contest, Tsinghua University	2018

SERVICE

Conference Reviewer:

• Conference on Neural Information Processing Systems (NeurIPS)	2023
• The Conference on Computer Vision and Pattern Recognition (CVPR)	2022, 2023
• International Conference on Computer Vision (ICCV)	2023
• The European Conference on Computer Vision (ECCV)	2022
• AAAI Conference on Artificial Intelligence (AAAI)	2023
• Winter Conference on Applications of Computer Vision (WACV)	2023

EXTRACURRICULAR ACTIVITES

- Vice president of Microsoft Club at Tsinghua University, member of Microsoft Summer Camp, 2019.
- Member of the football team in the school of Software Engineering and Department of Electronic Engineering.
- Champion of Yuehan Ma Campus Football Cup, 2018.