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

PUBLICATION

Pose Recognition with Cascade Transformers (CVPR 2021)
Ke Li*, Shijie Wang*, Xiang Zhang*, Yifan Xu, Weijian Xu, Zhuowen Tu
(*equal contribution)

RESEARCH

Prompt-based Object-centric Video Representation for Action Anticipation 09/2022 – 03/2023
Supervised by Prof. Chen Sun, Collaboration with Honda Research
- We propose to build object-centric video representations by leveraging visual-language pretrained models by 'object prompts', an approach to extract task-specific object-centric representations from general-purpose pretrained 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.

Study on Multimodal Robustness towards Missing Modality (Ongoing) 06/2022 – Now
Mentored by Dr. Yin Cui, Google Research
- Investigate the influence of missing modalities on multimodal models and different modality fusion methods in different training and inference settings.
- Proposed several methods to better learn cross-modality information and improve the robustness of multimodal models to missing modalities.

Study on RL-Based Vision-Language Navigation (Ongoing) 11/2021 - Now
Supervised by Prof. Chen Sun, Brown University
- Designing model-free Reinforcement Learning method with transformer structure for VLN tasks.
- Exploring methods for better Cross-Modal fusion on vision and language information and feature representation.

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.

Study of Transferability of Deep Neural Network for Regression 05/2020 - 08/2020
Supervised by Prof. Mingsheng Long, Tsinghua University
- The knowledge learned from the classification task can be partly used for regression, for the backbone networks, the lower layers have better transferability than the upper layers.
- We analyzed the difference between classification and regression and the reason why the regression task is hard to transfer. The state space is the essential difference between classification and regression.
- Replacing Batch Normalization with Instance Normalization can improve the transferability of DNN significantly, indicating regression transfer has some similarity with style transfer like a single image domain adaptation problem.
- Designing baseline models and doing more confirmatory experiments.

Transferable Attention for Domain Adaptation 07/2019 - 10/2019
Supervised by Prof. Mingsheng Long, Tsinghua University
- Presented the dimensional symmetry attention model for domain adaptation to improve the transferability of DNN.

- Used domain discriminative method to generate dimensional symmetry transferable attention: spatial, channel-wise and instance-wise transferable attention.
- Made transferable attention a standard and plug-in module suited for different domain adaptation models such as DANN and CDAN in different dataset like Office-Home and DomainNet, exceeding SOTA in some tasks on these datasets.

INTERNSHIP

Google Research | Student Researcher 05/2022 – Now

- 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.

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:

• The Conference on Computer Vision and Pattern Recognition (CVPR)	2022, 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

SELECTED COURSE PROJECT

Wechat Game: Doodle Gold Miner

- Course project for the course Web Front-end Technology. Using WeChat dev-tools and Cocos Creator.
- I work on UI design, main logic for the game, designing in-game animations, WeChat open domain ranking board, level system, and store system. I invited about 40 people to play the demo version.

C To LLVM Compiler

- Course project for course Principles of Compilation. Designing a compiler frontend to convert C language to LLVM IR.
- Use python and Antlr, the compiler supports most grammar in C, such as structure and array, some test codes are attached.

FTP Project & RTP Project

- Both are projects for the course Computer Network.
- In the FTP project, I complete an FTP server according to RFC 959 and an FTP client with a user-friendly GUI with support for resuming from break-point. The FTP server is compatible with many widely-used FTP clients like FileZilla.
- In the RTP project, I complete an RTP server according to RFC 1889 and a streaming media player client. The server and client support multiple video formats like avi, flv, mp4 and iso, lyrics display, and speed modification.

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.