YIWEI ZHANG

+86 18801311968 | email: yw_zhangthu@163.com
Department of Computer Science and Technology
Tsinghua University, P.R. China

EDUCATION

Tsinghua University Beijing, P.R. China
B.E. in Computer Science and Engineering *Aug. 2016 - Jul. 2020*

- GPA: 3.82/4.00(Rank 10/158); Major GPA: 3.90/4.00; Admitted on basis of excellent performance on national college admission exam(8/330,000)
- Selected Award: *President Jiang Nan-Xiang Scholarship*(**top 1/158**, awarded to those with outstanding academic performance and research experience)

Carnegie Mellon University Pittsburgh, PA
Visiting Student *Jul. 2019 - Dec. 2019*

- Visiting student to MultiComp Lab(Advised by Prof. Louis-Philippe Morency).

RESEARCH INTERESTS

In general, my research interest lies in the field of Machine Learning, in particular **multimodal machine learning** and its applications.

From a theoretical perspective, I am interested in understanding the computational and statistical principals in multimodal learning and building systems that can process and relate information from multiple modalities. From an application perspective, I apply these principals to solve problems in Computer Vision, Natural Language Processing, and Speech such as multimodal language modeling, multi-modality learning from videos, and audio-visual embodied indoor navigation.

PUBLICATIONS

1. Xuguang Duan, Qi Wu, Chuang Gan, **Yiwei Zhang**, Wenbing Huang, Anton van den Hengel, and Wenwu Zhu. 2019. "Watch, Reason and Code: Learning to Represent Videos Using Program." *Proceedings of the 27th ACM International Conference on Multimedia* (MM '19), p. 1543-1551, 2019
2. Gan Chuang[1], **Yiwei Zhang**[1], Jiajun Wu, Boqing Gong, and Joshua Tenenbaum. "Look, Listen, and Act: Towards Audio-Visual Embodied Navigation" (Accpeted in ICRA 2020.)

Under Review:

1. Amir Zadeh, Chengfeng Mao, Kelly Shi, **Yiwei Zhang**, Paul Pu Liang, Soujanya Poria, and Louis-Philippe Morency. "Factorized Multimodal Transformer for Multimodal Sequential Learning" (Under review in Information Fusion.)

RESEARCH EXPERIENCE

Language Technology Inititute, Carnegie Mellon University Pittsburgh, USA
Advisor: Prof. Louis-Philippe Morency *Jul. 2019 - Dec. 2019*

Incorporate Phoneme Information into Human Multimodal Language Modeling

- The first attempt to incorporate structural phoneme information into human multimodal language modeling.
- Found an accidental shift of opinion label segments and some misalignment between phonemes and words in CMU-MOSI, a frequently used dataset for multimodal research. Helped to generate the new data files.

[1]Equal Contribution

- Built a successful model that hierarchically attends to the nonverbal context at different levels and seamlessly makes use of phoneme information during inference.

Factorized Multimodal Transformer for Multimodal Sequential Learning
- Helped to generate the input data in the required format, searched for the best architecture for the model, and ran extensive experiments.
- Under review in the journal Information Fusion.

Spectral-Temporal Transformer in Human Multimodal Language Modeling
- Implemented this idea in the multimodal sentiment analysis task and introduced some effective modifications.
- Successfully boosted the performance by extracting the short and long-range relationships in both time domain and feature domain.
- Still in progress. Plan to submit it to a major AI conference in 2020.

Department of Computer Science and Technology, Tsinghua University Beijing, P.R. China
Advisor: Chuang Gan *Jul. 2018 - Jun. 2019*

Audio-Visual Embodied Navigation
- The first attempt to build a multimodal navigation environment with visual and audio inputs.
- Built a multimodal environment on top of the AI2-THOR platform. A spatial audio software development kit, Resonance Audio API, was integrated.
- Proposed a novel approach to the challenging task by combining three separate components: a visual perception mapper, an audio perception module, and a dynamic path planner.
- Under review in the ICRA 2020, one of the most important conferences in the field of robotics.

Learning to Represent Videos Using Program
- Proposed a better model for video-to-program synthesis, which expects to synthesis programs that describe the process carried out in a set of video sequences.
- This paper was published in ACM Multimedia 2019.

SELECTED AWARDS AND HONORS

· Huawei Scholarship at Tsinghua University 2019
Top 10%; Awarded to those with outstanding academic performance

· President Jiang Nan-Xiang Scholarship at Tsinghua University 2018
Top 1/158*; Awarded to those with outstanding academic performance and research experience*

· The Scholarship for Excellence in Social Practice Performance at Tsinghua University 2017
Top 10%; Awarded to those with outstanding social practice performance

· Zheng Geru Scholarship at Tsinghua University 2017
Top 10%; Awarded to those with outstanding academic performance

· Freshman Scholarship at Tsinghua University 2016
Top 10%; Awarded to those with excellent performance on national college admission exam

ENGLISH SKILLS

TOEFL 108(Reading 30; Listening 29; Speaking 22; Writing 27)
GRE Verbal: 157; Quantitative 170; Analytical Writing 3.0

PROGRAMMING SKILLS

Languages Python, JAVA, C/C++, Javascript, MATLAB, Bash
Machine Learning Pytorch, Tensorflow, Keras, scikit-learn