Zifan WANG

4742 Centre Ave Apt 505, Pittsburgh, PA, 15213
(+1) 408-242-6192 • zifan@cmu.edu • LinkedIn: Zifan Wang • Homepage

Education

Carnegie Mellon University • SV & Pitts • Pittsburgh, PA Aug 2019 – Present
Doctor of Philosophy • Electrical and Computer Engineering • GPA: 3.93/4.0

Carnegie Mellon University • Silicon Valley Campus • Mountain View, CA Aug 2018 – May 2019
Master of Science • Electrical and Computer Engineering • GPA: 3.93/4.0

Beijing Institute of Technology • Beijing, China Sept 2014 – July 2018
Bachelor of Science • Electronic Science and Technology • GPA: 88.7/100

Research Focus: Explainability and Robustness of Deep Neural Networks (in Vision Tasks)
Main Courses: Deep Learning; Computer Vision; Introduction to Machine Learning; Probability Graphical Models

Technical Skills

- Primary Programming Languages: Python
- Deap Learning Framework: Pytorch, Tensorflow, Keras
- Data Anlytics & Visualization: Pandas, Scipy, Sklearn, Skimage, OpenCV, Matplotlib, Plotly

Work Experience

Machine Learning Engineer – Internship, Truera, Menlo Park, CA May 2019 – Aug 2019, May 2020 – Aug 2020

- Developed neural network attribution library that produces local and global explanation for feed-forward networks. The library is still under testing and will be open-sourced soon.
- Collaborated closely with costumers' data science teams to improve the performance and reduce bias of existing models, e.g. GBM, GRU and image models, with feature-level explanations.
- Extracted features from exiting but complex models using explanations to augment features that help train simple models.
- Helped build multiple data visualizers in the front-end demo.

Publications

- Zifan Wang, Haofan Wang, Shakul Ramkumar, Matt Fredrikson, Piotr Mardziel, and Anupam Datta. Smoothed geometry for robust attribution, 2020 [**To Appear in NeurIPS 2020**]
- Zifan Wang, Piotr Mardziel, Anupam Datta, and Matt Fredrikson. Interpreting interpretations: Organizing attribution methods by criteria. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020
- Haofan Wang, Zifan Wang, Mengnan Du, Fan Yang, Zijian Zhang, Sirui Ding, Piotr Mardziel, and Xia Hu. Score-cam: Score-weighted visual explanations for convolutional neural networks. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2020

Pre-print

- Kaiji Lu, Zifan Wang, Piotr Mardzie, and Anupam Datta. Abstracting influence paths for explaining (contextualization of) bert models, 2020
- Zifan Wang, Yilin Yang, Ankit Shrivastava, Varun Rawal, and Zihao Ding. Towards frequency-based explanation for robust cnn, 2020
- Xuan Chen, Zifan Wang, Yucai Fan, Bonan Jin, Piotr Mardziel, Carlee Joe-Wong, and Anupam Datta. Towards behavior-level explanation for deep reinforcement learning, 2020

Tutorial

- From Explainability to Model Quality and Back Again. Anupam Datta, Matt Fredrikson, Klas Leino, Kaiji Lu, Shayak Sen, Zifan Wang [**To Appear in AAAI-21**]

Coursework & Projects

Human Joint Detection System on NVIDIA TX1 – Microelectronic Device Lab Mar 2018 - June 2018
Beijing Institute of Technology, Beijing, China

- Deployed the open source algorithms, OpenPose, on Nvidia TX1 to realize real-time human pose and human joint detection and recognition. The program reaches a 0.807 precision using MPII dataset with 9.7 FPS on NVIDIA TX1
- Added HOG+SVM human detector with ROS wrapper to pre-processs the input video.
- Ranked 1/32 in the evaluation group by the committee as outstanding thesis work.

Speech to Text Translation using Ecoder-Decoder model with Attention Nov 2018
Carnegie Mellon University-Silicon Valley, Mountain View, CA

- Built up a three-layer pBiLSTM network to encode the audio signals and an attention-based two-layer LSTM decoder to generate texts.
- Trained the network with WSJ dataset using character-based projection to learn unusual words.
- Improved the model with locked dropout, teacher force, weight decay in training and random search in inferring to achieve an average edit distance of 15.2 on the test set.