Peiyu **Yang**

(+61) 422188793 | yangpeiyv@gmail.com | ypeiyu | 10/19/1994

Research Interests: *Explainable Artificial Intelligence and Trustworthy Machine Learning.*

Education

PhD candidate of Computer Science — Perth, Australia

Department of Computer Science and Software Engineering, The University of Western Australia (**UWA**) — *Feb. 2021 - Dec. 2024*

- PhD Thesis: Explaining Deep Neural Networks to Establish Trust
- Supervisor: Professor Ajmal Mian, Dr. Naveed Ahktar

Research Assistant of Computer Science — Zhejiang, China

School of Computer Science, Zhejiang University (**ZJU**) — *May 2020 - Jan. 2021*

- Supervisor: Professor Zeke Wang

Master of Computer Technology — Shenyang, China

School of Computer Science and Engineering, Northeastern University (**NEU**) — *Sep. 2017 - Jan. 2020*

- Master's Thesis: Visible part prediction and temporal calibration for pedestrian detection
- Supervisor: Professor Lu Wang

Publications

Dynamic Model Editing to Rectify Unreliable Behavior in Neural Networks.

Peiyu Yang, Naveed Akhtar, Jiantong Jiang, Ajmal Mian

Under Review.

Backdoor-based Explainable AI Benchmark for High Fidelity Evaluation of Attribution Methods.

Peiyu Yang, Naveed Akhtar, Jiantong Jiang, Ajmal Mian

Under Review.

Regulating Model Reliance on Non-Robust Features by Smoothing Input Marginal Density.

Peiyu Yang, Naveed Akhtar, Mubarak Shah, Ajmal Mian

*European Conference on Computer Vision (**ECCV**)*, 2024.

Re-Calibrating Attributions for Model Interpretation.

Peiyu Yang, Naveed Akhtar, Zeyi Wen, Mubarak Shah, Ajmal Mian

*International Conference on Learning Representation (**ICLR Spotlight**)*, 2023.

Local Path Integration for Attribution.

Peiyu Yang, Naveed Akhtar, Zeyi Wen, Ajmal Mian

*AAAI Conference on Artificial Intelligence (**AAAI**)*, 2023.

Multi-Grained Interpretable Network for Image Recognition.

Peiyu Yang, Zeyi Wen, Ajmal Mian:

*International Conference on Pattern Recognition (**ICPR**)*, 2022.

A Part-Aware Multi-Scale Fully Convolutional Network for Pedestrian Detection.

Peiyu Yang, Guofeng Zhang, Lu Wang, Lisheng Xu, Qingxu Deng, Ming-Hsuan Yang

*IEEE Transactions on Intelligent Transportation Systems (**TITS**)*, 2021.

Visible Part Prediction and Temporal Calibration for Pedestrian Detection.

Peiyu Yang, Weixi Li, Lu Wang, Lisheng Xu, Qingxu Deng

*IET Image Processing (**IET-IP**)*, 2023.

Hardhat-Wearing Detection Based on a Lightweight Convolutional Neural Network with Multi-Scale Features and a Top-Down Module.

Lu Wang, Liangbin Xie, **Peiyu Yang**, Qingxu Deng, Shuo Du, Lisheng Xu

Sensors (2020)

Professional Activities

Reviewer ICLR 2025, NeurIPS 2024, ICML 2024-2025, CVPR 2024-2025, ECCV 2024, etc.