Yue Zhuo

PhD, Zhejiang University, China
zhuoy1995@zju.edu.cn — Google Scholar — +86 13685782929

RESEARCH INTERESTS

Reliable and Trustworthy ML, explainable AI, Adversarial Robustness, Optimization, Any-shot Learning, Intelligent Manufacturing, Anomaly/Fault Diagnosis

PERSONAL PROFILE

Yue Zhuo received PhD at Zhejiang University in 2023. He aims to address reliability and trustworthiness of machine learning. His research is primarily focused on explainable AI, adversarial robustness, optimization, and data augmentation. The resulting methods offer reliable solutions for learning from cross-domain datasets, particularly in manufacturing applications. As a first author, he has published seven high-quality papers in leading journals of his research area, with average impact factor of 10.0.

EDUCATION

Zhejiang University, Hangzhou, China 2018 — 2023
PhD in Control Science and Engineering
Thesis Title: Research on the Reliability of Data-driven Industrial System Fault Diagnosis Models

University of California, Riverside, USA 2016
Exchange program in Electrical Engineering

Zhejiang University of Technology, Hangzhou, China 2014 — 2018
B.Eng. in Automation and Electrical Engineering

ACADEMIC EXPERIENCE

Zhejiang University Hangzhou, China
PhD Research 2018 — 2023

- Developing advanced feature attribution methods for explaining deep models, involving adversarial, counterfactual and integrated gradient approaches.
- Exploring ML model resilience against perturbations, including adversarial attack and countermeasures, formal verification and robust learning.
- Designing optimization algorithms, encompassing multi-objective Bayesian optimization for AutoML and math programming for model formal verification.
- Addressing the model accuracy degradation caused by data scarcity, especially in the context of any-shot learning problem.
- Detecting, classifying and diagnosing anomaly and fault data for industrial process optimization and product quality diagnosis.

PUBLICATIONS

Reviewed and preprinted

- (Under 4th review) "IG^2: Integrated Gradient on Iterative Gradient Path for Feature Attribution," **Y. Zhuo** and Z. Ge, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023. preprint URL
- (Under 2nd review) "Feature Augmentation for Adversarial Robustness," **Y. Zhuo** and Z. Ge, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023. preprint URL
- (Submitted) "ABIGX: A Unified Framework for explainable Fault Detection and Classification." Automatica, **Y. Zhuo**, J. Qian, Z. Song, Z. Ge, 2023. preprint URL
- (Preprinted) "PatchProto Networks for Few-shot Visual Anomaly Classification." J. Wang and **Y. Zhuo**, 2023. preprint URL

Published paper

- "Security Versus Accuracy: Trade-Off Data Modeling to Safe Fault Classification Systems." **Y. Zhuo**, Z. Song and Z. Ge, IEEE Transactions on Neural Networks and Learning Systems, 2023.
- "Adversarial Security Verification of Data-Driven FDC Systems." **Y. Zhuo** and Z. Ge, IEEE Transactions on Reliability, 2022.
- "Attack and Defense: Adversarial Security of Data-Driven FDC Systems." **Y. Zhuo**, Z. Yin and Z. Ge, IEEE Transactions on Industrial Informatics, 2023, 19(1): 5-19.

- "One Variable Attack on the Industrial Fault Classification System and Its Defense." **Y. Zhuo**, Y. A.W. Shardt, Z. Ge, Engineering, 2022, 19: 240-251.
- "Data Guardian: A Data Protection Scheme for Industrial Monitoring Systems." **Y. Zhuo** and Z. Ge, IEEE Transactions on Industrial Informatics, 2022, 18(4): 2550-2559.
- "Auxiliary Information-Guided Industrial Data Augmentation for Any-Shot Fault Learning and Diagnosis." **Y. Zhuo** and Z. Ge, IEEE Transactions on Industrial Informatics, vol. 17, no. 11, pp. 7535-7545, Nov. 2021.
- "Gaussian Discriminative Analysis aided GAN for imbalanced big data augmentation and fault classification." **Y. Zhuo** and Z. Ge, in Journal of Process Control, vol. 92, pp. 271-287, Aug. 2020.
- "Transfer Adversarial Attacks Across Industrial Intelligent Systems." Z. Yin, **Y. Zhuo** and Z. Ge, Reliability Engineering & System Safety, 2023.

PROJECTS

Surface quality diagnosis of automotive steel 2022 — 2023

- Modeling steel surface quality with process control.
- Cross-process traceability analysis with explainable AI.
- Optimizing steel heating process based on surface quality.

Ammonia process optimization in chemical industry 2019 — 2022

- Data-driven fault diagnosis for reactors.
- Fault localization with feature attribution techniques.
- Soft sensing based on regression models.

SKILLS

- **AI explainability**: feature attribution, Shapely theory, counterfactuals
- **AI security**: adversarial attack and defense, formal verification
- **Optimization**: math programming, Bayesian optimization, multiple-objective optimization
- **Data modeling**: few-shot and zero-shot learning, anomaly detection
- **Data augmentation**: Generative Adversarial Networks
- **Coding and Softwares**: Pytorch, Tensorflow, Matplotlib, Latex, Adobe Illustrator, and etc.

REFERENCES

Prof. Zhihuang Song
Professor, College of Control Science and Engineering , Zhejiang University, China
E-mail: zhsong@iipc.zju.edu.cn — Scholar Profiles: Google Scholar

Zhiqiang Ge
Professor, College of Control Science and Engineering , Zhejiang University, China
E-mail: zhiqiang.ge@hotmail.com — Scholar Profiles: Google Scholar