Shaopeng Fu

Last Update: August 2024

shaopengfu.me
i7cWm4gAAAAJ
github.com/fshp971

shaopeng.fu@kaust.edu.sa
shaopengfu15@gmail.com
+966 (0) 56 534 0337

Education

King Abdullah University of Science and Technology — **Thuwal, Saudi Arabia**
Provable Responsible AI and Data Analytics (PRADA) Lab — Aug. 2023 – Present
Ph.D. Student in Computer Science
Advisor: Prof. Di Wang

The University of Sydney — **Sydney, Australia**
UBTECH Sydney Artificial Intelligence Centre — Oct. 2019 – Jan. 2021
Master of Philosophy (Engineering and IT)
Advisor: Prof. Dacheng Tao
Thesis Title: Bayesian Inference Forgetting

South China University of Technology — **Guangzhou, China**
B.Sc in Mathematics and Applied Mathematics — Sep. 2015 – Jun. 2019
Advisor: Prof. Chuhua Xian (Advising the Competitive Programming Group affiliated to School of CSE)

Employment

King Abdullah University of Science and Technology — **Thuwal, Saudi Arabia**
Provable Responsible AI and Data Analytics (PRADA) Lab — May 2023 – Aug. 2023
Research Intern (Topic: Adversarial Robustness; Advisor: Prof. Di Wang)

JD.com, Inc. — **Beijing, China**
JD Explore Academy — Mar. 2021 – Jul. 2022
Algorithm Engineer (Full-time)

- First author of two ICLR 2022 papers.
- Co-author of the *White Paper on Trustworthy Artificial Intelligence* (Chn Ver.) (Eng Ver.).
- Chief developer of **TAICore**, a trustworthy AI assessment toolkit powered by JD Explore Academy for assessing the robustness and privacy-preserving ability of white-box and black-box ML models.

The University of Sydney — **Sydney, Australia**
UBTECH Sydney Artificial Intelligence Centre — Oct. 2019 - Oct. 2020
Research Assistant (Topic: Machine Unlearning)

Research Interests

My research lies in trustworthy AI. I am interested in using mathematical principles to identify and mitigate security and privacy risks in real-world machine learning systems. Currently, I am working on:

- Adversarial Robustness of Pre-trained Models
- Privacy-preserving Ability of Pre-trained Models

Publications

Conferences & Journals

1. **Shaopeng Fu** and Di Wang. Theoretical Analysis of Robust Overfitting for Wide DNNs: An NTK Approach. In *International Conference on Learning Representation (ICLR)*, 2024.

2. **Shaopeng Fu**, Fengxiang He, Yang Liu, Li Shen, and Dacheng Tao. Robust Unlearnable Examples: Protecting Data Against Adversarial Learning. In *International Conference on Learning Representation (ICLR)*, 2022.
3. **Shaopeng Fu***, Fengxiang He*, and Dacheng Tao. Knowledge Removal in Sampling-based Bayesian Inference. In *International Conference on Learning Representation (ICLR)*, 2022.
4. Zeke Xie, Fengxiang He, **Shaopeng Fu**, Issei Sato, Dacheng Tao, and Masashi Sugiyama. Artificial Neural Variability for Deep Learning: On Overfitting, Noise Memorization, and Catastrophic Forgetting. *Neural Computation 33 (8)*, 2021.

Manuscripts

1. **Shaopeng Fu**, Xuexue Sun, Ke Qing, Tianhang Zheng, and Di Wang. Pre-trained Encoder Inference: Revealing Upstream Encoders In Downstream Machine Learning Services. *arXiv preprint arXiv:2408.02814*, 2024.
2. Fengxiang He*, **Shaopeng Fu***, Bohan Wang*, and Dacheng Tao. Robustness, Privacy, and Generalization of Adversarial Training. *arXiv preprint arXiv:2012.13573*, 2020.

Selected Awards

International Collegiate Programming Contest (ICPC)

• The ICPC Asia-East Continent Final Xi'an Site	Silver Medal, Dec. 2018
• The ICPC Asia Regional Contest Qingdao Site	Silver Medal, Nov. 2018
• The ICPC Asia Regional Contest Shenyang Site	Gold Medal (Rank: 6/186), Oct. 2018
• The ACM-ICPC Asia Regional Contest Xi'an Site	Silver Medal, Oct. 2017
2017-2018 China National Scholarship	Ministry of Education of P.R. China, Nov. 2018
2016-2017 China National Scholarship	Ministry of Education of P.R. China, Nov. 2017

Services

Conference Reviewer

- International Conference on Machine Learning (ICML): 2022, 2023, 2024
- International Conference on Learning Representations (ICLR): 2022, 2023, 2024
- Conference on Neural Information Processing Systems (NeurIPS): 2021, 2022, 2023, 2024
- International Conference on Artificial Intelligence and Statistics (AISTATS): 2021, 2024

Conference Committee Member

- ACM Conference on Computer and Communications Security (CCS): 2024 (Artifact Evaluation)
- AAAI Conference on Artificial Intelligence (AAAI): 2025

Journal Reviewer

- IEEE Transactions on Pattern Analysis and Machine Intelligence: 2024
- IEEE Transactions on Cybernetics: 2021
- Springer Neural Processing Letters: 2020

Teaching

Teaching Assistant of CS 229: Machine Learning, Spring 2024 @ KAUST

Miscellaneous

Competitive Programming: My Codeforces account is fshp971.

Others: C/C++, Python, PyTorch, JAX, Vim, Linux, Arch Linux.