Bingyin Zhao

Google Scholoar: Bingyin Zhao

Email : bingyiz@g.clemson.edu
Mobile : +1-585-748-1626

Summary

I am currently a PhD student working with Dr. Yingjie Lao at Clemson University. I have over five years of research experience in trustworthy AI, machine learning security and hardware-oriented machine learning system.

Academic Positions

- **Clemson University** — Clemson, SC
  Graduate Research Assistant (Advisor: Dr. Yingjie Lao) — *Jan. 2018 - Present*

Education

- **Clemson University** — Clemson, SC
  Ph.D. in Electrical Engineering; — *Jan. 2018 - Dec. 2023*
- **Rochester Institute of Technology** — Rochester, NY
  Master of Science in Electrical Engineering; — *Aug. 2012 - Sep. 2014*
- **East China University of Science and Technology** — Shanghai
  Bachelor of Science in Electrical Engineering; — *Sep. 2008 - Jun. 2012*

Skills Summary

- **Skills**: Trustworthy AI, Deep Learning, Computer Vision, Adversarial Machine Learning, Model Compression, ASIC Design
- **Language & Tools**: Python, Pytorch, TensorFlow/Keras, Numpy, Scikit-learn, Pandas, Docker, Vim, Synopsys, HSPICE, VCS, Design Complier, Xilinx Vivado

Research Interest

My research interests include trustworthy AI, computer vision, adversarial machine learning, and hardware oriented machine learning systems.

Research Experience

- **Research Assistant** — Clemson, SC
  Clemson University — *Jan. 2018 - Present*
  - **Advisors**: Prof. Yingjie Lao
  - **Objective 1**: Exploit new adversarial attacks on deep neural network systems, featuring the design of an algorithm-hardware collaborative backdoor attack.
  - **Objective 2**: Develop algorithms that incorporate the hardware aspect into defense for enhancing adversarial robustness against vulnerabilities in the untrusted semiconductor supply chain.
  - **Objective 3**: Model recovery strategies as an innovative approach to mitigate hardware-oriented fault attacks in the untrusted user-space.
- **Deep Learning Research and Software Intern** — Remote
  Nvidia Corporation — *May. 2022 - Feb. 2023*
  - **Advisors**: Dr. José M. Álvarez and Dr. Zhiding Yu
  - **Objective 1**: Explore new training paradigms to improve the robustness of vision transformers against out-of-distribution scenarios and natural corruptions.
  - **Objective 2**: Enhance the performance of foundation deep learning models to facilitate the reliability and safety of the perception systems of autonomous vehicles.

Research Projects

- **Robust Vision Transformers for Perception Systems**
  Pytorch/Python/Shell — *May. 2022 — Mar. 2023*
  - Proposed a novel training paradigm that jointly incorporates self-emerging token labels and image-level labels and significantly enhanced clean accuracy and zero-shot robustness of Fully Attentioinal Networks on image classification and segmentation tasks.
  - Achieved SOTA zero-shot robustness on ImageNet-A, ImageNet-R and Cityscape-C with model size of 77.3M.
  - Experience with distributed training and parameter tuning of neural networks on GPU clustering such as NGC and Maglev.

- **Design for Deep Neural Networks Testing**
  Pytorch/Python *Jan. 2022 — Feb. 2023*
  - Developed a novel defensive framework for detecting hardware-oriented fault attacks against deep neural networks and recovering the models.
  - Achieve up to 94.76% detection success rate with only 140 test vectors on the CIFAR-10 dataset.

- **Machine Learning for Approximate Circuits Design**
  Verilog/Python/HSPICE *May. 2021 — Jan. 2023*
  - Proposed a new data-driven feature selection framework for approximate circuits design. The proposed method is applicable to both voltage over-scaling and approximate logic design and achieves a high compensated accuracy.

- **Robust DNNs against Poisoning Attacks**
  Pytorch/TensorFlow/Python *Sep. 2018 — May. 2022*
  - Devised a general and scalable defensive framework against clean-label backdoor attacks towards image classification tasks. Achieved up to 100% detection rate and reduced attack success rate from ∼90% to 0% against three widespread attacks.
  - Proposed a novel defense against poisoning attacks using gradual magnitude pruning. Analyzed the correlation between pruning and model robustness and improved the post-attack accuracy from 5% to over 50%.

- **Poisoning Attacks towards DNNs**
  Pytorch/TensorFlow/Python *Sep. 2018 — May. 2022*
  - Designed a generative adversarial net (GAN)-based framework for clean-label poisoned data generation that degrades the overall model accuracy.
  - Built the framework using BigGAN architecture and devised a triplet loss function to improve the effectiveness and fidelity of poisoned data.
  - Achieved 18% accuracy drop with only 20% poisoning ratio and 55% accuracy drop with full poisoning on modern neural networks such as ResNet, VGG and Inception-V3.
  - Proposed an innovative poisoning attack that manipulates the predictions of neural networks on a per-class basis.
  - Designed gradient-based and class-oriented algorithms to efficiently generate poisoned data at a large scale.

Publications

[1] **Bingyin Zhao** and Yingjie Lao. Resilience of pruned neural network against poisoning attack. In *2018 13th International Conference on Malicious and Unwanted Software (MALWARE)*, pages 78–83. IEEE, 2018.

[2] **Bingyin Zhao** and Yingjie Lao. Clpa: Clean-label poisoning availability attacks using generative adversarial nets. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 9162–9170, 2022.

[3] **Bingyin Zhao** and Yingjie Lao. Towards class-oriented poisoning attacks against neural networks. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3741–3750, 2022.

[4] **Bingyin Zhao** and Yingjie Lao. Ultraclean: A simple framework to train robust neural networks against backdoor attacks. 2023.

[5] **Bingyin Zhao**, Ling Qiu, and Yingjie Lao. Data-driven feature selection framework for approximate circuit design. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2023.

[6] **Bingyin Zhao**, Zhiding Yu, Shiyi Lan, Yutao Cheng, Anima Anandkumar, Jose Alvarez, and Yingjie Lao. Fully attentional networks with self-emerging token labeling. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2023.

[7] Antian Wang, **Bingyin Zhao**, Weihang Tan, and Yingjie Lao. Neural network fault attacks detection using gradient-based test vector generation. In *Proceedings of the 60th Annual Design Automation Conference*, 2023.