Jingyang Zhang

☎ (984)245-5792 • ✉ zhjy227@gmail.com • 🌐 zjysteven.github.io

Summary

Jingyang is a machine learning engineer with in-depth experience in designing and implementing advanced and robust training algorithms for **machine learning** models. During his Ph.D. study at Duke, he has worked on **adversarial robustness** and **out-of-distribution detection**. He also has hands-on experience with **diffusion models** and **multi-modal LLMs**. He combines *1)* outstanding research capabilities, with publications in top-tier ML conferences, and *2)* strong engineering skills, demonstrated through open-source implementations of ML models and algorithms.

Work Experience

- **Machine Learning Engineer @ Sciforium** *Jan 2025 - now*
  - Works as a full-stack ML engineer to develop latest multi-modal LLMs. Focuses on various aspects including data pipeline, model design and implementation, efficient scaling methods like MoE and quantization, and training infrastructure.
- **Machine Learning Intern @ Tesla** *Jun 2023 - Sep 2023*
  - Implemented and adapted state-of-the-art deep learning models for trajectory prediction. Showed the efficacy of this method over baselines with proof-of-concept experiments in different scenarios.
- **Machine Learning Research Intern @ Bosch Center for AI** *Jun 2022 - Dec 2022*
  - Developed a "universal" adversarial defense using diffusion model that is robust to both ℓ_p (digital) and patch (physical) adversarial attacks against images. Demonstrated the effectiveness and potential of the defense through extensive experiments, which resulted in a patent.

Education

- **Duke University (Durham, NC)** **Aug 2019 - Dec 2024**
  Ph.D., Dept. of Electrical and Computer Engineering *GPA: 3.96/4.0*
  Advisor: Prof. Hai (Helen) Li, Prof. Yiran Chen
- **Tsinghua University (Beijing, China)** **Sep 2015 - Jul 2019**
  B.Eng., Dept. of Electronic Engineering

Selected First-Author Publications

- **DVERGE: Diversifying Vulnerabilities for Enhanced Robust Generation of Ensembles**
  - Proposed DVERGE, a novel ensemble training methodology for Deep Neural Networks (DNNs) that diversifies the learnt features of sub-models. With little degradation in clean accuracy, DVERGE was once the state-of-the-art ensemble-based defense against black-box transfer attacks.
  - *Supported by DARPA QED-RML program and was accepted by NeurIPS'20 (oral).* [Paper][Code]
- **Privacy Leakage of Adversarial Training Models in Federated Learning Systems**
  - Developed a privacy attack such that for any user that performs adversarial training in a federated learning system, an attacker can eavesdrop to accurately reconstruct the user's private training images at scale (i.e., even when the training batch size is large).
  - *Accepted by CVPR'22 The Art of Robustness workshop (oral).* [Paper][Code]
- **Min-K%++: Improved Baseline for Detecting Pre-Training Data from Large Language Models**
  - Developed a novel membership inference attack for LLMs with theoretical insights that improves the detection rate by a large margin.
  - *Accepted by ICLR'25 (spotlight).* [Paper][Code]
- **Mixture Outlier Exposure: Towards Out-of-Distribution in Fine-Grained Environments**
  - Proposed MixOE, a new DNN training algorithm that leads to 4%-13% improvement in true negative rate in large-scale, fine-grained OOD detection.
  - *Supported by AFRL and was accepted by WACV'23.* [Paper][Code]

Selected Open-Source Projects

- **OpenOOD**
  - The largest development and evaluation codebase for out-of-distribution detection. 960+ stars. [Code]
- **lmms-finetune**
  - A lightweight codebase for easily fine-tuning vision LLMs (LLaVA, Qwen-VL, etc.) with custom user-specified data. 300+ stars. [Code]
- **VLM-Visualizer**
  - A tool that visualizes the attention of vision LLMs on the input image. 180+ stars. [Code]

Technical Skills

- Programming Languages: **Python**, C++, Matlab.
- Deep Learning Frameworks: **PyTorch**, **JAX**, TensorFlow.