Jiawei ZHANG

jiaweiz7@illinois.edu ◇ javyduck@gmail.com ◇ javyduck.github.io
(+1) 217-200-3511 ◇ 1007 W University Ave, Urbana, IL 61801, USA

Education

University of Illinois Urbana-Champaign (UIUC) Advisor: *Prof.* Bo Li
Master of Science in Computer Science (Research-Oriented), GPA: 4.0/4.0. *August. 2021 – May. 2023 (Expected)*

Zhejiang University (ZJU) Hangzhou, China
Bachelor of Engineering (Honor Class), GPA: 89.4/100 (Rank: 6/130). *Sept. 2017 – Jun. 2021*

Research Interest:

Security & Privacy, Trustworthy Machine Learning, Probabilistic Graphical Model, Explainable AI. Specifically, I aim to develop more certifiably reliable machine learning systems in areas like computer vision and reinforcement learning and try to incorporate more prior domain knowledge into these systems via PGM to make them more interpretable and controllable.

Publication

- **Jiawei Zhang**, Zhongzhu Chen, Huan Zhang, Chaowei Xiao, Bo Li. DiffSmooth: Certifiably Robust Learning via Diffusion Models and Local Smoothing. *32th USENIX Security Symposium 2023.* (Under Review) [draft]
- **Jiawei Zhang**, Linyi Li, Ce Zhang, Bo Li. CARE: Certifiably Robust Learning with Reasoning via Variational Inference. *IEEE Conference on Secure and Trustworthy Machine Learning (SatML) 2023.* [arxiv]
- Zhuolin Yang*, Zhikuan Zhao*, Boxin Wang, **Jiawei Zhang**, Linyi Li, Hengzhi Pei, Bojan Karlas, Ji Liu, Heng Guo, Ce Zhang, Bo Li. Improving Certified Robustness via Statistical Learning with Logical Reasoning. *Advances in Neural Information Processing Systems (NIPS) 2022.* [arxiv]
- Linyi Li, **Jiawei Zhang**, Tao Xie, Bo Li. Double Sampling Randomized Smoothing. *International Conference on Machine Learning (ICML) 2022.* [arxiv]
- **Jiawei Zhang***, Linyi Li*, Huichen Li, Xiaolu Zhang, Shuang Yang, Bo Li. Progressive-Scale Boundary Blackbox Attack via Projective Gradient Estimation. *International Conference on Machine Learning (ICML) 2021.* [arxiv]

Research Experience

Safety-Critical Driving Scenario Generation *Oct. 2022 – Present*
Research Assistant | Secure Learning Lab, UIUC *Advised by Prof. Bo Li and Prof. Ding Zhao (CMU)*

- Aim to enrich the safety-critical testing scenarios in SafeBench [link] for Autonomous Vehicles.
- Train an adversarial agent (vehicle/pedestrian/bicyclist) via specifically designed multi-agent reinforcement learning to cause the unexpected collision of the ego vehicle.

Enhance Robustness via Diffusion Models and Local Smoothing *Jun. 2022 – Oct. 2022*
Research Assistant | Secure Learning Lab, UIUC *Advised by Prof. Bo Li & Postdoc. Huan Zhang (CMU)*

- Prove that the "one-shot" denoising of DDPM can approximate the mean of the generated posterior distribution by continuous-time diffusion models, which is an approximation of the original instance under mild conditions.
- Propose a local smoothing technique based on the diffusion models, achieve the SOTA 43.6% certified accuracy on CIFAR-10 under ℓ_2 radius 1.0 and the SOTA 53.0% certified accuracy on ImageNet under the ℓ_2 radius 1.5.

Certifiably Robust Learning with Reasoning via Variational Inference *May. 2022 – Sep. 2022*
Research Assistant | Secure Learning Lab, UIUC *Advised by Prof. Bo Li & Prof. Ce Zhang (ETH Zürich)*

- Propose a scalable and certifiably robust learning with reasoning pipeline CARE, which is able to integrate knowledge rules to enable reasoning ability for reliable prediction
- Propose an efficient Expectation Maximization (EM) algorithm to approximate the reasoning based on Markov Logic Network (MLN) via variational inference using Graph Convolutional Network (GCN).
- Extensive experiments on different datasets show that the proposed method achieves significantly higher certified robustness than SOTA baselines, for example, the certified accuracy could be improved from 36.0% (SOTA) to 61.8% under ℓ_2 radius 2.0 on AwA2.

Boundary Blackbox Attack via Projection Based Gradient Estimation *Jun. 2020 – May. 2021*
Research Intern | Cooperate with Ant Financial *Advised by Prof. Bo Li*

- Propose the first theoretical framework to analyze boundary blackbox attacks with general projection functions.
- Characterize the key characteristics and trade-offs for a good projective gradient estimator.
- Propose Progressive-Scale based projective Boundary Attack via progressively searching for the optimal scale in a self-adaptive way under spatial, frequency, and spectrum scales.
- The extensive experiments show that our method outperforms the state-of-the-art boundary attacks on MNIST, CIFAR-10, CelebA, and ImageNet against different blackbox models and an online API (MEGVII Face++).

Short-Term Traffic Flow Prediction *Apr. 2019 – Apr. 2020*
Research Assistant | Huawei 2012 lab & ZJU Intelligent Transportation Systems lab *Advised by Prof. Xiqun Chen*

- Conclude the limitations of the traditional time series prediction model like LSTM, DBN & CNN.
- Propose a state-of-the-art short-term traffic speed prediction method which utilizes Graph Convolutional Network (GCN) to extract the topological relation of the Maryland roads in space domain while combined with the speed correlation between road and road extracted by Transformer.

6DoF Pose Estimation *Sep. 2018 – Apr. 2019*
Research Intern | State Key Laboratory of CAD&CG, ZJU *Advised by Prof. Xiaowei Zhou*

- Use Unity to build a aircraft carrier deck and estimate the 6DoF pose of the moving planes on it with PVNet.
- Employ a coarse-to-fine prediction scheme to predict per voxel likelihoods for each human joint by ConvNet.

Selected Course Projects

Dynamic Matching *Aug. 2022 – Nov. 2022*
Team Leader | Course - Algorithmic Game Theory, UIUC *Advised by Prof. Ruta Mehta*

- Explore the mechanisms for the dynamic matching on ride-sharing systems like Uber and Lyft.
- Delve into the general BATCHING mechanisms and explain how could we introduce stochastic future information into the current batch matching in a bipartite graph.

Verification of Neural Networks Based on DeepPoly *Sep. 2021 – Nov. 2021*
Team Leader | Course - Logic and Artificial Intelligence, UIUC *Advised by Prof. Gagandeep Singh*

- Propose a much mores efficient certification method based on the widely used bound-propagation algorithm DeepPoly (CROWN-IBP) via a recursive refinement of the linear constraints.
- Achieve the *Top-1* certification performance in class, and achieve 100% verification accuracy on all testing cases.

Natural Language to SQL to Visual *Sep. 2019 – Jan. 2020*
Team Leader | Course - Distributed System, ZJU *Advised by Prof. Renjun Xu*

- Propose a quasi-PageRank algorithm to extract key textual data which crawled from various mainstream websites with conducting sentimental analysis and importance sequencing in the meantime.
- Establish a SQL database with respect to stock market and financial realm, supporting online natural language searching in Chinese, which could be converted to SQL language automatically in the back-end and presented by visualized charts for better demonstration.

Teaching

CS 307 - Modeling and Learning in Data Science (Spring 2022)
- Course Assistant with Prof. Bo Li and Prof. David Forsyth

Selected Honor & Awards

Meritorious Winner, MCM COMAP's Mathematical Contest in Modeling *Apr. 2020*
First Prize, China Harbour Scholarship (1/130) *Jan. 2020*
Second Class Scholarship, ZJU Merit-based Scholarship *Jan. 2020*
First Prize, Zhejiang Intelligent Construction and Management Innovation Competition *Dec. 2019*
Second Prize, Gad Scholarship *May. 2019*
Honorary Title of Model Student for Academic Excellence, ZJU-wide *Sept. 2018/19/20*
First Prize, The Chinese Mathematics Competitions (non-math major), Zhejiang Province *Nov. 2018*
First Prize, National High School Mathematics League, Zhejiang Province *Nov. 2016*

SKILLS & OTHERS

Computer Languages	Python, C/C++, Java, Matlab, Shell script, Markdown
Frameworks & Packages	PyTorch, TensorFlow, OpenCV, MySQL, Hadoop, Unity, SPSS