YUHAO MAO

School of Mathematical Science, Zhejiang University, P.R. China
+86 13757408769 | e: myh821746176@outlook.com

EDUCATION

Zhejiang University Hangzhou, China
Applied Mathematics & Finance September 2017 – Present

- GPA: 3.94/4.00
- Selected Courses: Mathematical Analysis II (98), Ordinary Differential Equations (98), Stochastic Processes (96), Multivariate Statistical Analysis (97), Advanced Data Structure & Algorithm Analysis (91).
- Selected Awards: 2017-2018 and 2018-2019 Provincial Scholarship
- Admitted to Cho Kochen Honors College (5% freshmen admitted annually)

RESEARCH EXPERIENCE

The Helmholtz Center for Information Security Saarland, Germany
Research Assistant to Professor Yang Zhang August 2020 – present
Topic: Exploiting unforgettability of neural networks to inject a backdoor without poisoning

- Most successful existing backdoor attacks to neural networks involve poisoning, which run a multi-object optimization to make models deceive users when a trigger is present. We illustrate that without poisoning, the feature extractors of neural networks already preserve enough irrelevant information which can be further exploited to inject a neuron-level backdoor.
- Independent research, under the supervision of Professor Zhang, to develop a neuron-level ensemble-based classifier with careful mathematical design. 100% separability achieved with only ten neurons and a trigger consisting of eleven pixels in the dataset CIFAR10. We are currently exploring more extensions and applications on unforgettability.

The Helmholtz Center for Information Security Saarland, Germany
Research Assistant to Professor Yang Zhang July 2020 – August 2020
Topic: Efficient and Generalized Artificial Brain Stimulation for Detecting Backdoors in Neural Networks

- Recent studies have shown that a neural network can be trojaned to inject backdoors, *i.e*. a normal behavior on benign inputs but making malicious decisions when a trigger is invoked. This project improved a recently proposed novel defense enabling it to work efficiently on large networks (scanning and analysis currently takes hours). This work generalized techniques that, until now, only studied a limited range of backdoors, and has been used to study standardly trained models to discuss natural backdoors.
- Independent research, under the supervision of Professor Zhang, to accelerate the technique on large networks by quadratic magnitudes using layer-subsampling (*e.g*. approximately 100x for ResNet-50 without loss of detection precision).

Zhejiang University, College of Computer Science and Technology Hangzhou, China
Research Assistant to Professor Shouling Ji, ZJU 100-Young Professor December 2019 – June 2020
Title: Transfer Attacks Revisited: A Large-Scale Empirical Study in Real Settings

- Neural networks are vulnerable to crafted inputs, known as adversarial examples (AEs), that possess a mysterious property called transferability, *i.e*. AEs crafted to fool one network are likely to also fool another independent model. This project studies that property in real-world settings, contrasting and extending previous studies set in simplistic and unrealistic lab settings.
- We empirically consider the following questions: 1) Are real systems vulnerable to transfer attacks? 2) Which attack transfers better in real settings? 3) How is the transferability influenced by surrogate settings? 4) How do sample-level properties contribute to the transferability? To emphasize, we overturn two conclusions previously made in lab settings and extend many others.
- First author to a paper currently in press, conducting the majority of experiments, analyses, visualizations and paper preparation.

Zhejiang University, College of Computer Science and Technology Hangzhou, China
Research Assistant to Professor Shouling Ji, ZJU 100-Young Professor August 2019 – November 2019
Topic: Noncentral and Nonuniform Robustness Certification for Neural Networks

- Attacking and defending neural networks via adversarial examples is a thriving area of research. One way to eliminate this endless competition is to provide theoretical bounds of robustness. This project extends former certification methods to noncentral and nonuniform cases, based on which we compare and discuss the robust space of raw models and robust models and find an interesting correlation between interpretability of the model and its robust space.
- Assistant to a graduate student, actively participating in critical analysis and coding for the interpretation of results.

Massachusetts Institute of Technology, Department of Electrical Engineering & Computer Science Boston, USA
Machine Learning Summer Program July 2019

- Participated in immersive academic courses in traditional machine learning, deep learning, and reinforcement learning. Led a diverse and multi-institution team of students to complete a project on artist style classification.
- Utilizing ensemble learning, we developed a model that achieved almost the same performance as a state-of-art model but with much simpler architecture. With additional data augmentation skills, we were awarded best team performance and a final score of 96/100.

ADDITIONAL INFORMATION

Additional Professional and Extracurricular Experiences

- Delivered oral presentation and attend lectures as one of four specially invited students at the *4th Annual Honors International Faculty Institute Workshop* at Texas Christian University, USA, in June 2019.
- Spend approximately 200 hours volunteering as an undergraduate student, including as an assistant for the *11th International Chinese Statistics Association (ICSA) International Conference*.

Interests

- Teaching math classes at elementary schools, high schools and as a part-time calculus tutor to first-year university students.
- During the first half of 2020, achieved a 20% yield rate in the funds that I invested in.
- Member of the CKC College volleyball team for the past three years.

Computer and Language Skills

- Fluent: Python (95/100 course score, multiple course and research projects in Python, familiar with Pytorch, Pandas, Matplotlib, etc.), Latex, C, R, Markdown
- Experienced: MATLAB, HTML