YUHAO MAO

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

EDUCATION

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

- GPA: 3.96/4.00
- Selected awards: 2017-2018 and 2018-2019 Provincial Scholarship
- Admitted to CKC Honors College (only 5% out of all freshmen are admitted every year)

RESEARCH EXPERIENCE

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

- It is known in recent years that a neural network can be trojaned to inject backdoors, i.e. having a normal behavior on benign inputs but making malicious decisions when a trigger is invoked. This project aims to improve one novel defense proposed recently so that it can work efficiently on large networks which currently takes hours to scan and analyze. Besides, current techniques only study a limited range of backdoors, which we aim to generalize. We will use the developed technique to study standardly trained models to discuss natural backdoors.
- Already speed up the technique on large networks by quadratic magnitudes using careful discussion and 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

- It is known that neural networks are vulnerable to crafted inputs called adversarial examples (AEs) and those AEs somehow possess a mysterious property called transferability: AEs crafted to fool one network are probable to fool another independent model as well. This project studies that property in the real settings while all previous findings are drawn in the simple and unrealistic lab settings and found many insights that are refinements or corrections to previous conclusions. Specifically, from a theoretical perspective we study 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?
- First author, in collaboration with another student in a leading position under supervision of multiple professors. Design and undertake most of the experiments, do all the result analysis, prepare most of the visualization, write the paper draft solely and actively participate in revising and polishing. The paper is being in submission of USENIX 2021 and is currently under second round of review.

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 has been a prospering area of research and one way to eliminate that endless competition is to provide theoretical bounds of robustness. This project aims to generalize former methods to noncentral and nonuniform case. Based on the generalized method, we compare and discuss robustness space of raw models and robust models and find that robustness space of robust models exhibits some interpretability of that model.
- As an assistant to a graduate student, actively participate in brainstorm and coding and provide valuable clues for interpreting the result.

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

- Attend academic courses and get experiences about traditional machine learning, deep learning and reinforcement learning. Lead a team consisting of students from different backgrounds and universities to finish a project considering artist style classification. Utilizing ensemble learning, we develop a model which achieves almost the same performance as state-of-art model with much simpler architecture. With additional data augmentation skills, we achieve best performance among all teams.
- Get a final score 96 out of 100.

ADDITIONAL INFORMATION

Additional Professional and Extracurricular Experiences

- Speak, give presentation and attend lectures as one of four students specially invited to the 4th Annual Honors International Faculty Institute Workshop at Texas Christian University (Texas, the U.S.) in June, 2019.
- Spend about 200 hours on volunteering during undergraduate.

Interests

- As a math and finance double-major, I love math education and financial analysis. I have taught math classes at elementary schools, high schools and served as a part-time Calculus tutor to first-years. During the first half year of 2020, I achieved a 20% yield rate in my investment in the funds.
- I love volleyball too and have been a volleyball team member of the college for three years.

Computer and Language Skills

- Extremely familiar: Python (got 95/100 in the course, wrote multiple course and research project in Python, familiar with Pytorch), Latex, C, R
- Familiar: MATLAB, HTML, Markdown