Defense Frameworks Against Adversarial Attacks For Deep Learning Models

Deep Neural Networks (DNNs) models have been used successfully in various tasks. However, they are particularly susceptible to adversarial attacks, that some tiny changes in the input’s value will cause incorrect predictions and raise security risk to real-world deep learning applications, e.g., autonomous driving and surveillance systems. One direction in adversarial defense approaches is to reduce the adversarial noise from the model input. In this talk, Zhixun will present his two solutions for adversarial noise reduction and defense framework against adversarial attacks. He will also present empirical results which show that his methods outperform the state-of-the-art results in terms of the deep learning models’ prediction accuracy under adversarial attacks and their consistency across wide range of adversarial noise levels.

Date: Friday, March 17, 2023  
Time: 12:00 PM  
Location: Zoom  
Zoom link: https://ucm.edu/EECS03-17

Biography

Zhixun "Jason" He, is a Ph.D. candidate in EECS department at UC Merced, advised by Prof. Mukesh Singhal. He received B.E. degree from Nanjing Tech University (Nanjing, China). Zhixun’s research is in the domain of computer vision, with focus on adversarial defense for deep learning models. Specifically, his recent work includes adversarial noise reduction using generative network based approach, domain adaptation and probability theory. His publications appeared in MVA, PRML and IROS etc.