מכון ויצמן למדע

THE WEIZMANN INSTITUTE OF SCIENCE

FACULTY OF MATHEMATICS AND COMPUTER SCIENCE

Machine Learning and Statistics Seminar

Room 1 ,Ziskind Building
on Wednesday, Apr 10, 2019
at 11:15

Yossi Keshet
BIU

Deep Learning: Vulnerabilities, Defenses and Beyond

Abstract:

Deep learning has been amongst the most emerging fields in computer science and engineering. In recent years it has been shown that deep networks are vulnerable to attacks by adversarial examples. I will introduce a novel flexible approach named Houdini for generating adversarial examples for complex and structured tasks and demonstrate successful attacks on different applications such as speech recognition, pose estimation, semantic image segmentation, speaker verification, and malware detection. Then I will discuss how this weakness can be turned into two secure applications. The first is a new technique for watermarking deep network models in a black-box way. That is, concealing information within the model that can be used by the owner of the model to claim ownership. The second application is a novel method to Speech Steganography, namely hiding a secret spoken message within an ordinary public spoken message. I will conclude the talk by a brief discussion of our attempts to detect such adversarial attacks, based on multiple semantic label representations.