BlurNet: Defense by Filtering the Feature Maps

Ravi Raju

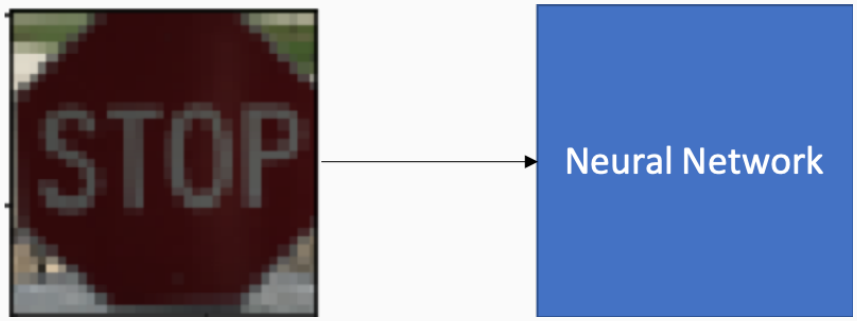
September 25, 2019

Outline

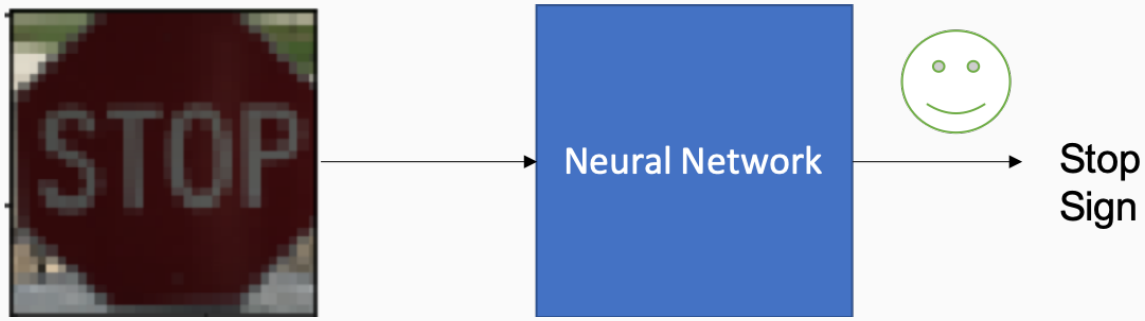
Introduction

Introduction

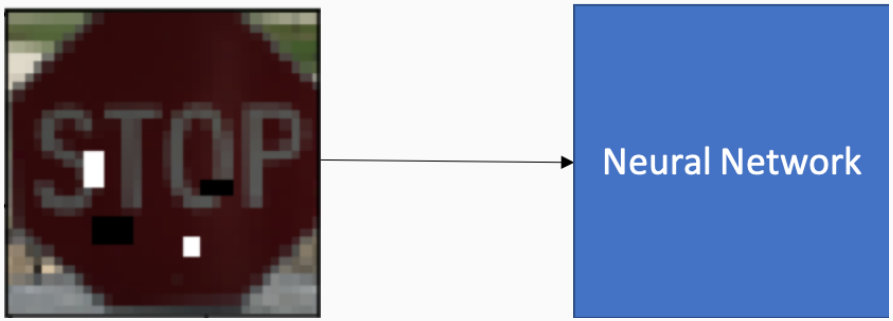
Vulnerabilities in NNs



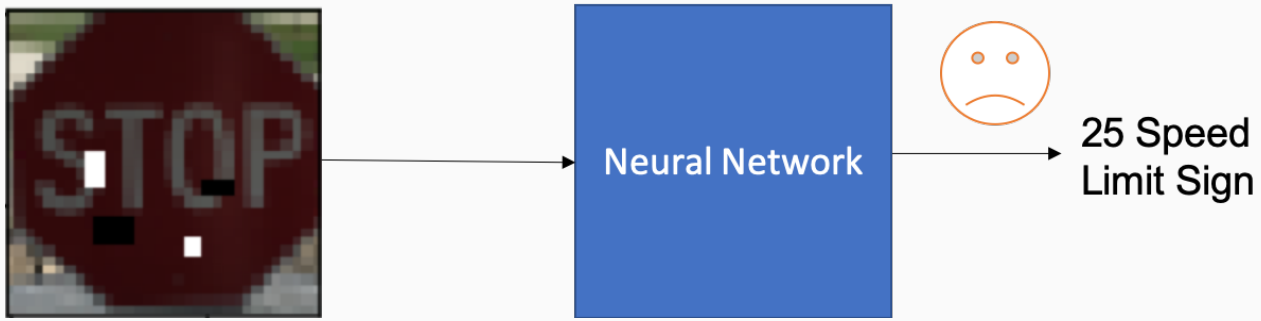
Vulnerabilities in NNs



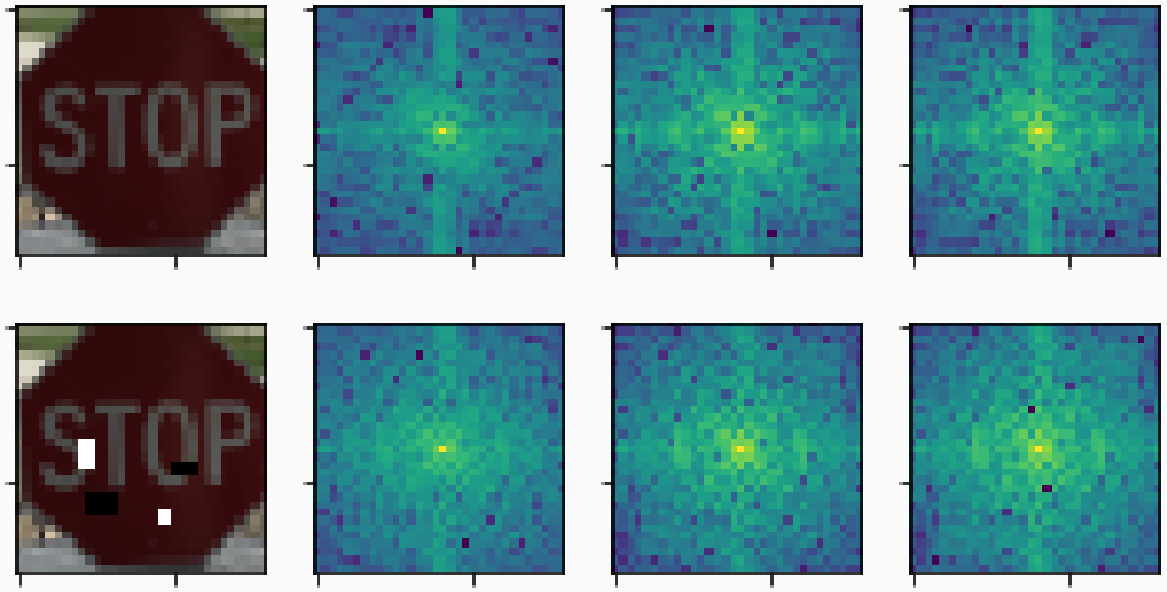
Vulnerabilities in NNs cont.



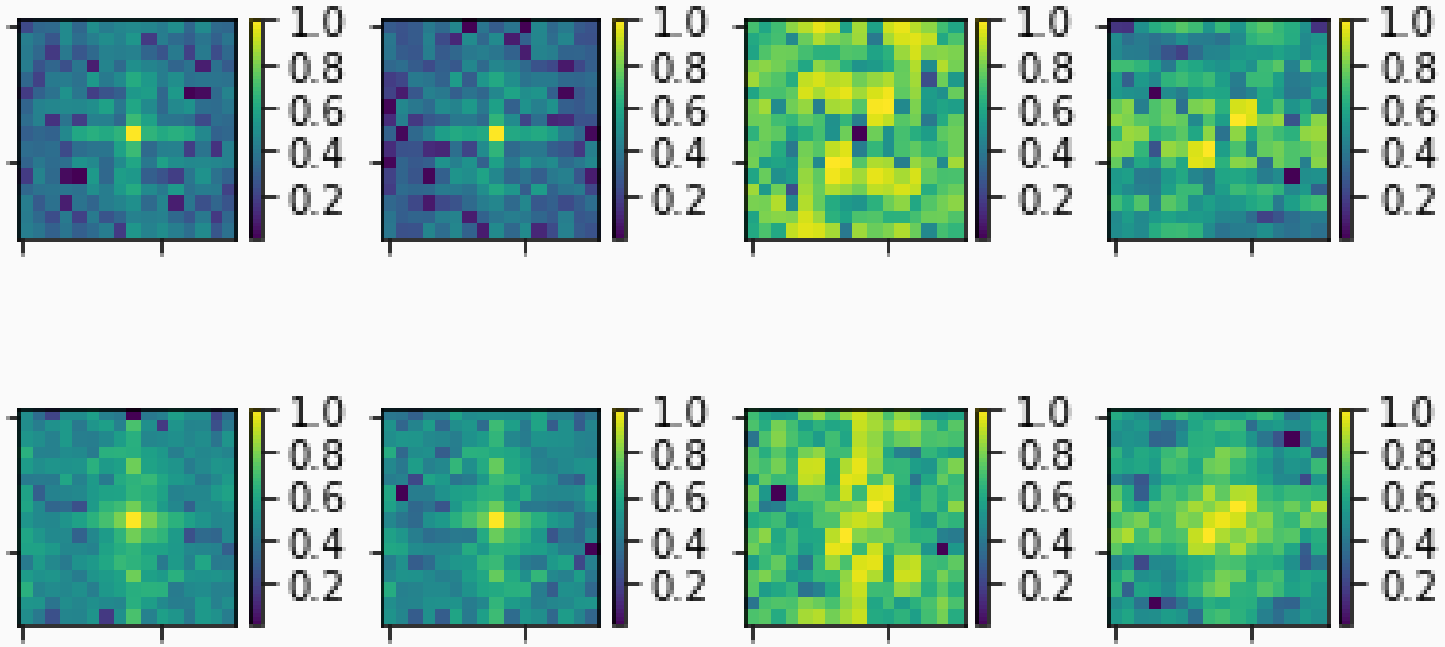
Vulnerabilities in NNs cont.



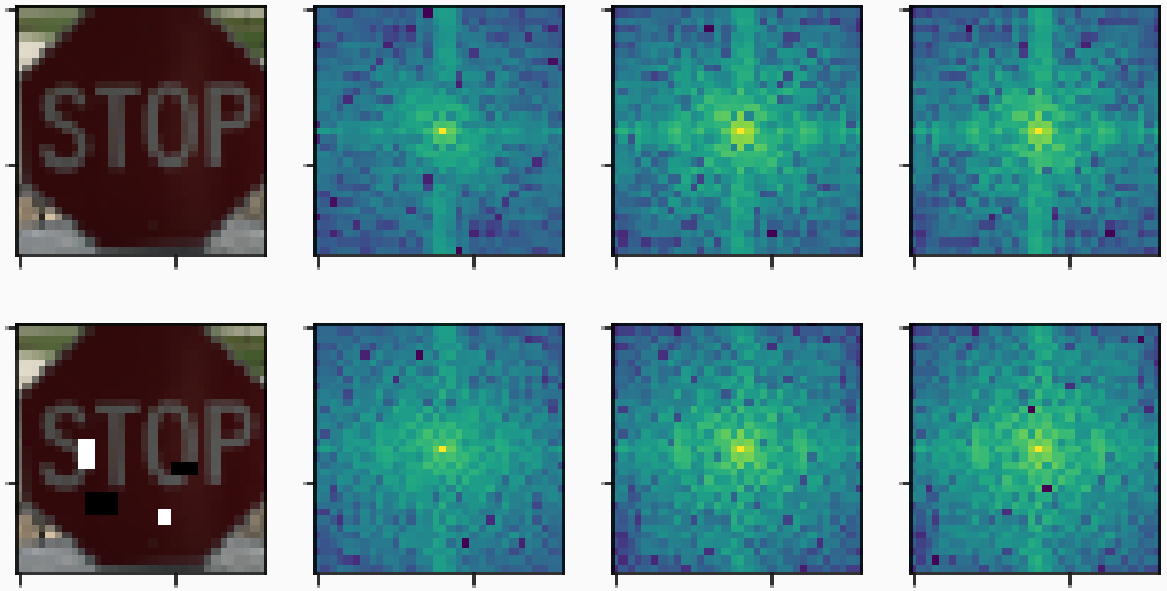
FFT Spectrum of channels



FFT of First Layer



FFT Spectrum of channels



L2 vs Attack Plot