Bo Huang

1995-04-08
(+86)182-0013-3822
bhuangas@connect.ust.hk

Education Background

2022.02-present **The Hong Kong University of Science and Technology (Guangzhou)** **PhD DSA**

Research interests: Adversarial machine learning

2017.09 – 2020.06 **Shenzhen University** **MEng CS**

Major courses: Machine Learning, Object-Oriented Technology and Methods, Theory of Algorithms, Combinatorial Mathematics

Master's thesis: Defense Methods Against Adversarial Examples in Deep Learning

2013.09 - 2017.07 **University of Electronic Science and Technology of China** **BSci Biotechnology**

Publications

- **Bo Huang**, Zhiwei Ke, Yi Wang, Wei Wang, Linlin Shen, Feng Liu. Adversarial Defense by Diversified Simultaneous Training of Deep Ensembles. (**AAAI-2021**)
- **Bo Huang**, Yi Wang, and Wei Wang. Model-Agnostic Adversarial Detection by Random Perturbations. (**IJCAI-2019**)
- Yi Wang and **Bo Huang**, Adversarial example detection method and apparatus, computing device, and non-volatile computer-readable storage medium, **United States Patent**, 2020.07.27, **Patent No**. 10936973.

Research Experience

2020.11-present **Certified robustness in Deep ensembles**

- We conduct research on how to approximately calculate the certified robustness of deep ensembles.

2019.08-2020.10 **Adversarial defense by diversified simultaneous training of deep ensembles**

- A novel strategy of diversified learning of high-level feature representations by ensemble networks was proposed;
- Two regularization schemes in simultaneous training to facilitate the proposed diversified learning were developed;
- Three measures of ensemble diversity were analyzed for adversarial defense in deep ensembles.

2019.03-2019.07 **A comparative study of the robustness of single and ensemble model**

- We investigated the robust performance of ensemble DNNs based on traditional ensemble methods;
- Ensemble SVM was firstly found to be less robust than single SVM in the black-box attack scenario;
- We proposed the concept of gradient correlation, which can be used to evaluate the adversarial robustness;

2018.07-2019.02 **Model-Agnostic Adversarial Detection by Random Perturbations**

- We proposed an effective adversarial detection method based on statistical analysis of model responses;
- A theoretical analysis by relating the bound of random perturbations to the adversarial distortions was given;

Awards

- The First Prize of Postgraduate Entrance Scholarship
- Outstanding Graduate

Professional Skills

- **English:** IELTS **7.0**, CET-6;
- **Programming Language:** Python, Matlab, C, HTML;
- **OS&Platform:** Windows, Linux, Docker;
- **Tools:** Pytorch, Keras, Tensorflow, Cleverhans;