Vikash Sehwag

Ph.D. Candidate
Princeton University, Princeton, NJ 08544

(609)-216-6036
vsehwag.github.io vvikash@princeton.edu

Research Interest

I am interested in research problems at the intersection of *security, privacy, and machine learning*. Some topics I have worked on are adversarial robust supervised/self-supervised learning, adversarial robustness in compressed neural networks, open-world machine learning, self-supervised detection of outliers, adversarial robust outlier detection, privacy leakage in large scale deep learning, and faster-federated learning.

Education

Program	Institution	Years
Ph.D., Electrical Engineering *Advisors – Prateek Mittal, Mung Chiang*	Princeton University NJ, USA	2017 - Present
M.A., Electrical Engineering	Princeton University NJ, USA	2017 - 2019
B.Tech., Electronics and Electrical Communication Engg.	Indian Institute of Technology (IIT) Kharagpur, INDIA	2013 - 2017

Honors and Awards

- Winner of Qualcomm Innovation Fellowship, North America Region 2019
- Received a departmental nomination for Microsoft Research PhD Fellowship 2019
- Received best undergraduate thesis award (1 from 72 students) at IIT Kharagpur 2017
- IEEE student award from IEEE student branch of IIT Kharagpur 2016
- Awarded the WISE scholarship from German Academic Exchange Service (DAAD), Germany 2016
- Received Merit-cum-Means Scholarship from MHRD, Government of India 2013-17

Publications

Preprints and papers under review

- SSD: A Unified Framework for Self-Supervised Outlier Detection
  Vikash Sehwag, Mung Chiang, Prateek Mittal
  Preprint under review at International Conference on Learning Representations (**ICLR**), 2021
- AdvBench: Tracking the Progress in Adversarial Robustness
  Francesco Croce*, Maksym Andriushchenko*, **Vikash Sehwag***, Nicolas Flammarion, Mung Chiang, Prateek Mittal, Matthias Hein
  Arxiv preprint, 2020
- Beyond ℓ_p Norms: Delving Deeper into Robustness to Physical Image Transformations
  Vikash Sehwag, Jay Stokes, Cha Zhang
  Under review at AAAI 2021
- Fast-Convergent Federated Learning
  Hung T. Nguyen, **Vikash Sehwag**, Seyyedali Hosseinalipour, Christopher G. Brinton, Mung Chiang, H. Vincent Poor
  Preprint under review at IEEE JSAC Series on Machine Learning for Communications and Networks
- PatchGuard: Provable Defense against Adversarial Patches Using Masks on Small Receptive Fields
  Chong Xiang, Arjun Nitin Bhagoji, **Vikash Sehwag**, Prateek Mittal
  Arxiv preprint, 2020

* refers to equal contribution.

- Towards compact and robust deep neural networks
  Vikash Sehwag*, Shiqi Wang*, Prateek Mittal, Suman Jana
  Arxiv preprint, 2019

Peer-reviewed papers

- HYDRA: Pruning Adversarially Robust Neural Networks
  Vikash Sehwag, Shiqi Wang, Prateek Mittal, Suman Jana
  Neural Information Processing Systems (**NeurIPS**), 2020 (*to appear*)
- Time for a Background Check! Uncovering the impact of Background Features on Deep Neural Networks
  Vikash Sehwag, Rajvardhan Oak, Mung Chiang, Prateek Mittal
  ICML workshop on Object-Oriented Learning, 2020
- On separability of self-supervised representations
  Vikash Sehwag, Mung Chiang, Prateek Mittal
  ICML workshop on Uncertainty & Robustness in Deep Learning, 2020
- On Pruning Adversarially Robust Neural Networks
  Vikash Sehwag, Shiqi Wang, Prateek Mittal, Suman Jana
  ICLR workshop on Towards Trustworthy ML, 2020
- Analyzing the robustness of open-world machine learning
  Vikash Sehwag*, Arjun Nitin Bhagoji*, Liwei Song*, Chawin Sitawarin, Daniel Cullina, Mung Chiang, Prateek Mittal
  In Proceedings of the 12th ACM Workshop on Artificial Intelligence and Security (**AISec**), 2019
- Not All Pixels are Born Equal: An Analysis of Evasion Attacks under Locality Constraints
  Vikash Sehwag, Chawin Sitawarin, Arjun Nitin Bhagoji, Arsalan Mosenia, Mung Chiang, Prateek Mittal
  Poster at ACM SIGSAC Conference on Computer and Communications Security (**CCS**), 2018.

Work Experience

- Summer Research Internship – Microsoft Research, Redmond, USA — Summer 2019
  Advisors – Jay Stokes, Cha Zhang
  Project: Adversarial attacks and defenses beyond ℓ_p norms
- Research Assistant – IIT Kharagpur, India — Fall 2016
  Advisors – Indrajit Chakrabarti, Santanu Chattopadhyay
  Project: Implementing physical unclonable functions with Network-on-chip routers
- Summer Research Internship – Technische Universität Darmstadt, Germany — Summer 2016
  Advisor – Heinz Koeppl
  Project: A study of stochastic SIS disease spreading on random graphs

Academic services

Teaching and Metoring

- Taught a mini-course on adversarial attacks & defenses — Wintersession 2020
- Teaching assistant for ELE 535: Machine Learning and Pattern Recognition — Fall 2019
- Mentoring Princeton undergraduates for their senior independent research work
  - Tinashe Handina (B.S.E., Electrical Engineering 2021)
  - Matteo Russo (B.S.E., Computer Science 2020)

Other Services

- One of three core maintainers of Adversarial Robustness Benchmark (advbench.github.io) — 2020
- Volunteered as junior mentor at Princeton-OLCF-NVIDIA GPU Hackathon — 2020
- Reviewer for ACM Transactions on Privacy and Security (TOPS), PLOS One — 2019, 2020
- Sub-reviewer for USENIX Security — 2018, 2019