Yash Sharma

ysharma1126@gmail.com
http://www.yash-sharma.com
https://scholar.google.com/citations?user=AlGCn8wAAAAJ&hl=en
http://www.github.com/ysharma1126
http://www.kaggle.com/ysharma1126
http://www.linkedin.com/in/yashjsharma

Education

Sep. 2014 - May. 2018 **Bachelor & Master in Electrical Engineering** - Cooper Union
New York, NY, USA
Thesis: "Gradient-based Adversarial Attacks in Limited Access Settings"
Masters GPA: 3.8/4.0

Organized and Taught the Saturday STEM Program, which exposes high-school students from underrepresented groups to engineering and coding.

Research Experience

Sep. 2018 - Present **Borealis AI** - *ML Researcher*, Toronto, ON, Canada

- Working on developing more robust deep learning training methods [WSY+18], understanding adversarial examples [SWB18], and other research topics.

Jun. 2017 - Aug. 2017 **IBM Research** - *Research Intern*, Yorktown Heights, NY, USA

- Worked with the AI Foundations Learning Group on adversarial examples.
- Improved the state-of-the-art in black-box attacks (can query target model) and transfer attacks (cannot access target model) with [ZCS+17] and [SCZ+18].

Technical Skills

Programming	Python, C++, Java
Machine Learning	TensorFlow, PyTorch, Keras, Scikit-Learn

Honors

2018 | **CAAD Competition Winner**
2018 | **DEFCON 26 Presenter**
2017 | **Kaggle Competitions Master**
2016 | **IBM Blockchain Hackathon Winner**
2016 | **CodeSuisse Winner**
2015 | **HackRU Prize**
2014 | **Half-Tuition Scholarship**

Publications

Conference Proceedings

[SAE+18] Y. Sharma, M. Alzantot, A. Elgohary, B.-J. Ho, M. Srivastava, and K.-W. Chang. "Generating Natural Language Adversarial Examples". In *Conference on Empirical Methods in Natural Language Processing (EMNLP); NIPS Workshop on Security in Machine Learning Encore Track*. https://arxiv.org/abs/1804.07998. 2018.

[SCZ+18] Y. Sharma, P.-Y. Chen, H. Zhang, J. Yi, and C.-J. Hsieh. "EAD: Elastic-Net Attacks to Deep Neural Networks via Adversarial Examples". In *AAAI Conference on Artificial Intelligence (Oral)*. https://arxiv.org/abs/1709.04114. 2018.

[ZCS+17] H. Zhang, P.-Y. Chen, Y. Sharma, J. Yi, and C.-J. Hsieh. "ZOO: Zeroth Order Optimization based Black-box Attacks to Deep Neural Networks without Training Substitute Models". In *ACM Workshop on Artificial Intelligence and Security (AISec) (Best Paper Nominee)*. https://arxiv.org/abs/1708.03999. 2017.

Workshops

[LBS18] Y. Li, J. Bradshaw, and Y. Sharma. "Are Generative Classifiers More Robust to Adversarial Attacks?" In *International Conference on Machine Learning (ICML) Theoretical Foundations and Applcations of Deep Generative Models Workshop (Oral)*. https://arxiv.org/abs/1802.06552. 2018.

[SC18a] Y. Sharma and P.-Y. Chen. "Attacking the Madry Defense Model with L1-based Adversarial Examples". In *International Conference on Learning Representations (ICLR) Workshops*. https://arxiv.org/abs/1710.10733. 2018.

Manuscripts

[ASC+18] M. Alzantot, Y. Sharma, S. Chakraborty, and M. Srivastava. "GenAttack: Practical Black-box Attacks with Gradient-Free Optimization". *arXiv preprint arXiv:1805.11090* (2018). https://arxiv.org/abs/1805.11090.

[PFC+18] N. Papernot, F. Faghri, N. Carlini, I. Goodfellow, R. Feinman, A. Kurakin, C. Xie, Y. Sharma, et al. "Technical Report on the CleverHans v2.1.0 Adversarial Examples Library". *arXiv preprint arXiv:1610.00768* (2018). https://arxiv.org/abs/1610.00768v6.

[SLA18] Y. Sharma, T.-D. Le, and M. Alzantot. "CAAD 2018: Generating Transferable Adversarial Examples". *arXiv preprint arXiv:1810.01268* (2018). https://arxiv.org/abs/1810.01268.

[SC18b] Y. Sharma and P.-Y. Chen. "Bypassing Feature Squeezing by Increasing Adversary Strength". *arXiv preprint arXiv:1803.09868* (2018). https://arxiv.org/abs/1803.09868.

[SWB18] Y. Sharma, G. Weiguang Ding, and M. Brubaker. "On the Effectiveness of Low Frequency Perturbations". *To appear on arXiv.* (2018).

[WSY+18] G. Weiguang Ding, Y. Sharma, K. Yik Chau Liu, and R. Huang. "Max-Margin Adversarial (MMA) Training: Direct Input Space Margin Maximization through Adversarial Training". *arXiv preprint arXiv:1812.02637* (2018). https://arxiv.org/abs/1812.02637.

Projects

Oct. 2018 — **NIPS 2018 Adversarial Vision Challenge**
Top-10 Participant; 5th in Targeted Attack
Participated in competition pitting submitted adversarial attacks against submitted defenses. Maximizing score requires minimizing L_2 distortion on TinyImageNet. Allowed limited number of queries. Placed 5th in Targeted Attack competition, invited to present at workshop.

June. 2018 - Aug. 2018 — **Competition on Adversarial Attacks and Defenses (2018)**
Overall Competition Winner
Participated in competition pitting submitted adversarial attacks against submitted defenses. Enforced L_∞ distortion constraint on ImageNet. Achieved **1st**, **1st**, and **3rd** in the Targeted Attack, Non-Targeted Attack and Defense competitions, respectively. Prize: $38,000.

Sep. 2017 - May. 2018 — **Lane Keeping and Navigation Assist System**
IEEE Student Paper + Senior Project, 2018
Built a miniature autonomous vehicle which can navigate through maps consisting of various road topologies. System is comprised of a perception module, for detecting lanes and intersections, and a control module, for lane keeping and turn making.

Aug. 2017 - Oct. 2017 — **NIPS 2017 Competition: Adversarial Attacks and Defenses**
Achieved 1 Gold and 2 Silver Medals
Participated in competition pitting submitted adversarial attacks against submitted defenses. Achieved 6th, 11th, and 14th in the Targeted Attack, Defense, and Non-Targeted Attack competitions, respectively.

Mar. 2017 - May. 2017 — **Learning to Play Super Smash Bros. Melee with Delayed Actions**
Used recurrent neural networks to teach a computer to play Super Smash Bros. Melee in a more humanlike way. Solution stabilized the training of competitive agents with reinforcement learning under human-level delay, as evidenced by qualitative and quantitative results against both the built-in AI and other trained agents.

Jan. 2017 - May. 2017 — **Using Macroeconomic Forecasts to Improve Mean Reverting Trading Strategies**
Improved a multiple pairs trading strategy on major currency pairs using machine learning forecasts of a series of pertinent macroeconomic variables. This addition resulted in a clear improvement in the APR over the evaluation period.

Mar. 2017 - May. 2017 — **The Game of Set**
Completed a client-server application which allows users to play the game of SET against each other over the internet. Used JavaFX8 for the UI and MySQL for the database. On the server-side, pipes were used to communicate between threads. Used the publish-subscribe messaging pattern for server-client communication.