YUN-YUN (ALICE) TSAI

(+1) 917-498-2709 • yt2781@columbia.edu • https://yunyuntsai.github.io

RESEARCH INTERESTS

My research focuses on **robust machine learning** and **self-supervised learning** toward the robustness of deep neural networks. My recent projects and papers are related to test-time domain adaptation, adversarial attack and defense. My research goal is to improve trustworthy, reliability over ML algorithms and their practical applications.

EDUCATION

Columbia University in the City of New York
Ph.D. student in Computer Science *Sept., 2021 – Present*
– Advisor: Prof. Junfeng Yang

National Tsing Hua University
Master of Science in Computer Science *Sept., 2018 – June, 2020*
– Overall GPA: 3.9/4.0, Advisor: Prof. Tsung-Yi Ho
Bachelor of Science in Computer Science *Sept., 2014 – June, 2018*
– Last 60 GPA: 3.8/4.0

PUBLICATIONS

Conference and Workshop Papers

[C1] **Yun-Yun Tsai**, Lei Hsiung, Pin-Yu Chen, and Tsung-Yi Ho, "Generalizing Adversarial Training to Composite Semantic Perturbations," in Proceeding of International Conference on Machine Learning AdvML Workshop (**ICML**), 2021.

[C2] Lei Hsiung, **Yun-Yun Tsai**, Pin-Yu Chen, and Tsung-Yi Ho, "CARBEN: Composite Adversarial Robustness Benchmark," in Proceeding of International Joint Conference on Artificial Intelligence (**IJCAI**), 2022.

[C3] Chao-Han Huck Yang, **Yun-Yun Tsai**, and Pin-Yu Chen "Voice2Series: Reprogramming Acoustic Models for Time Series Classification," in Proceeding of International Conference on Machine Learning (**ICML**), 2021.

[C4] **Yun-Yun Tsai**, Pin-Yu Chen, and Tsung-Yi Ho, "Transfer Learning without Knowing, Reprogramming black box machine learning model with scarce data and limited resources," in Proceeding of International Conference on Machine Learning (**ICML**), 2020.

[C5] **Yun-Yun Tsai**, Pin-Yu Chen, Tsung-Yi Ho, "Adversarial Machine Learning for Social Good: Reprogramming black box machine learning model with scarce data and limited resources," Advances in Neural Information Processing Systems (**NeurIPS**) NewInML Workshop, Poster, 2019.

[C6] Honggang Yu, Kaichen Yang, Teng Zhang, **Yun-Yun Tsai**, Tsung-Yi Ho, Yier Jin, "CloudLeak: Large-Scale Models Extraction Through Adversarial Examples," in Proceeding of Network and Distributed System Security Symposium (**NDSS**), 2020.

[C7] Ta-Wei Huang, **Yun-Yun Tsai**, Chung-Wei Lin, Tsung-Yi Ho, "Vehicle Sequence Reordering with Cooperative Adaptive Cruise Control," in Proceeding of Design, Automation and Test in Europe Conference and Exhibition (**DATE**), 2019.

Patents

[P1] Pin-Yu Chen, **Yun-Yun Tsai**, Sijia Liu, Chia-Yu Chen, I-Hsin Chung, Tsung-Yi Ho. "Transfer Learning With Machine Learning Systems", U.S. Patent Application No: 17/029506, Application Date: September 23, 2020.

HONORS, AWARDS, AND GRANTS

PhD Dean's Fellowship, Fu Foundation School of Engineering and Applied Sciences, Columbia University, 2021.

Best Presenter, Blackhat Award Forum in CyberSec, Taiwan, 2020.

Research and Working Experience

Graduate Research Assistant — **Columbia University**, NY, New York
Advisor: Prof. Junfeng Yang — *Sept, 2021 – present*

- **Contrastive learning toward Reversing Data for Robust Test-time Adaptation**
  - Proposed an effective input reversal method at inference time for model adaptation via optimizing auxiliary perturbations and minimizing contrastive loss.
  - Improving 15∼20% robust acc. on 15 types of ImageNet-Corruption which is better than SOTA.
- **A Robust Transformer for Defense against Unknown Adversarial with Masked Auto Encoder (MAE)**
  - Proposed a robust model pipeline for defense against unforeseen adversarial attacks, including three steps: detection, reconstruction and classification.
  - Utilizing reconstruction loss of MAE to repair the input from malicious to benign samples.
  - Up to 95% detection acc. and 10%∼25% classification acc. improvement after reconstruction.

Graduate Research Assistant — **NTHU**, Hsinchu, Taiwan
Advisor: Prof. Tsung-Yi Ho — *July, 2020 – June, 2021*
Co-Advisor: Dr. Pin-Yu Chen, IBM Research, NY, USA

- **Generalized Adversarial Training to Composite Semantic Perturbations**
  - Exploring the optimal combinations of pixel-/semantic-based perturbations and spatial transformation, including various ordering and magnitude of bias, to enhance model robustness.
- **Limited-data Transfer Learning on ML Models**
  - Utilizing adversarial reprogramming techniques and zeroth-order gradient estimation algorithms to transfer model functionality from general ImageNet classifier to specific medical imaging classifiers with limited training data, and the results outperform both traditional and SOTA transfer learning methodologies.
- **Autonomous vehicle sequence optimization for Cooperative Adaptive Cruise Control**
  - Proposed a clique-based partition and merge algorithm for vehicle reordering to find the desired platoon sequence and minimize platoon length determined by braking factors.
  - Improved 1.3x∼2.2x on total operation times comparing with baselines and reduced the platoon length by 20%.

Visiting Scholar — **University of Florida**, Gainesville, FL, USA
Advisor: Prof. Yier Jin — *Mar., 2019 – Aug., 2019*

- **Large-scale Model Extraction from MLaaS**
  - Reducing 30x queries for copying a given black-box ML model using adversarial active learning and local approximation of decision boundary with different model architectures while remaining same performance.
  - Utilized adversarial active learning to copycat targeted black-box ML services with substitute retrained model and locally approximate the decision boundary.

Research Intern — **Microsoft Cloud + AI Team**, Taipei, Taiwan
Mentor: Cheryl Hsu — *July, 2017 – June, 2018*

- Prototyping IoT solutions to 20+ business partners (e.g., NEC, Nexcom) by leveraging Azure IoT Cloud and Edge computing in a wide range of scenarios including Smart City, Retail, Home and Manufacturing.
  - Designed/Created Azure Machine Learning hands-on materials with Microsoft MVPs and delivered the workshop in coding angel event to 100+ college STEM female students.

Professional Service and Skills

Paper Review	IEEE Access, ICLR 2021, AAAI 2021, ICPAI 2020, CVPR 2022, NeurIPS 2022
Programming Languages	Python, Java, Ruby, C/C++, Verilog
Packages	Tensorflow, Pytorch, Sci-kit Learn, Keras, Caffe
Languages	English (fluent), Chinese (native)
TOEFL iBT	102/120 (Reading 28, Listening 25, Writing 26, Speaking 23)
GRE	Verbal 157/170, Quantitative 166/170, Analytical Writing 3.0/6.0
Certificate	Microsoft Data Science Certificate