Ryan Feng

U.S. Citizen | rtfeng@umich.edu | https://websites.umich.edu/~rtfeng/ | ryan-feng-b0b205133 | ryan-feng

Education

- **University of Michigan** *2019-Present (Expected: Dec. 2025)*
  Ph.D, Computer Science and Engineering, Advisors: Atul Prakash and Stella X. Yu Ann Arbor, MI
  - **Topic:** Machine Learning Robustness in the Physical World
- **University of Michigan** *2019-2021*
  M.S., Computer Science and Engineering GPA: 4.0
- **University of Washington** *2015-2019*
  B.S., Computer Engineering, Summa Cum Laude GPA: 3.95

Selected Projects

- **FoCal: Foundation Model Robustness via Test-Time Search [ICML 2025]** *University of Michigan*
  Tools: CLIP, SAM, Stable Diffusion, 3D NVS Generators, TRELLIS, Zero-1-to-3 [ ]
  - Developed FoCal, a test-time search method to make foundation models (CLIP, SAM) more robust to input variations (3D viewpoints, lighting, contrast, 2D rotations, day-night).
  - Used 3D generators (Trellis, Zero-1-to-3) and foundation model priors to canonicalize inputs at test-time.
- **GRAPHITE: Automatic, Physical, Black-Box Attack Generator [EuroS&P 2022]** *University of Michigan*
  Tools: CNNs, Zeroth-Order Optimization, Adversarial Attacks, Traffic Sign Recognizers [ ]
  - Developed GRAPHITE, the first physical, black-box hard-label attack on deep learning computer vision models such as traffic sign recognizers. Includes a stop sign to speed limit sticker attack with 95.7% ASR.
  - Contributed it to the Adversarial Robustness Toolbox, version 1.12.0.
- **OARS: Adaptive Black-Box Attack [CCS 2023]** *University of Michigan*
  Tools: MLaaS, Stateful Defenses, Zeroth-Order Optimization, Adversarial Attacks [ ]
  - Developed OARS, an adaptive black-box attack that breaks MLaaS stateful defenses (ASR 0% to 99%).
- **SPANet: Generalizing Assistive Feeding Robots to Unseen Food [ISRR 2019]** *University of Washington*
  Tools: CNNs, ROS
  - Led data collection for ML-based skewering generalization to unseen food.
  - Developed manipulation strategies for robotic, computer vision assistive feeding system.
- **(Ongoing): Jailbreak Defense Against Embodied LLMs / VLA Agents** *University of Michigan*
  Tools: VLMs, LLMs, VLA agents, Jailbreaking
  - Developing a defense for jailbreaks against embodied VLA agents.

Industry Experience

- **KLA** *Summer 2023*
  Algorithm Engineering Intern, AI / Algorithms Group Ann Arbor, MI
  - Improved OOD robustness in ML metrology applications with adversarial examples and cGANs.
- **Xevo, Inc.** *2017-2019*
  Software Development Engineer, AI Group Seattle, WA
  - Built deep learning computer vision-based eye-tracking system for driver attentiveness and incorporated it into our team's immersive driving demo.
  - Trained lightweight CNNs for traffic sign recognition for autonomous vehicle technology.

Honors and Awards

- **J. Robert Beyster Computational Innovation Graduate Fellowship** *2023*
  Awarded to 4 students in the College of Engineering (University of Michigan)
- **Outstanding Computer Engineering Senior Award** *2019*
  Awarded to top male and female student in Computer Engineering (University of Washington)
- **Microsoft Endowed Scholarship** *2018*
  University of Washington

Software Releases

- **GRAPHITE:** https://github.com/ryan-feng/GRAPHITE and https://github.com/Trusted-AI/adversarial-robustness-toolbox/tree/1.12.0
- **OARS:** https://github.com/nmangaokar/ccs_23_oars_stateful_attacks
- **FoCal:** https://github.com/sutkarsh/focal

Publications

C=Conference, W=Workshop

[C.9] *Test-Time Canonicalization by Foundation Models for Robust Perception*
Utkarsh Singhal*, **Ryan Feng***, Stella X. Yu, Atul Prakash
International Conference on Machine Learning (ICML), 2025.

[C.8] *D4: Detection of Adversarial Diffusion Deepfakes Using Disjoint Ensembles*
Ashish Hooda*, Neal Mangaokar*, **Ryan Feng**, Kassem Fawaz, Somesh Jha, Atul Prakash
IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), 2024.

[C.7] *Stateful Defenses for Machine Learning Models Are Not Yet Secure Against Black-box Attacks*
Ryan Feng*, Ashish Hooda*, Neal Mangaokar*, Kassem Fawaz, Somesh Jha, Atul Prakash
ACM Conference on Computer and Communications Security (CCS), 2023.

[C.6] *Concept-based Explanations for Out-Of-Distribution Detectors*
Jihye Choi, Jayaram Raghuram, **Ryan Feng**, Jiefeng Chen, Somesh Jha, Atul Prakash
International Conference on Machine Learning (ICML), 2023.

[W.2] *Theoretically Principled Trade-off for Stateful Defenses against Query-Based Black-Box Attacks*
Ashish Hooda*, Neal Mangaokar*, **Ryan Feng**, Kassem Fawaz, Somesh Jha, Atul Prakash
ICML 2023 Workshop on New Frontiers in Adversarial Machine Learning (AdvML), 2023.

[C.5] *GRAPHITE: Generating Automatic Physical Examples for Machine-Learning Attacks on Computer Vision Systems*
Ryan Feng, Neal Mangaokar, Jiefeng Chen, Earlence Fernandes, Somesh Jha, Atul Prakash
IEEE European Symposium on Security and Privacy (EuroS&P), 2022.

[W.1] *Using Anomaly Feature Vectors for Detecting, Classifying and Warning of Outlier Adversarial Examples*
Nelson Manohar-Alers, **Ryan Feng**, Sahib Singh, Jiguo Song, Atul Prakash
ICML 2021 Workshop on Adversarial Machine Learning, 2021.

[C.4] *Leveraging Image Processing Techniques to Thwart Adversarial Attacks in Image Classification*
Yeganeh Jalalpour, Li-Yun Wang, **Ryan Feng**, Wu-chi Feng
IEEE International Symposium on Multimedia (ISM), 2019.

[C.3] *Robot-Assisted Feeding: Generalizing Skewering Strategies across Food Items on a Realistic Plate*
Ryan Feng*, Youngsun Kim*, Gilwoo Lee*, Ethan K. Gordon, Matt Schmittle, Shivaum Kumar, Tapomayukh Bhattacharjee, Siddhartha S. Srinivasa
International Symposium on Robotics Research (ISRR), 2019.

[C.2] *ISIFT: Extracting Incremental Results from SIFT*
Ben Hamlin, Wu-chi Feng, **Ryan Feng**
ACM Multimedia Systems Conference (MMSys), 2018.

[C.1] *Understanding the Impact of Compression on Feature Detection and Matching in Computer Vision*
Wu-chi Feng, **Ryan Feng**, Paul Wyatt, Feng Liu
IEEE International Symposium on Multimedia (ISM), 2016.

*denotes equal contribution.

Teaching, Service and Activities

- **Co-Instructor:** Co-led EECS 598-012: Special Topics in Secure and Trustworthy Machine Learning (Winter 2023)
- **GSI:** EECS 442: Computer Vision (Winter 2025), EECS 542: Advanced Topics in Computer Vision (Fall 2024)
- **TA:** CSE 142: Comp. Programming I (Fall 2016), CSE 143: Comp. Programming II (Winter 2017, Spring 2017)
- **Reviewer:** ICML (2022, 2023, 2024, 2025), NeurIPS (2022, 2023, 2024, 2025), CVPR (2024, 2025), ECCV (2024), ICCV (2025), AAAI (2024, 2025), TIP (2021), EuroS&P (2021 - external), IJCV (2022), IEEE S&P Posters (2024, 2025)
- **Area Chair:** R2HCAI (2023 - selected as top area chair)
- **Grad Student Mentor:** College of Engineering (CoE) Lunch and Lab, University of Michigan, Fall 2021, Winter 2022, Computer Science and Engineering Graduate Student Organization (CSEG) Buddy Program, 2022-2024
- **Vice-President and Treasurer:** Computer Science and Engineering Graduate Student Organization (CSEG), University of Michigan, Winter-Spring 2020
- **Lab Tour Guide:** Discover Engineering, University of Michigan, Summer 2022

Skills

- Python, C, C++, Java, Shell Scripting, PyTorch, TensorFlow, OpenCV, Numpy, Android, Git, Slurm, Unity

Research Experience

- **University of Michigan** *2019 - Present*
  Graduate Student Research Assistant, Prakash and Yu Groups Ann Arbor, MI
  - Machine learning robustness in the physical world. Includes foundation model robustness via test-time search (FoCal), physical black-box attacks (GRAPHITE), adaptive black-box attacks (OARS).

- **Portland State University** *Summer 2019*
  Research Intern, Intel Systems and Networking Lab Portland, OR
  - Designed adversarial machine learning image cleaning algorithms for CNNs.

- **University of Washington** *2018-2019*
  Undergraduate Research Assistant, Personal Robotics Lab Seattle, WA
  - Developed manipulation strategies for robotic, computer vision assistive feeding system.
  - Led data collection for ML-based skewering generalization to unseen food.

- **Portland State University** *Summer 2016, Summer 2015*
  Research Intern, Intel Systems and Networking Lab Portland, OR
  - Evaluated efficient image keypoint detection (ISIFT) and matching for vision and multimedia.

- **Portland State University** *Summer 2014*
  Research Intern, Graphics and Computer Vision Lab Portland, OR
  - Developed Android apps for computational photography applications in changing color distributions.