Yaoteng TAN

Tel: (951)-367-7913 Email: yaoteng.tan@email.ucr.edu Homepage: https://ytengtan.github.io/

EDUCATION

University of California, Riverside (On going) Sep. 2022 - Present
Ph.D. student in Electrical Computer Engineering (ECE) Riverside, CA
Advised by Dr. M. Salman Asif

Huazhong University of Science and Technology Sep. 2018 - Jun. 2022
B.S. in Electrical Engineering Wuhan, China
Graduated with the honors.

EXPERIENCE

ECE Department, University of California, Riverside Oct. 2022 - Present
Graduate research assistant Riverside, CA

- Research topics involve trustworthy machine learning and inverse problems.
- Teaching assistant for graduate and undergraduate courses.

CloudMinds Technology Inc. Jul. 2021 - Aug. 2021
Research intern Beijing, China

- Project summary: Service Robot vision based on YOLOv5.
- Implement object detection for the service robot and deploy it on the UE4-based simulation platform.

RESEARCH INTEREST

My research focuses primarily on trustworthy AI, with a emphasis on adversarial ML and machine unlearning, contributing to the ethical deployment of AI systems. I am also interested in computational imagining, exploring efficient optimization techniques, and better priors for general image reconstruction from limited measurements. While these are my core interests, I remain open to various research questions across computer vision and machine learning fields.

PUBLICATION

Targeted Unlearning with Single Layer Unlearning Gradient, *ICML 2025*.
*Zikui Cai, ***Yaoteng Tan**, M. Salman Asif. (*Equal contribution)
Large foundation models, such as Stable Diffusion and Vision-Language Models, are widely deployed but pose risks of generating harmful content. Existing unlearning techniques attempt to remove sensitive or harmful information, yet they often remain inefficient. In this work, we proposed an efficient unlearning technique for foundation models that requires only one-time gradient computation with one-step partial model modification (one layer).

Ensemble-based Blackbox Attacks on Dense Prediction, *CVPR 2023*.
*Zikui Cai, ***Yaoteng Tan**, M. Salman Asif. (*Equal contribution)
We propose an ensemble method for generating effective black-box adversarial attacks against models for dense prediction vision tasks (e.g., object detection and semantic segmentation). Our method reduces the black-box query space from the pixel space to the ensemble model space, thereby increasing query efficiency significantly. Moreover, our method can generate a single perturbation that effectively fools multiple black-box detection and segmentation models simultaneously in a target-consistent manner, which is demonstrated for the first time.

Transform-dependent adversarial attacks, *Under submission*.
Yaoteng Tan, Zikui Cai, M. Salman Asif. (Preprint: https://arxiv.org/abs/2406.08443)
Many properties of adversarial attacks are well-studied today. In this work, we explore an under-researched property of

adversarial attacks — transform-dependent, which the optimization process of additive adversarial perturbations can be combined with various image transformations to produce versatile, transform-dependent attack effects. We demonstrate that the transform-dependent property of attacks holds for many differentiable image transformations. Additionally, we show that this property can be leveraged to attack object detection systems, with implications for real-world applications.

PROJECT

Discovery of Sparsity-Constrained Optimization Algorithms, $Sep.2024 - Present.$
Traditionally, optimization algorithms, such as momentum acceleration, have been meticulously developed by mathematicians using well-established mathematical principles and rigorous theoretical foundations. In this project, we leveraging the power of data to uncover innovative solvers. With gradient descent as the sole domain knowledge, we aim to discover algorithms that can solve optimization problems efficiently and adaptively, potentially surpassing traditional methods in performance and scalability. By doing so, we seek to bridge the gap between classical mathematical approaches and modern data-driven methodologies, pushing the boundaries of what can be achieved in optimization.

Adversarial robustness study of data-driven computational imaging systems, $Feb. - Jun.2022.$
Data-driven methods such as deep learning are widely used in computational imaging pipelines and outperform traditional model-based methods on many benchmarks. However, as deep models are well-known to be vulnerable to adversarial attacks, we hypothesize that deep imaging pipelines are more vulnerable to adversarials than those imaging processes are explicitly modeled. In this project, we discovered a deep image super-resolution model recovers more artifacts than interpolation algorithms (e.g., bilinear, bicubic) when adversarial perturbation is injected.

SERVICE

Conference reviewer

- WACV 24′, ICIP 25′, ICCV 25′

Teaching assistant

- EE240 Pattern Recognition, UC Riverside, Spring 2023, Spring 2024
  Duties: holding office hours, grading students' homework and final projects.
- CS171/EE142 Introduction to Machine Learning and Data Mining, UC Riverside, Fall 2023
  Duties: short teaching in discussion sections, holding office hours, designing exam questions, gradings.

SKILLS

Technical

- Adversarial machine learning, machine unlearning, convex optimization, inverse problem, image (signal) processing.
- Programming language: Python, MatLab, Java, familiar with C-like, HTML.
- Familiar with basic source control tools (e.g., Git), writing tools (e.g., LaTeX)

Language

- English, Mandarin

MISC

Activities and societies

- Student leader of UCR student organization *Happy English Corner*.