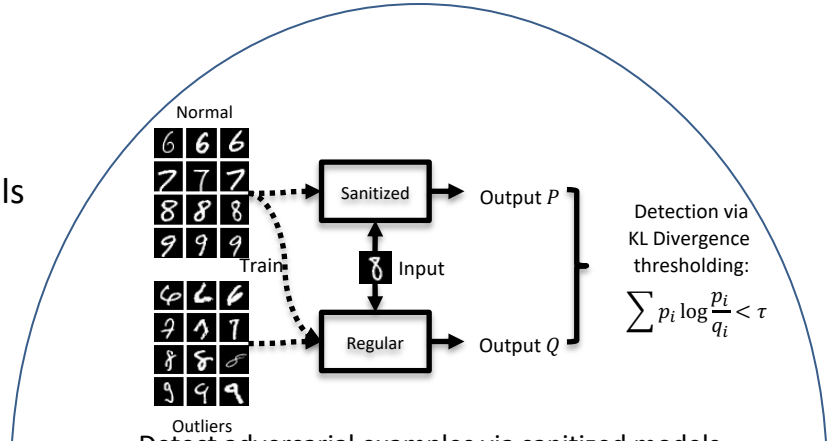
SaTC: CORE: Medium: Collaborative: Towards Robust Machine Learning Systems

Challenges:

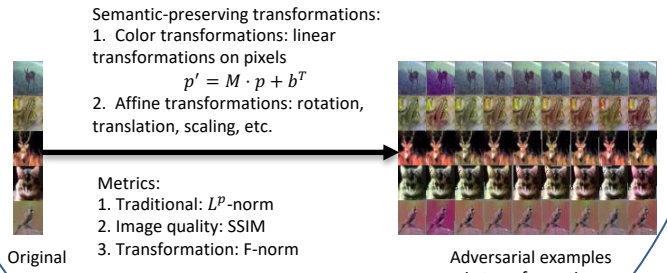
- Lack of fundamental understanding of machine learning models
- Hard to measure the boundary of machine perception



Detect adversarial examples via sanitized models

Solutions:

- Study the relationship between dataset quality and model robustness
- Explore larger attack spaces in real-world scenarios



Adversarial attack in broader attack spaces

Scientific Impact:

- Help real-world assessment of adversarial threats and build systems with better performance-robustness trade-offs
- Facilitate related research topics such as interpretable and privacy-preserving ML

Broader Impact:

- Make sources public for courses and research, with particular efforts to include students from underrepresented groups
- Support high school outreach programs and summer camps

NSF Award ID: 1801751

PIs: Hao Chen (University of California, Davis)
Neil Zhenqiang Gong (Duke University)