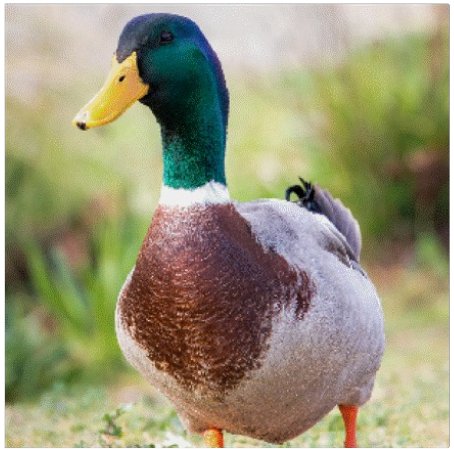
Adversarial

Eric Wong

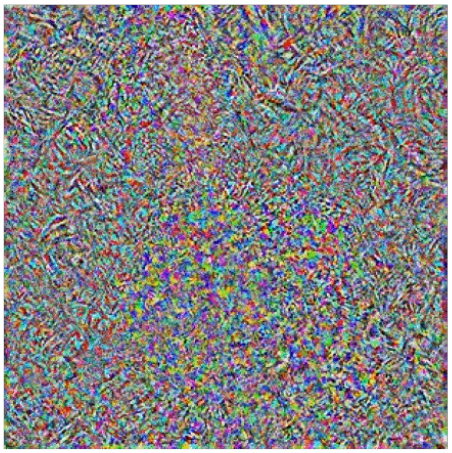
9/8/2022

Noise attack

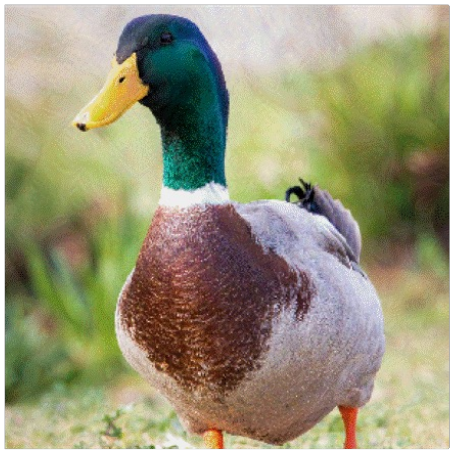


“Duck”

+

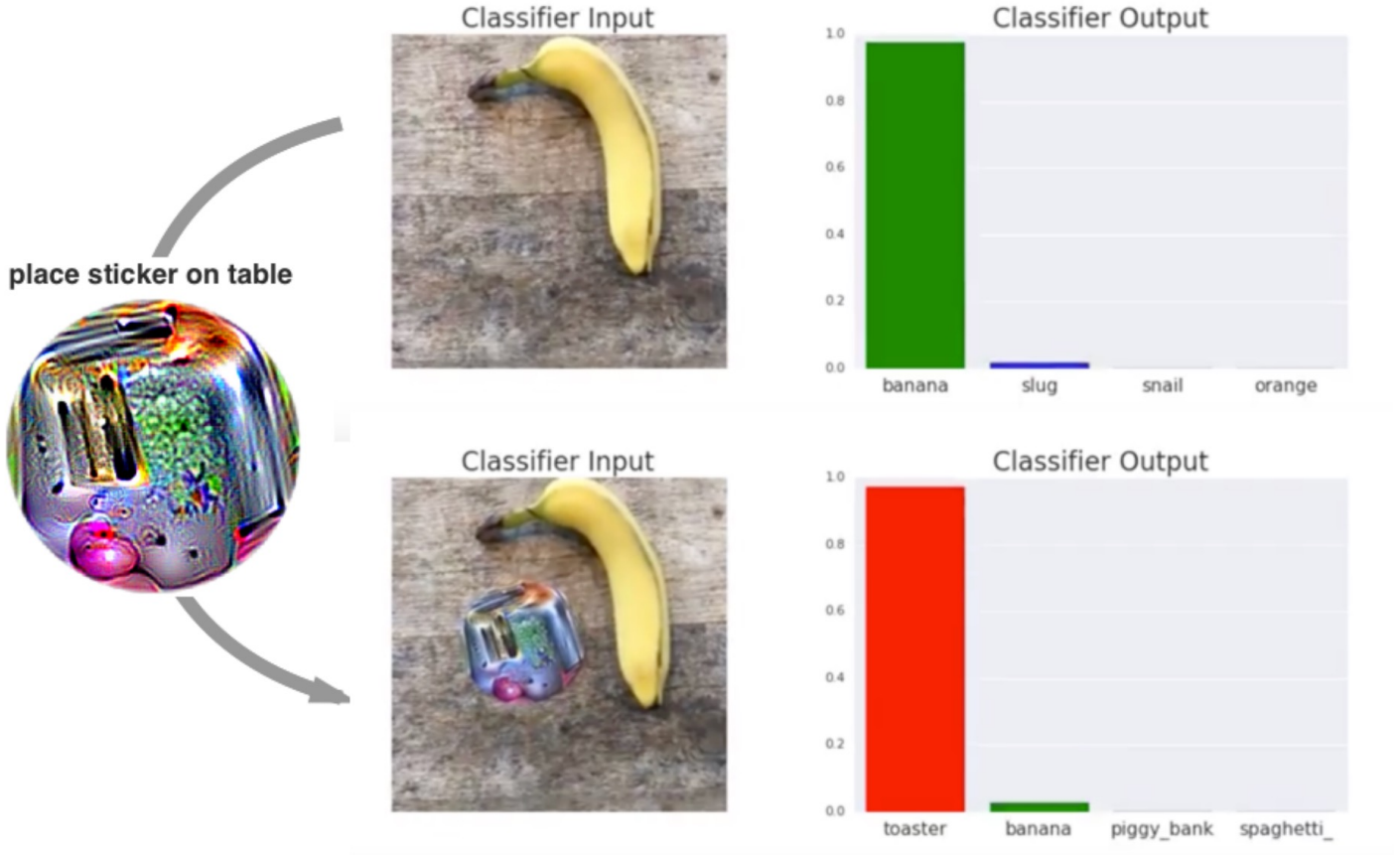


=



“Hermit crab”

Patch attack



Brown et al. 2017 “Adversarial Patch”

3D printed textures

classified as turtle

classified as rifle

classified as other

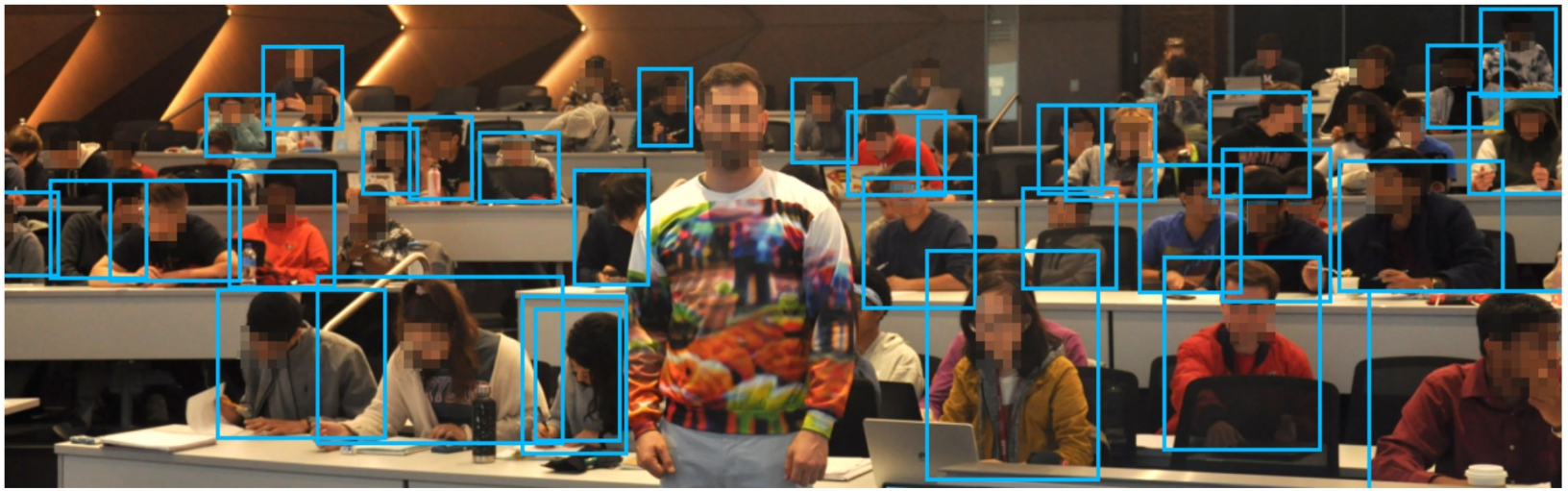
Athalye et al. 2017 “Synthesizing Robust Adversarial Examples”

Glasses

(a) (b) (c)

Sharif et al. 2018 “A General Framework for Adversarial Examples with Objectives”

Clothing

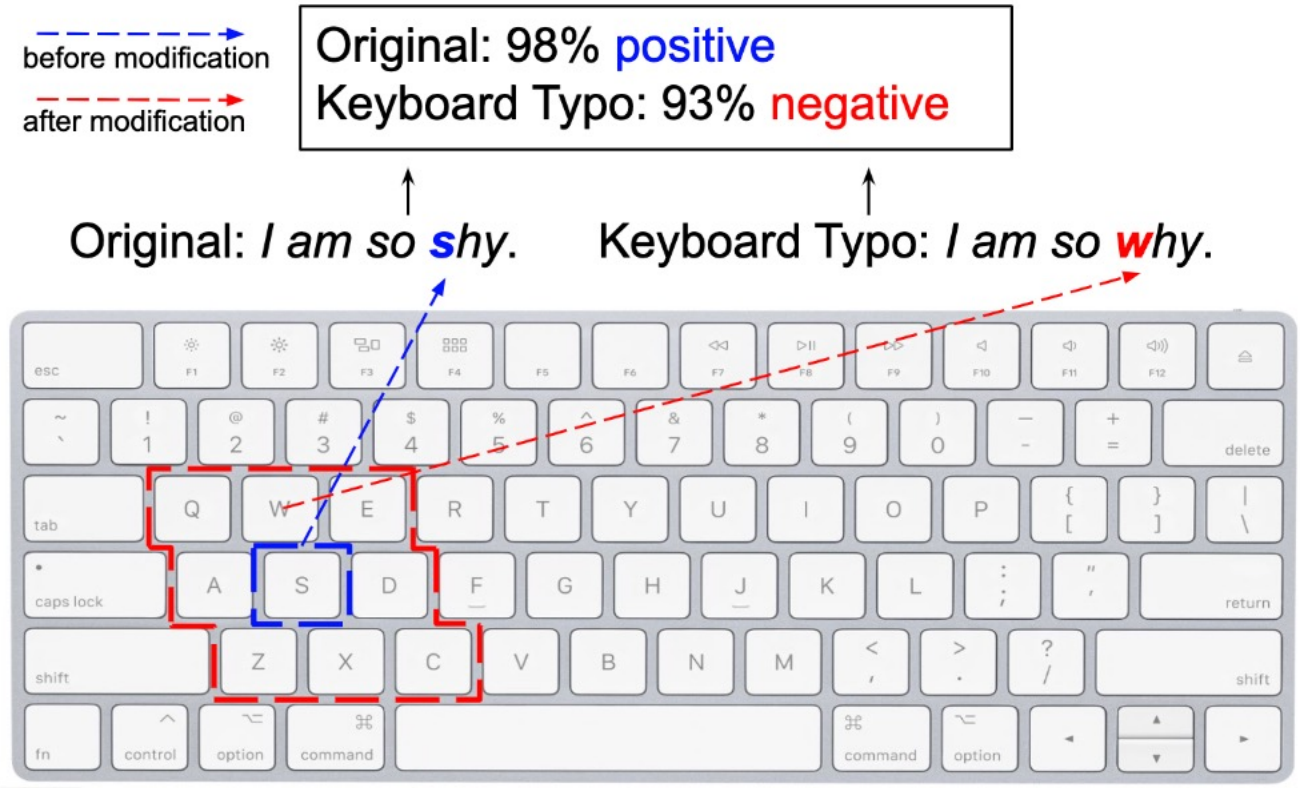


Wu et al. 2019 “Making an Invisibility Cloak: Real World Adversarial Attacks on Object Detectors”

Camera stickers

Li et al. 2019 “Adversarial camera stickers: A physical camera based attack on deep learning sytems”

Typos



Sun et al. 2020 “Adv-BERT: BERT is not robust on misspellings! Generating nature adversarial samples on BERT”

Sentences

Label	Sentence
P	I am currently trying to give this company another chance. I have had the same scheduling experience as others have written about. Wrote to them today
N	I am currently trying to give this company another review. I have had the same dental experience about others or written with a name. Thanks to them today

Hsieh et al. 2019 “Natural Adversarial Sentence Generation with Graident based Perturbation”

Speech recognition

Abdullah et al. 2020