NANYANG TECHNOLOGICAL UNIVERSITY
SINGAPORE

School of Computer Science and Engineering
College of Engineering

Developing AI Attacks/Defenses

Artificial Intelligence (AI) systems are very commonly used in our everyday lives. These systems help simplify tasks that us humans require to do. Dependence on such systems will be greatly increased in the future. However these systems are not fully secured to malicious acts of attackers. By the addition of perturbation to the input, the system may not be able to work as it is supposed to. Even if the perturbation to the source input is not distinguishable to humans, it is still able to fool the AI.

Project Objective

To develop an adversarial attack on the SinGAN model which learns from a single input image, which is able to fool the classifier, assigning a wrong label to the image.

Student: Terence Chan Chin Leng
Supervisor: Asst Professor Zhao Jun