Reading list for attacking

[1] Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey. 英文综述 link

[2] 知乎专栏，前几章有介绍样本对抗攻防的基础的 link

[3] pytorch官方文档----60分钟入门 link

[4] 样本对抗的来龙去脉和本质 link

[5] 样本对抗的一个相关比赛 link

[6] Awesome ML Attack link

[7] 简单易懂的人脸识别过程和原理介绍 link

[8] 一种鲁棒的神经网络架构(防御) link

[9] 对抗训练论文一(防御) link

[10] Ian GoodFellow机器学习的博客 link

Open Source about ADVERSARIAL EXAMPLE GENERATION

[1] PyTorch FGSM Tutorial link

[2] PyTorch C&W Attack link

[3] PyTorch DDN Attack(CVPR2019) link

Face Recognition

[1] Loss Function for training Face Recognition Model link

[2] Face Recognition Model: ZhaoJ9014/face.evoLVe.PyTorch （默认白盒模型） link

[3] Face Recognition Model: ageitgey/face_recognition （第一次老师给的白盒模型） link

Neural network backdoor

[0] 浙大的一篇调研 link

[1] Neural Cleanse: Identifying and Mitigating Backdoor Attacks in Neural Networks. link 翻译 link

[2] Targeted Backdoor Attacks on Deep Learning Systems Using Data Poisoning.(key pattern) link

[3] A General Framework for Adversarial Examples with Objectives.(AGN方法) link 机器之心的解读link

[4] Accessorize to a Crime: Real and Stealthy Attacks on State-of-the-Art Face Recognition. link 源码link

[5] Robust Physical-World Attacks on Deep Learning Visual Classification.(对路牌攻击) link