R10943181 電子所碩一 黃詩瑜

1. 根據你最好的實驗結果，簡述你是如何產生transferable noises, Judge

Boi上Accuracy降到多少？

Ans:

我選擇使用MI-FGSM attack method與Ensemble attack，其中用到的model為：

model\_names = ['resnet56\_cifar10','nin\_cifar10','densenet100\_k12\_cifar10','pyramidnet110\_a48\_cifar10','ror3\_110\_cifar10','wrn16\_10\_cifar10','wrn28\_10\_cifar10','seresnet56\_cifar10','rir\_cifar10','preresnet110\_cifar10','preresnet164bn\_cifar10','diapreresnet56\_cifar10','resnet1202\_cifar10','resnet1001\_cifar10','resnet110\_cifar10','diapreresnet110\_cifar10']

另外也加入Diverse input (DIM)。

Judge Boi上的Accuracy達到0.14。

1. 當source model為resnet110\_cifar10(from Pytorchcv), 使用最原始的fgsm

攻擊在dog2.png的圖片。

1. 請問被攻擊後的預測的class是錯誤的嗎？

錯誤，變成cat

2. 實作jpeg compression (compression rate=70%) 前處理圖片, 請問

prediction class是錯誤的嗎？同第一題作答

正確

1. Jpeg compression為什麼可以抵擋adversarial attack, 讓模型維持高正確率？

a. 圖片壓縮時讓色彩更鮮豔 b. 圖片壓縮時把雜訊減少 c. 圖片壓縮讓圖片品質下降 d. 圖片壓縮時雜訊反而變大

b. 圖片壓縮時把雜訊減少