Q1 Attack
1 分數

Depending on your best experimental results, briefly explain how you generate the transferable noises and the resulting accuracy on Judge Boi. (Only report accuracy without explanation can't earn credit)

我使用 MI-FGSM 在 JudgeBoi 獲得最佳結果，使用的 model 是 Query-Free Adversarial Transfer via Undertrained Surrogates，這篇 paper 有提到的 ResNet, SENet, DenseNet，我把 pytorchcv.model_provider.get_model 中有提供 ResNet, SENet, DenseNet CIFAR-10 版本的對應 models 拿來作 ensemble，其中有 resnet20_cifar10, resnet56_cifar10, resnet110_cifar10, resnet164bn_cifar10, resnet272bn_cifar10, resnet542bn_cifar10, resnet1001_cifar10, resnet1202_cifar10, resnext272_1x64d_cifar10, resnext272_2x32d_cifar10, seresnet20_cifar10, seresnet56_cifar10, seresnet110_cifar10, seresnet164bn_cifar10, seresnet272bn_cifar10, seresnet542bn_cifar10, densenet40_k12_cifar10, densenet40_k12_bc_cifar10, densenet40_k24_bc_cifar10, densenet40_k36_bc_cifar10, densenet100_k12_cifar10, densenet100_k24_cifar10, densenet100_k12_bc_cifar10, densenet190_k40_bc_cifar10, densenet250_k24_bc_cifar10。

ACC = 0.11

Q2
3 分數

When the source model is resnet110_cifar10 (from Pytorchcv), adopt the vanilla fgsm attack on image “dog/dog2.png” in data.zip.

Q2.1 Is the predicted class wrong after fgsm attack?
1 分數

No

✔ Yes

If Yes:
Change to class

cat

Q2.2 Implement the pre-processing method jpeg compression (compression rate=70%). Is the predicted class wrong after defense?
1 分數

- Yes
- ✔ No

If Yes:

Class after jpeg compression is:

Q2.3 Why jpeg compression method can defend the adversarial attack, improving the model accuracy?
1 分數

- jpeg compression enlarges the noise level
- jpeg compression makes images more colorful
- jpeg compression degrades the image qualities
- ✔ jpeg compression reduces the noise level

HW10

● 已批改

總分
4 / 4 pts

問題 1
Attack 1 / 1 pt

問題 2
(no title) 3 / 3 pts

2.1	Is the predicted class wrong after fgsm attack?	1 / 1 pt
2.2	Implement the pre-processing method jpeg compression (compression rate=70%). Is the predicted class wrong after defense?	1 / 1 pt
2.3	Why jpeg compression method can defend the adversarial attack, improving the model accuracy?	1 / 1 pt