Q1 Attack

1 Point

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)

```
model_names = [
    'nin_cifar10',
    'resnet20_cifar10',
    'preresnet20_cifar10',
    'seresnet20_cifar10',
    'sepreresnet20_cifar10',
    'wrn16_10_cifar10',
    'wrn20_10_1bit_cifar10',
    'rir_cifar10',
    'diaresnet20_cifar10',
    'diapreresnet20_cifar10',
]
```

use mifgsm to generate noise.

Because my GPU memory is not big, I choose the model of simple architecture and less number of models.

0.13

Q2

3 Points

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 Point

● Yes

○ No

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 Point

● No

○ Yes

If Yes:

Class after jpeg compression is:

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

1 Point

○ jpeg compression enlarges the noise level

○ jpeg compression degrades the image qualities

○ jpeg compression makes images more colorful

● jpeg compression reduces the noise level

STUDENT

方郁婷

TOTAL POINTS

- / 4 pts

QUESTION 1

Attack 1 pt

QUESTION 2

(no title) 3 pts

2.1 Is the predicted class wrong after fgsm attack? 1 pt

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

2.3 Why jpeg compression method can defend the adversarial attack, improving the model accuracy? 1 pt