## **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)

Attack method:
Ensemble Attack(13 models)+MIFGSM, as hints (2).

Proxy model:
ensemble model('nin_cifar10', 'resnet20_cifar10', 'preresnet20_cifar10', 'resnet1001_cifar10', 'seresnet20_cifar10', 'sepreresnet20_cifar10', 'pyramidnet110_a48_cifar10', 'densenet40_k12_cifar10', 'xdensenet40_2_k24_bc_cifar10', 'wrn16_10_cifar10', 'ror3_56_cifar10', 'rir_cifar10', 'shakeshakeresnet20_2x16d_cifar10').
[ Ensemble way: the average of each logit from the model(x). ]

Finally, MIFGSM is the normalized version of IFGSM. For the 4-D gradient array, I normalize it under each (x,y) tuple combination. The only adjustment to the default parameters is the number of iterations from 20 to 40.

0.120

## **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

- [x] 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 degrades the image qualities
- [x] jpeg compression reduces the noise level
- [ ] jpeg compression makes images more colorful
- [ ] jpeg compression enlarges the noise level

## HW10

**GRADED**

**STUDENT**
梁峻瑋

**TOTAL POINTS**
**4 / 4 pts**

**QUESTION 1**
Attack — **1** / 1 pt

**QUESTION 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 |