Assignment 2: Adversarial Attack on NN (Due: 2023/10/01 Sunday 23:59)

---

q1. (30%) Please read the **fast gradient sign method** (FGSM) paper and write pseudocode similar to the following format

Pseudocode Format Example :

Algorithm 1 Backpropagation Algorithm

```
procedure TRAIN
    X ← Training Data Set of size mxn
    y ← Labels for records in X
    w ← The weights for respective layers
    l ← The number of layers in the neural network, 1...L
    D_ij^(l) ← The error for all l,i,j
    t_ij^(l) ← 0. For all l,i,j
    For  i = 1 to m
        a^l ← feedforward(x^(i), w)
        d^l ← a(L) − y(i)
        t_ij^(l) ← t_ij^(l) + a_j^(l) · t_i^(l+1)
    if j ≠ 0 then
        D_ij^(l) ← (1/m) t_ij^(l) + λ w_ij^(l)
    else
        D_ij^(l) ← (1/m) t_ij^(l)
        where ∂/∂w_ij^(l) J(w) = D_ij^(l)
```

q2. (40%) FGSM Attack

1. Download the model weights and model reading program form here
2. Attack this model by FGSM
3. Download your testing data from here
4. To convert testing data to images and save them as JPG files (.jpg).
5. Experiment with FGSM Attack using testing data, and record the results along with the parameters used.

The **noise magnitude** of q2 will **not affect** the score, please provide a detailed description of parameter settings and implementation in **q3**.

q3. (15%) Write down your experiment setting in English. The setting should include but not be limited to (1) hardware specification, (2) package version and (3) all the experiment parameters and details in **q2**.

The font size is 12, and the page limit is 2 pages.

Submission Guideline

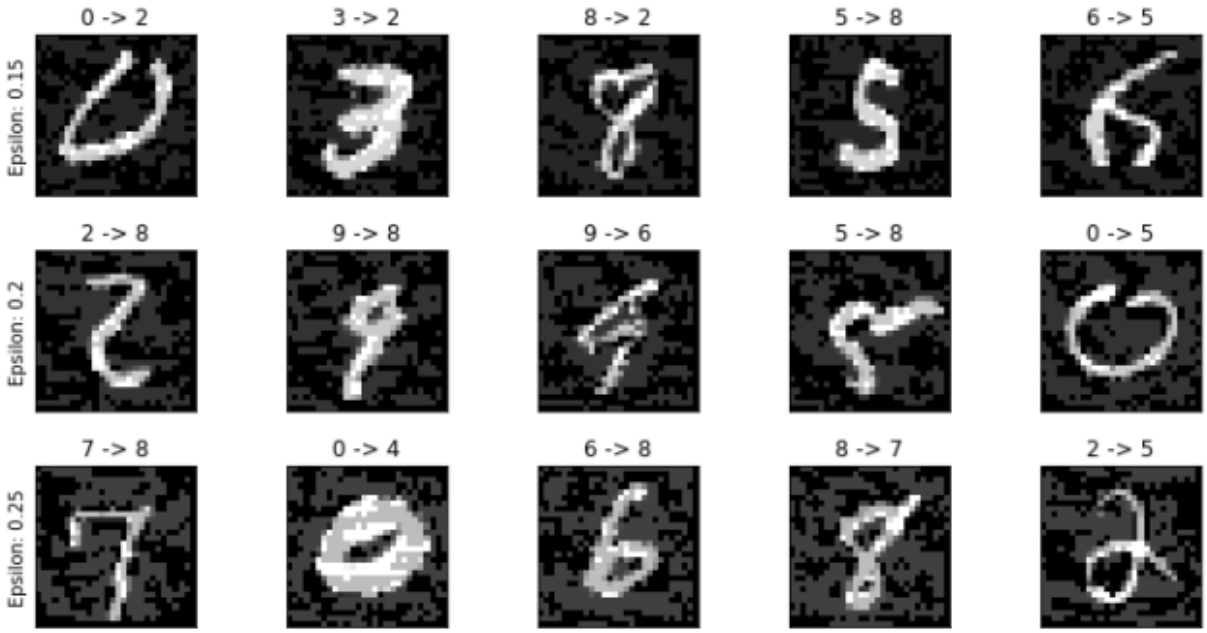
Please compress your files named {SID}_a2.zip (SID in upper case) to the COOL System, such as D111111_a2.zip, with two required files

file 1. {SID}_a2.py

Please submit your source code to the COOL system. The following command can exclude the code

The execution results of {SID}_a2.py
The output of your code is the generated attacked images (testing data original image + the noise), ground truth label and predict of FGSM attack.

Output Format Example



files 2. {SID}_a2_report.pdf

Supplementary Materials

PyTorch installation: https://pytorch.org/
FGSM paper 1412.6572.pdf (arxiv.org)