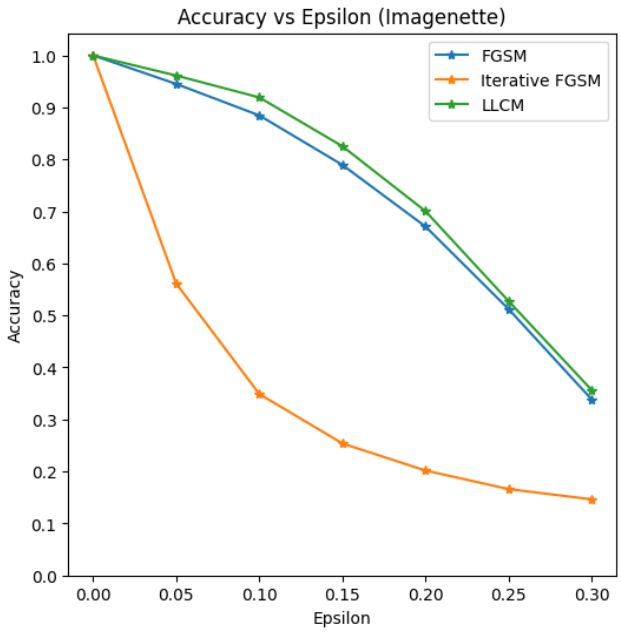
I implemented some standard Adversarial Example Generations algorithms, including Fast Gradient Sign, Iterative - Fast Gradient Sign, and Least Likely Class methods.

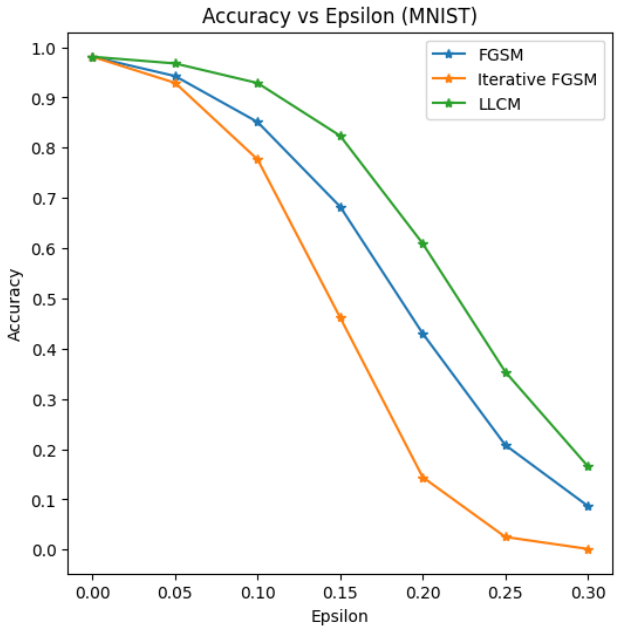
I utilized the Pytorch tutorial and this paper. The Pytorch tutorial provided the framework for the methods and provided inspiration for the visualizations. I chose to attack the ResNet50 model (with Pytorch's default training weights), due to its simplicity, which allowed for relatively fast training times.

I used most of the standard transformations as provided by the Pytorch weight specification, which includes a resize of [232,232] followed by a random crop of [224,224]. The images are then normalized. The weights originally also included interpolation, but I did not use it. I also filtered the final output of ResNet to select the max probability between the 10 classes that are included in the ImageNette dataset.

There were some differences between the paper and my implementation. The MNIST and Imagenette data are normalized between [0,1], while the paper has pixel values ranging from [0,255]. To account for this, I used Pytorch's suggested epsilon values of [0, .05, .1, .15, .2, .25, .3]. For the Iterative Gradient Sign method, I unnormalized some of the inputs, including epsilon, which I multiplied by 255 as the original paper had pixels values in the range [0,255]. Further, I also used the α value of 1/255 for the same reason.

Plots

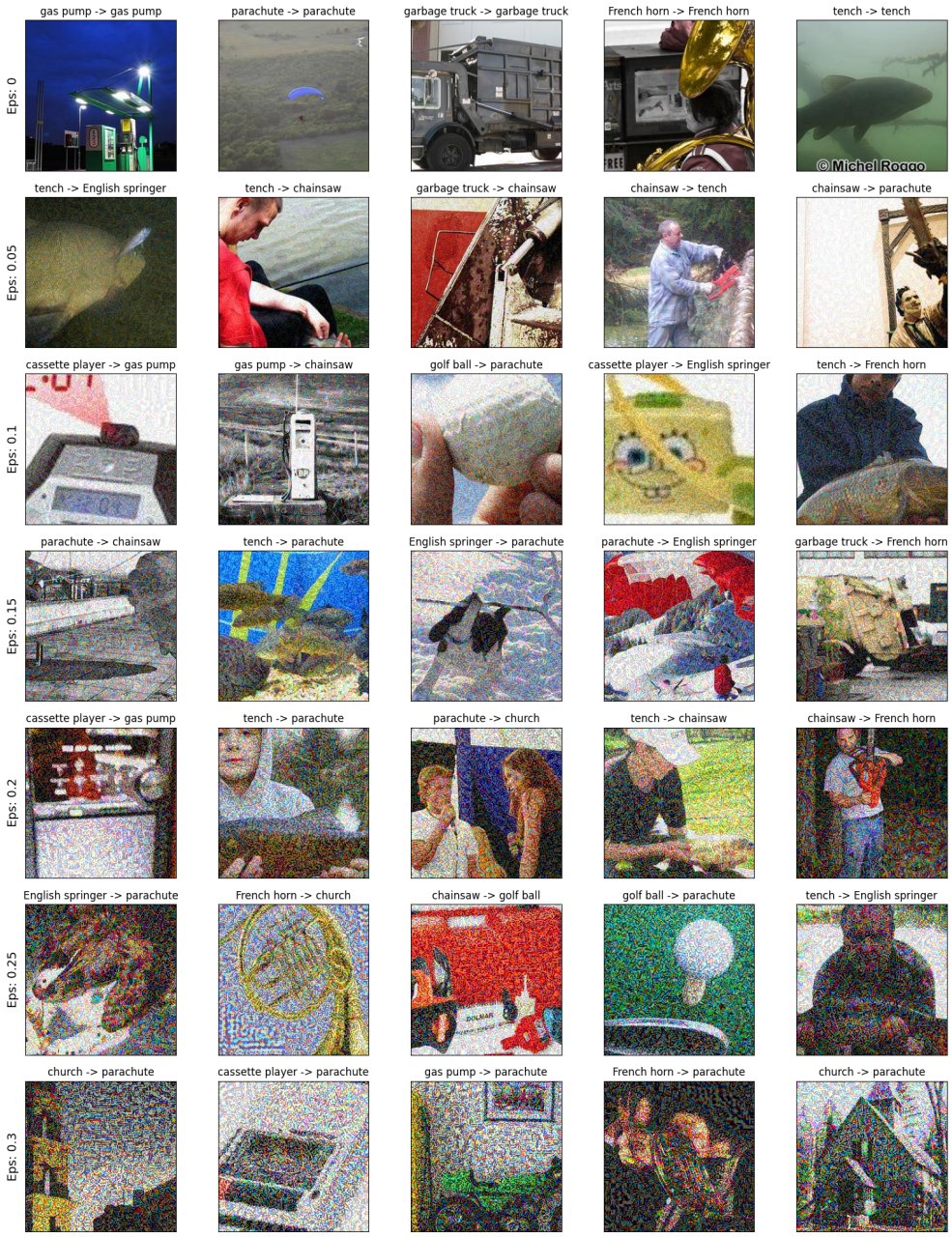


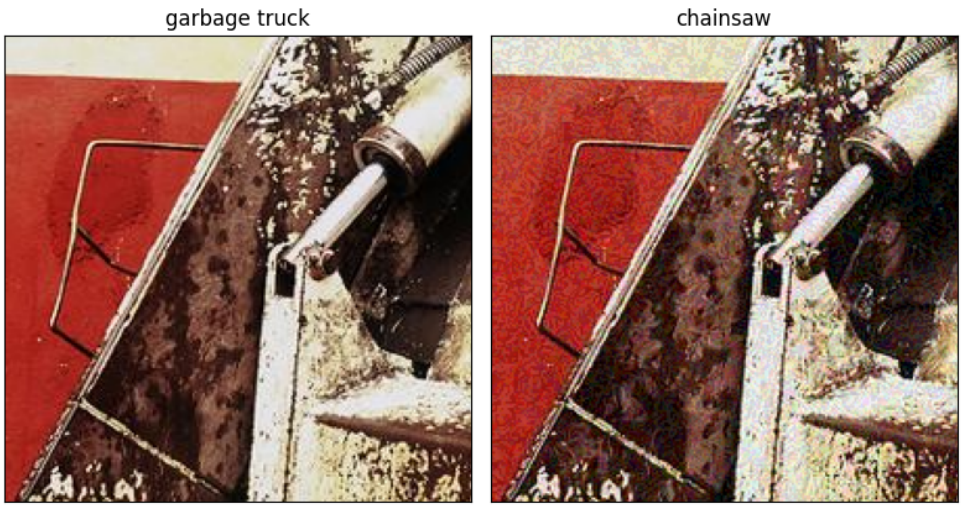


(iii) Visualize examples of attacks, where the original image and original (correctly predicted) label and the attacked image and falsely predicted label are shown for both MNIST and ImageNette

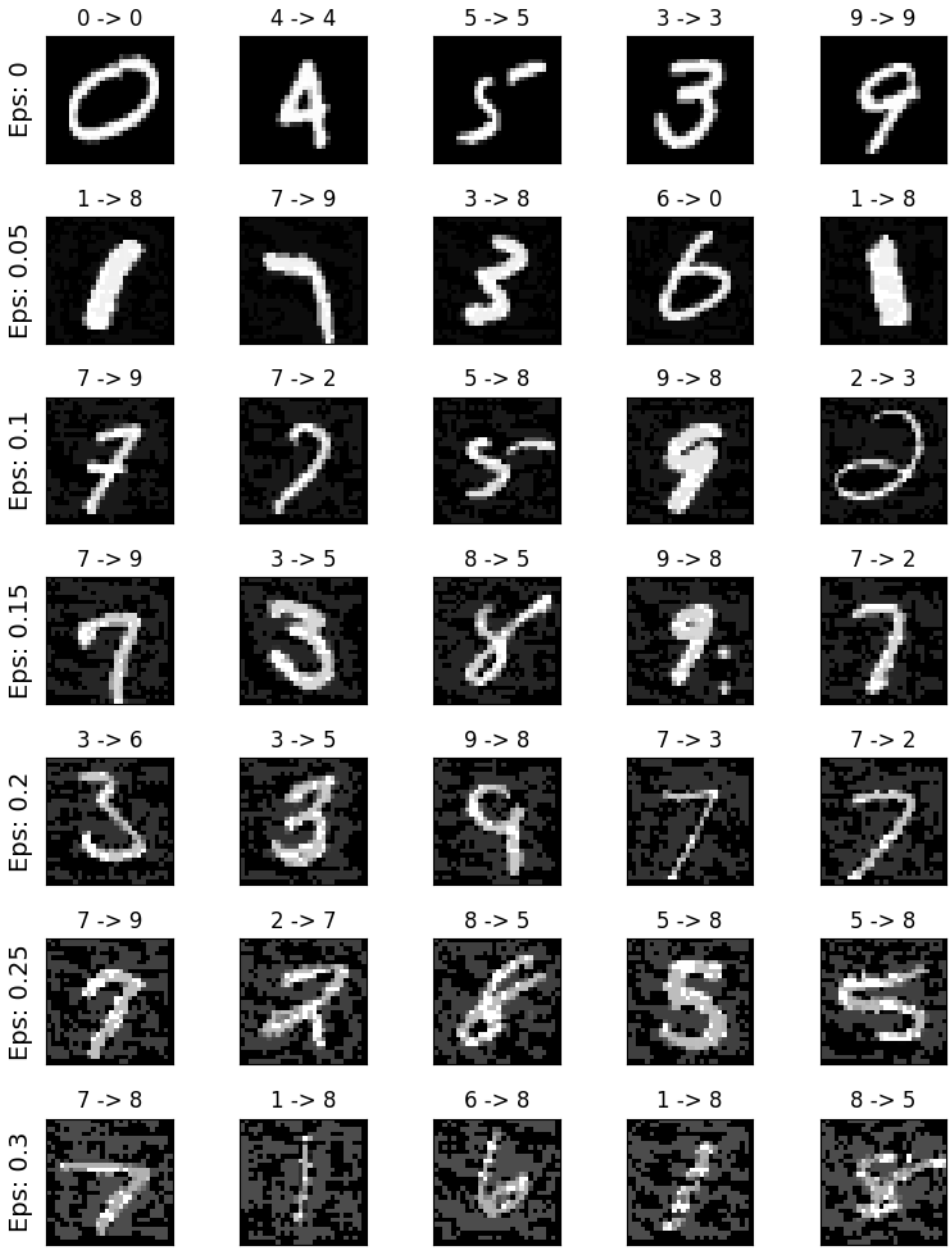
FGSM

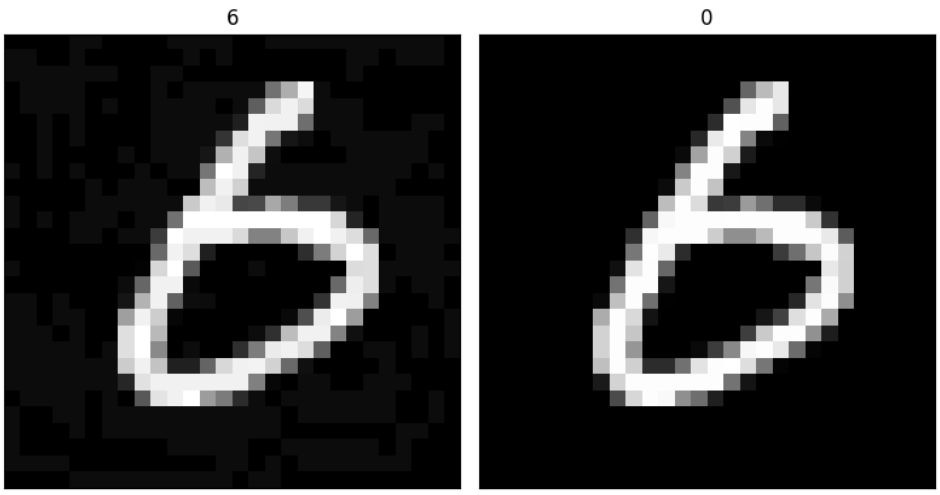
ImageNette


Eps: 0
gas pump -> gas pump
parachute -> parachute
garbage truck -> garbage truck
French horn -> French horn
tench -> tench
Eps: 0.05
tench -> English springer
tench -> chainsaw
garbage truck -> chainsaw
chainsaw -> tench
chainsaw -> parachute
Eps: 0.1
cassette player -> gas pump
gas pump -> chainsaw
golf ball -> parachute
cassette player -> English springer
tench -> French horn
Eps: 0.15
parachute -> chainsaw
tench -> parachute
English springer -> parachute
parachute -> English springer
garbage truck -> French horn
Eps: 0.2
cassette player -> gas pump
tench -> parachute
parachute -> church
tench -> chainsaw
chainsaw -> French horn
Eps: 0.25
English springer -> parachute
French horn -> church
chainsaw -> golf ball
golf ball -> parachute
tench -> English springer
Eps: 0.3
church -> parachute
cassette player -> parachute
gas pump -> parachute
French horn -> parachute
church -> parachute



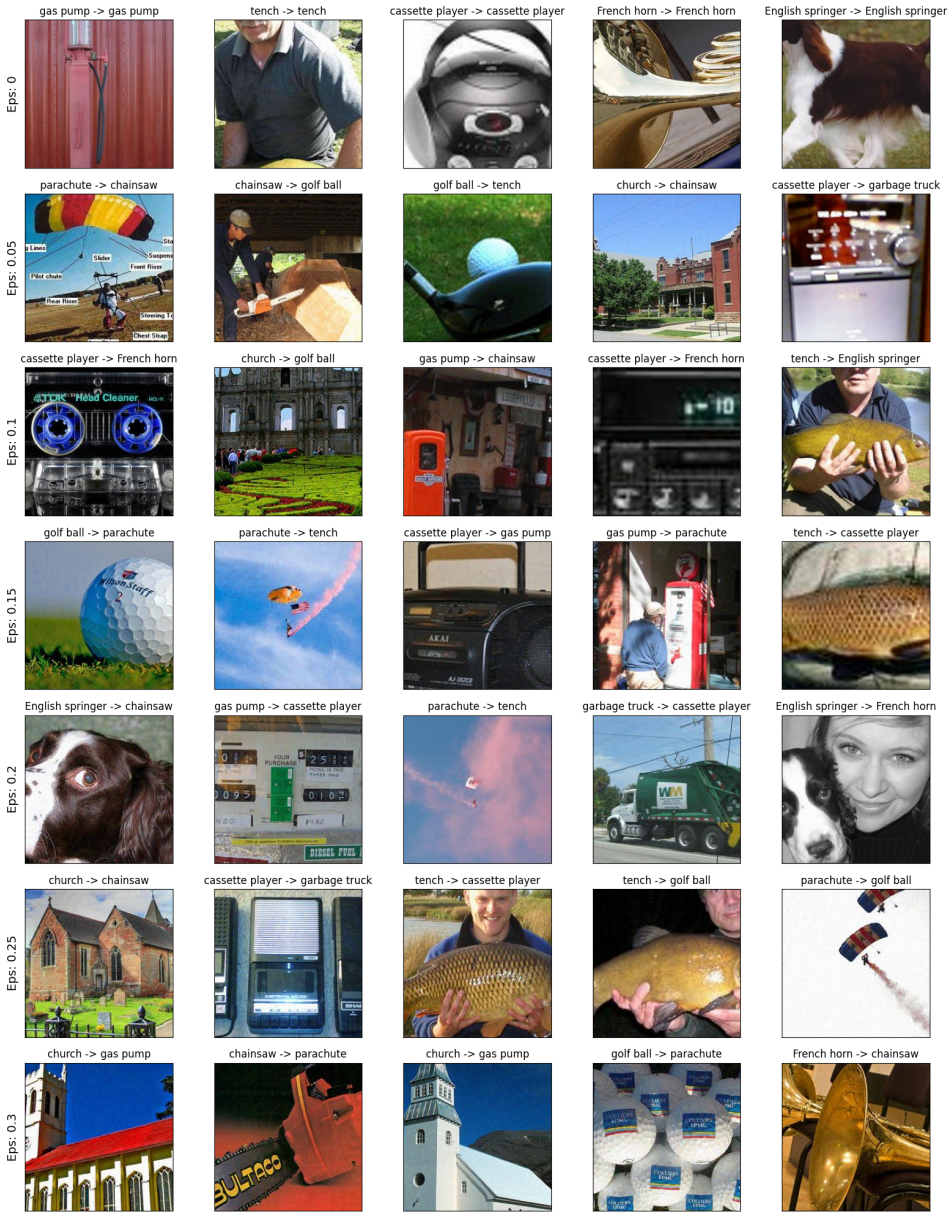
MNIST


Eps: 0
0 -> 0
4 -> 4
5 -> 5
3 -> 3
9 -> 9
Eps: 0.05
1 -> 8
7 -> 9
3 -> 8
6 -> 0
1 -> 8
Eps: 0.1
7 -> 9
7 -> 2
5 -> 8
9 -> 8
2 -> 3
Eps: 0.15
7 -> 9
3 -> 5
8 -> 5
9 -> 8
7 -> 2
Eps: 0.2
3 -> 6
3 -> 5
9 -> 8
7 -> 3
7 -> 2
Eps: 0.25
7 -> 9
2 -> 7
8 -> 5
5 -> 8
5 -> 8
Eps: 0.3
7 -> 8
1 -> 8
6 -> 8
1 -> 8
8 -> 5



IGS

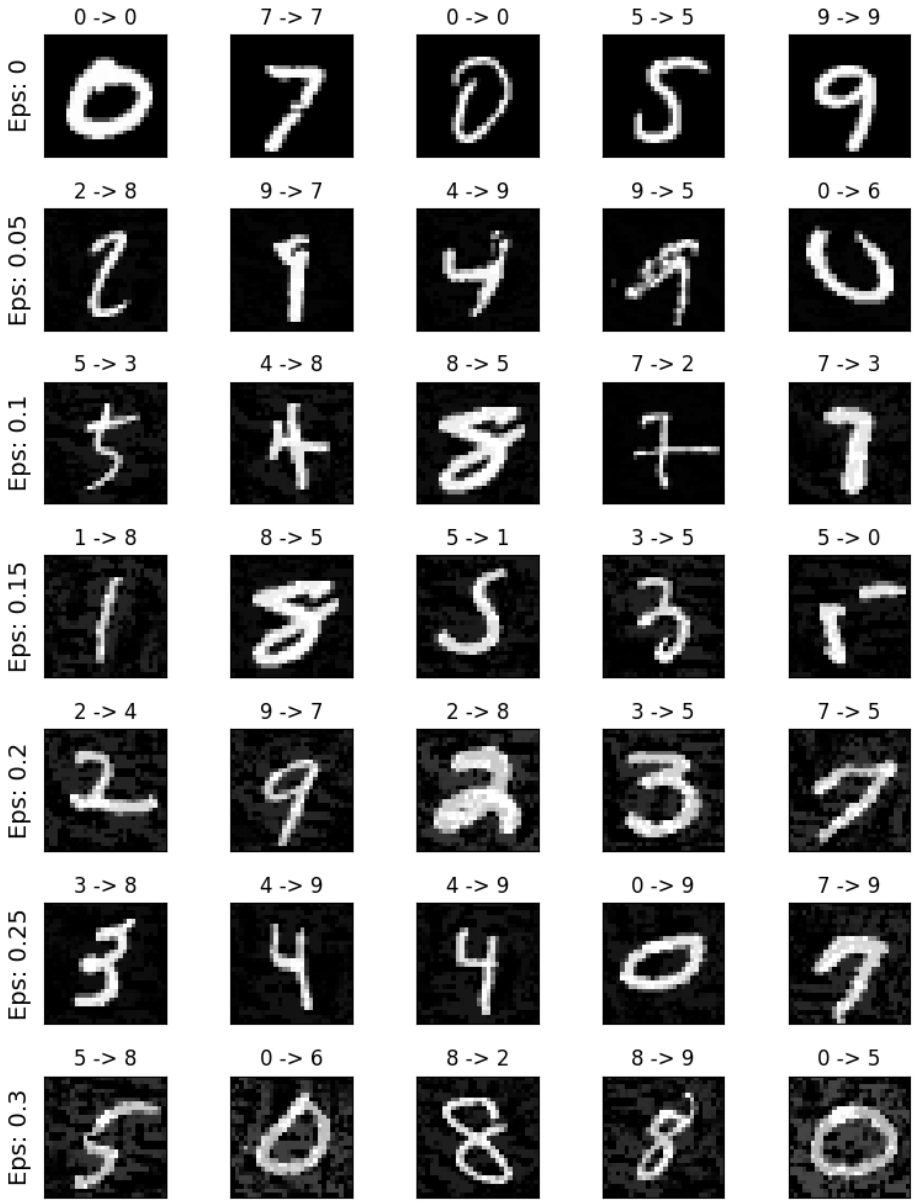
ImageNette

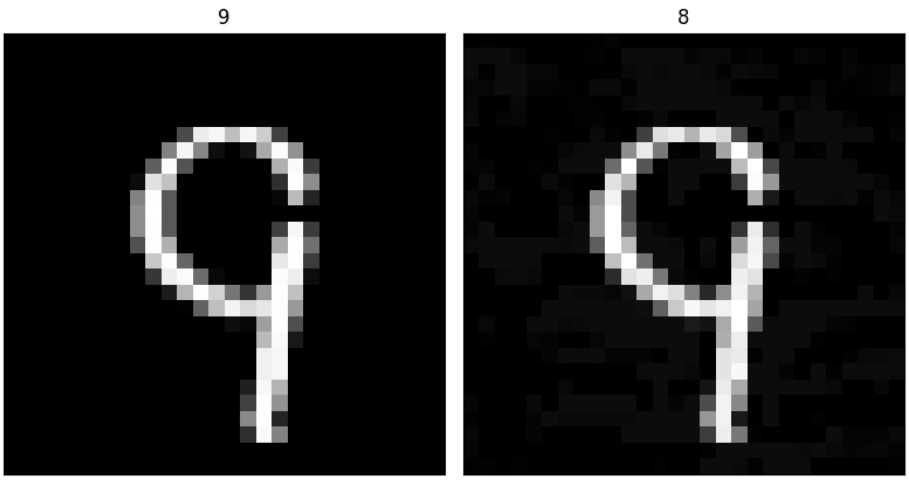

Eps: 0
gas pump -> gas pump
tench -> tench
cassette player -> cassette player
French horn -> French horn
English springer -> English springer
Eps: 0.05
parachute -> chainsaw
chainsaw -> golf ball
golf ball -> tench
church -> chainsaw
cassette player -> garbage truck
Eps: 0.1
cassette player -> French horn
church -> golf ball
gas pump -> chainsaw
cassette player -> French horn
tench -> English springer
Eps: 0.15
golf ball -> parachute
parachute -> tench
cassette player -> gas pump
gas pump -> parachute
tench -> cassette player
Eps: 0.2
English springer -> chainsaw
gas pump -> cassette player
parachute -> tench
garbage truck -> cassette player
English springer -> French horn
Eps: 0.25
church -> chainsaw
cassette player -> garbage truck
tench -> cassette player
tench -> golf ball
parachute -> golf ball
Eps: 0.3
church -> gas pump
chainsaw -> parachute
church -> gas pump
golf ball -> parachute
French horn -> chainsaw

gas pump

parachute

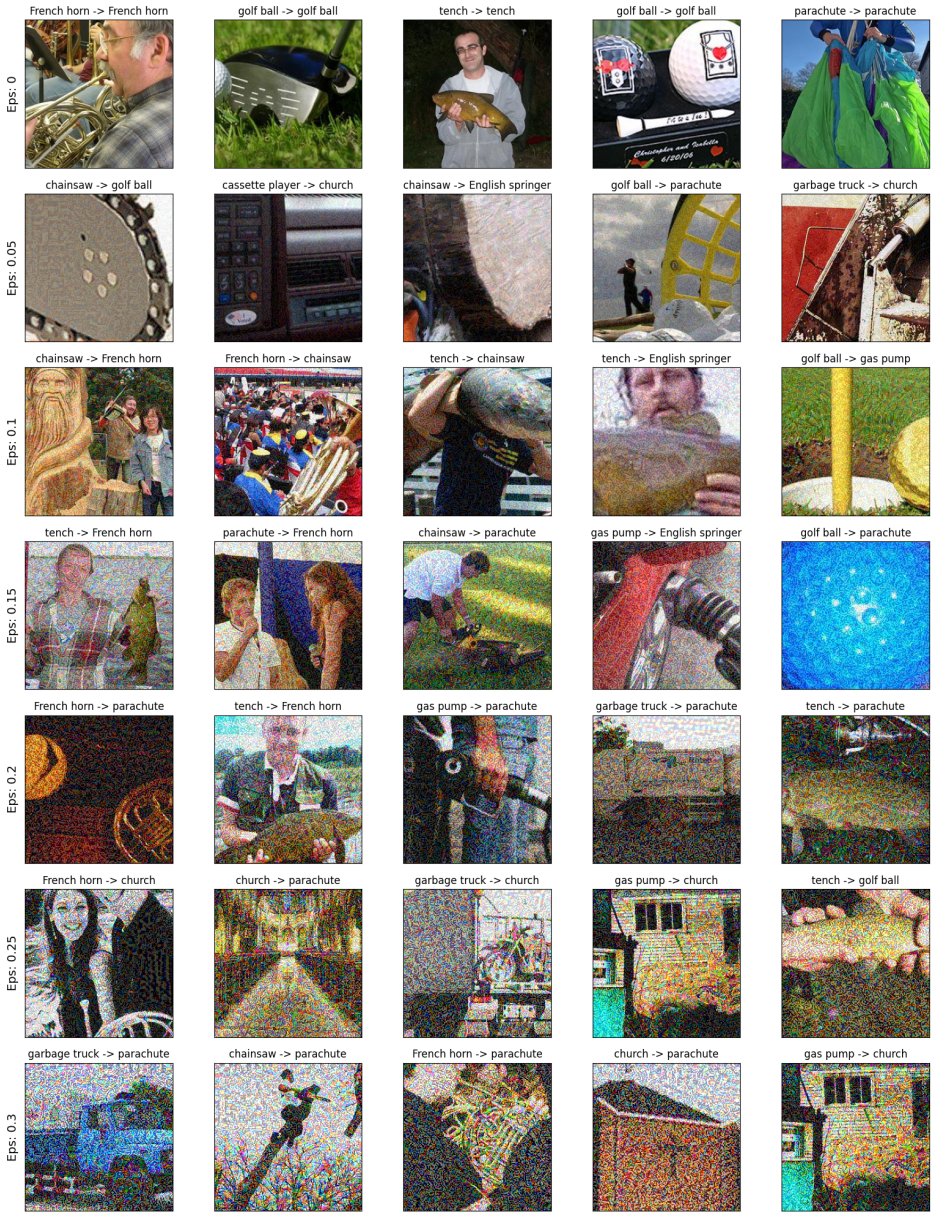
MNIST


Eps: 0
0 -> 0
7 -> 7
0 -> 0
5 -> 5
9 -> 9
Eps: 0.05
2 -> 8
9 -> 7
4 -> 9
9 -> 5
0 -> 6
Eps: 0.1
5 -> 3
4 -> 8
8 -> 5
7 -> 2
7 -> 3
Eps: 0.15
1 -> 8
8 -> 5
5 -> 1
3 -> 5
5 -> 0
Eps: 0.2
2 -> 4
9 -> 7
2 -> 8
3 -> 5
7 -> 5
Eps: 0.25
3 -> 8
4 -> 9
4 -> 9
0 -> 9
7 -> 9
Eps: 0.3
5 -> 8
0 -> 6
8 -> 2
8 -> 9
0 -> 5

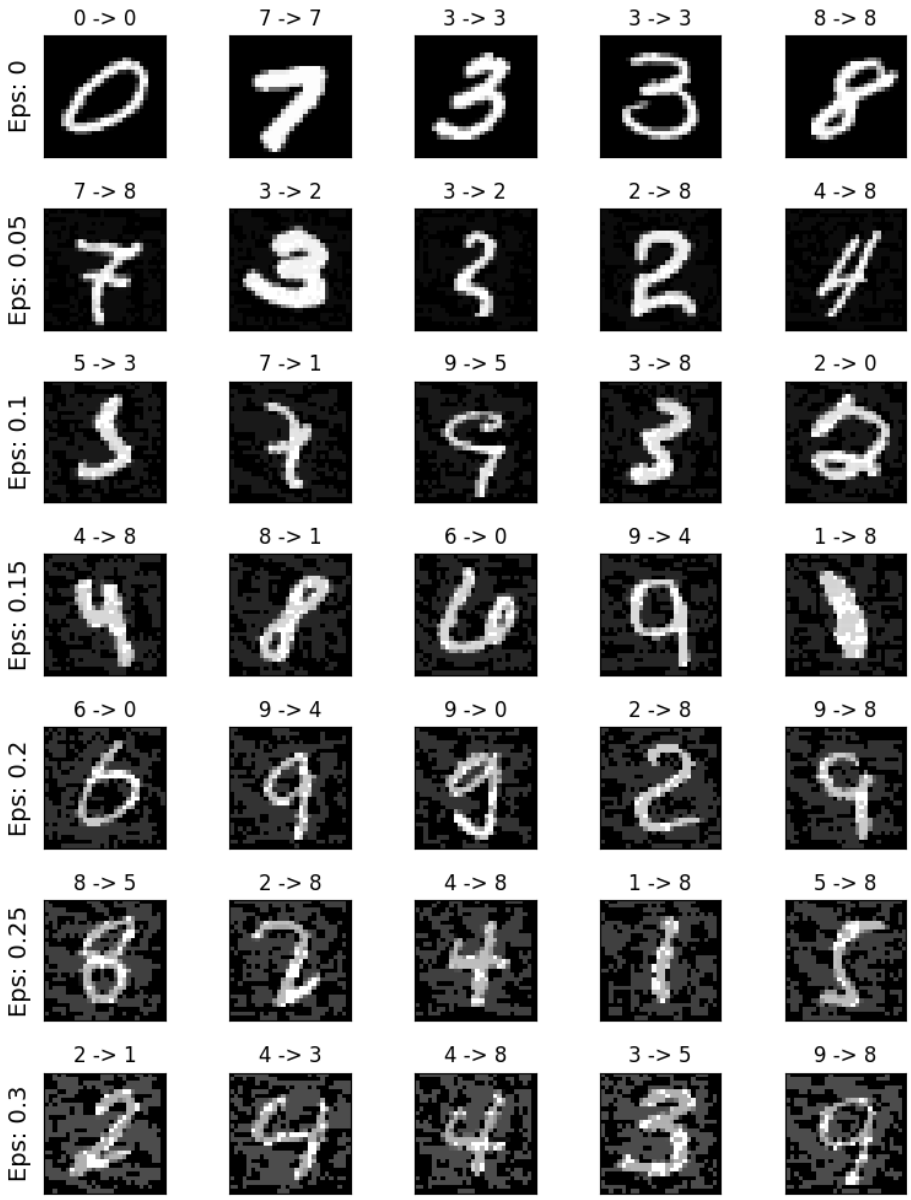


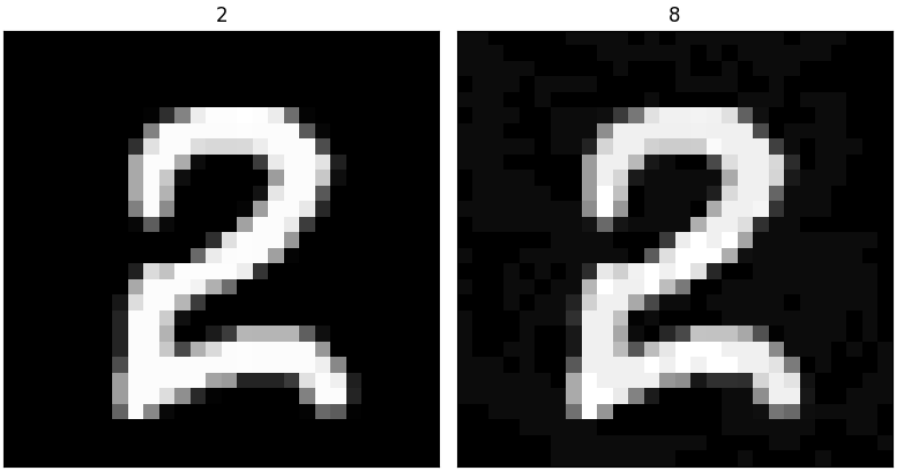
LLC

ImageNette


Eps: 0
French horn -> French horn
golf ball -> golf ball
tench -> tench
golf ball -> golf ball
parachute -> parachute
Eps: 0.05
chainsaw -> golf ball
cassette player -> church
chainsaw -> English springer
golf ball -> parachute
garbage truck -> church
Eps: 0.1
chainsaw -> French horn
French horn -> chainsaw
tench -> chainsaw
tench -> English springer
golf ball -> gas pump
Eps: 0.15
tench -> French horn
parachute -> French horn
chainsaw -> parachute
gas pump -> English springer
golf ball -> parachute
Eps: 0.2
French horn -> parachute
tench -> French horn
gas pump -> parachute
garbage truck -> parachute
tench -> parachute
Eps: 0.25
French horn -> church
church -> parachute
garbage truck -> church
gas pump -> church
tench -> golf ball
Eps: 0.3
garbage truck -> parachute
chainsaw -> parachute
French horn -> parachute
church -> parachute
gas pump -> church

MNIST


0 -> 0
7 -> 7
3 -> 3
3 -> 3
8 -> 8
Eps: 0
7 -> 8
3 -> 2
3 -> 2
2 -> 8
4 -> 8
Eps: 0.05
5 -> 3
7 -> 1
9 -> 5
3 -> 8
2 -> 0
Eps: 0.1
4 -> 8
8 -> 1
6 -> 0
9 -> 4
1 -> 8
Eps: 0.15
6 -> 0
9 -> 4
9 -> 0
2 -> 8
9 -> 8
Eps: 0.2
8 -> 5
2 -> 8
4 -> 8
1 -> 8
5 -> 8
Eps: 0.25
2 -> 1
4 -> 3
4 -> 8
3 -> 5
9 -> 8
Eps: 0.3


2
8