

(a) Superimposition (2 $\times$ ) on Original Logit

(b) Superimposition (2 $\times$ ) on Noisy Logit

Figure 1. Adversarial Images using Superimposition (2 $\times$ ) MNIST Sample 1

## Appendix

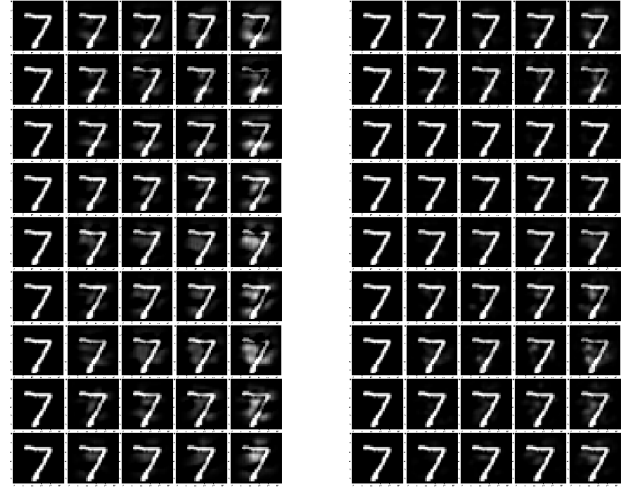
### Superimposition attacks

In Fig. 1a - Fig. 8b samples of resulting images with adversarial perturbations are shown. The leftmost column displays the original images, the middle columns display the adversarial examples with the two or three smallest perturbations, the last column shows the superimposition of the two or three adversarial examples. The rows correspond to different targets being applied in the adversarial examples. Classifications of these images are provided in Tables 1 - 4.

The images provided were generated from the superimposition attacks.

### Source codes

Source codes to generate the images are provided in the Supplementary Material folder and the steps to generate the images are described in the README file.



(a) Superimposition (3 $\times$ ) on Original Logit

(b) Superimposition (3 $\times$ ) on Noisy Logit

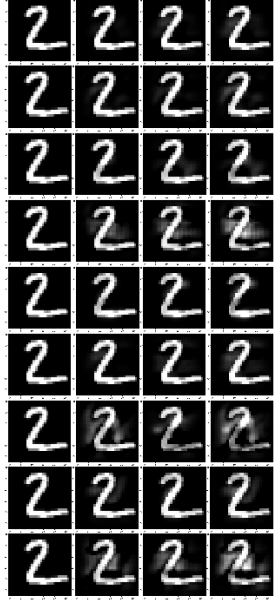
Figure 2. Adversarial Images using Superimposition (3 $\times$ ) MNIST Sample 1

Target	SI (2 $\times$ )	SI-NL (2 $\times$ )	SI (3 $\times$ )	SI-NL (3 $\times$ )
0	7	7	0	7
1	2	7	2	7
2	7	7	2	7
3	7	7	3	7
4	4	7	4	7
5	7	7	5	7
6	6	7	6	7
8	7	7	8	7
9	9	7	9	7

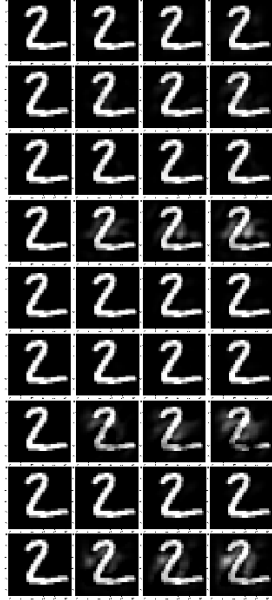
Table 1. Classifications of MNIST Sample 1, corresponding to Fig. 1a,1b,2a, 2b.

Target	SI (2 $\times$ )	SI-NL (2 $\times$ )	SI (3 $\times$ )	SI-NL (3 $\times$ )
0	2	2	2	2
1	2	2	2	2
3	2	2	3	2
4	8	2	4	4
5	2	2	5	2
6	2	2	6	2
7	7	8	7	3
8	2	2	8	2
9	8	2	9	9

Table 2. Classifications of MNIST Sample 2, corresponding to Fig. 3a,3b,4a, 4b.

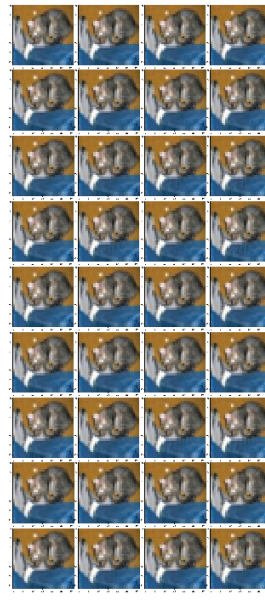


(a) Superimposition (2 $\times$ ) on Original Logit

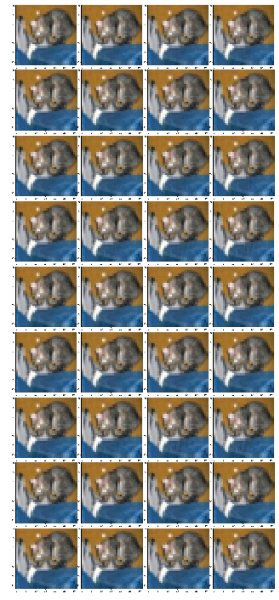


(b) Superimposition (2 $\times$ ) on Noisy Logit

Figure 3. Adversarial Images using Superimposition (2 $\times$ ) MNIST Sample 2

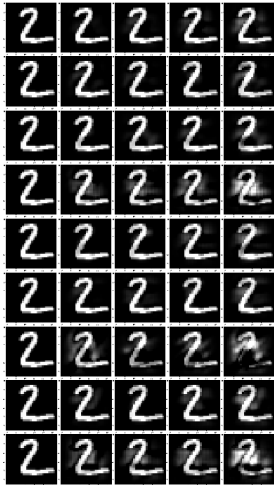


(a) Superimposition (2 $\times$ ) on Original Logit

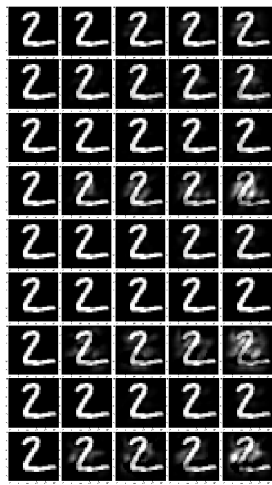


(b) Superimposition (2 $\times$ ) on Noisy Logit

Figure 5. Adversarial Images using Superimposition (2 $\times$ ) CIFAR Sample 1

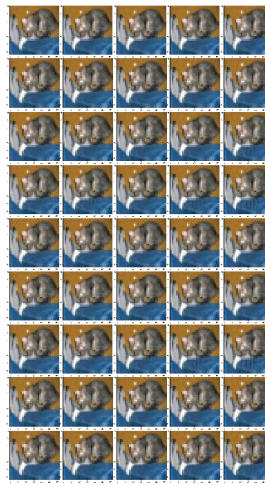


(a) Superimposition (3 $\times$ ) on Original Logit

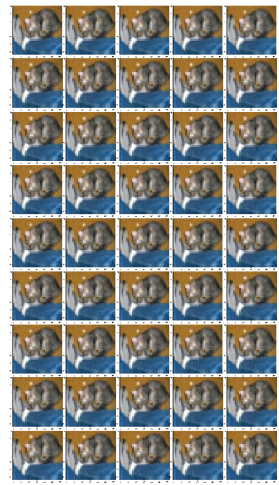


(b) Superimposition (3 $\times$ ) on Noisy Logit

Figure 4. Adversarial Images using Superimposition (3 $\times$ ) MNIST Sample 2



(a) Superimposition (3 $\times$ ) on Original Logit

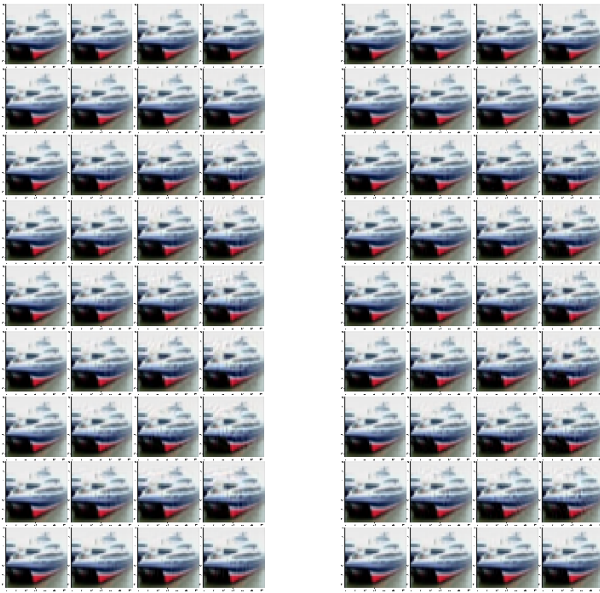


(b) Superimposition (3 $\times$ ) on Noisy Logit

Figure 6. Adversarial Images using Superimposition (3 $\times$ ) CIFAR Sample 1

Target	SI (2×)	SI-NL (2×)	SI (3×)	SI-NL (3×)
0	3	3	3	3
1	3	3	3	3
2	3	3	3	3
4	3	3	3	3
5	3	3	3	3
6	3	3	3	3
7	3	3	3	3
8	3	3	3	3
9	3	3	3	3

Table 3. Classifications of CIFAR Sample 1, corresponding to Fig. 5a,5b,6a, 6b.

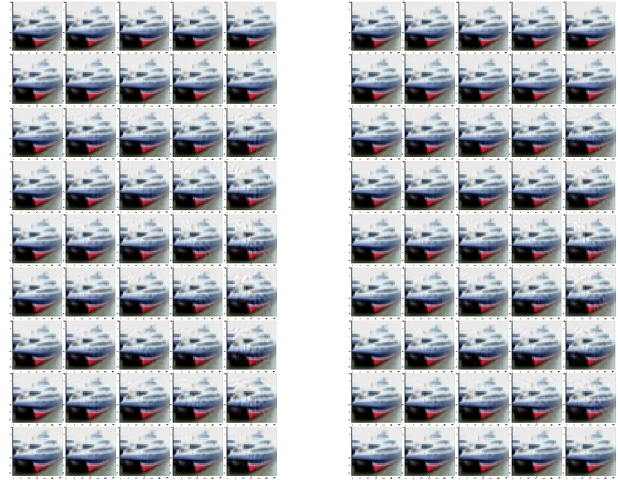


(a) Superimposition (2×) on Original Logit (b) Superimposition (2×) on Noisy Logit

Figure 7. Adversarial Images using Superimposition (2×) CIFAR Sample 2

Target	SI (2×)	SI-NL (2×)	SI (3×)	SI-NL (3×)
0	8	8	8	8
1	8	8	8	8
2	8	8	8	8
3	8	8	8	8
4	8	8	8	8
5	8	8	8	1
6	8	8	8	8
7	8	8	0	8
9	1	8	1	8

Table 4. Classifications of CIFAR Sample 2, corresponding to Fig. 7a,7b,8a, 8b.



(a) Superimposition (3×) on Original Logit (b) Superimposition (3×) on Noisy Logit

Figure 8. Adversarial Images using Superimposition (3×) CIFAR Sample 2