Classification error (%) on the first 1000 test samples

			mnist			fashion_mnist		
			original	attacked		original	attacked	
				scenario 1	scenario 2		scenario 1	scenario 2
UNENCRYPTED	CW l_2		0.97	100		8.66	100	
	CW l_0			100			100	
	CW l_∞			100			100	
	FGSM		1.5	82.94		10.62	94.25	
PERMUTATED	CW l_2		3.63	100	4.5	12.4	100	12.7
	CW l_0			100	7.3		100	
	CW l_∞			100			100	
	FGSM		3.02	89.14		12.04	91.82	
ECB	CW l_2	encrypt v1	16.58			55.66	irrelevant	irrelevant
		encrypt v2	18.11			41.97		
	FGSM	encrypt v1	20.88			59.23		
		encrypt v2	19.95			46.25		
CBC	CW l_2	encrypt v1	64.07	irrelevant	irrelevant	72.12	irrelevant	irrelevant
		encrypt v2	69.12			64.47		
	FGSM	encrypt v1	88.65			90		
		encrypt v2	88.65			90		
CTR	CW l_2	encrypt v1	88.65	irrelevant	irrelevant	90	irrelevant	irrelevant
		encrypt v2	88.65			90		
	FGSM	encrypt v1	88.65			90		
		encrypt v2	88.65			90		

scenario 1: the attacker gets an oracle to the attacked model which for given an image, performs the permutation, feeds through the model's layers and returns the logits

scenario 2: the attacker gets to knows only the architecture, i.e. he gets an oracle to the unencrypted version

accuracies of permutation on different image sizes
(padding done with 0's around the original)

	image size	error rate	min/epoch
mnist	28x28	3.63	4
	40x40	2.65	5
	60x60	2.69	12
	100x100	2.3	14
fashion_mnist	28x28	12.4	
	40x40	12.07	13
	60x60		
	100x100		