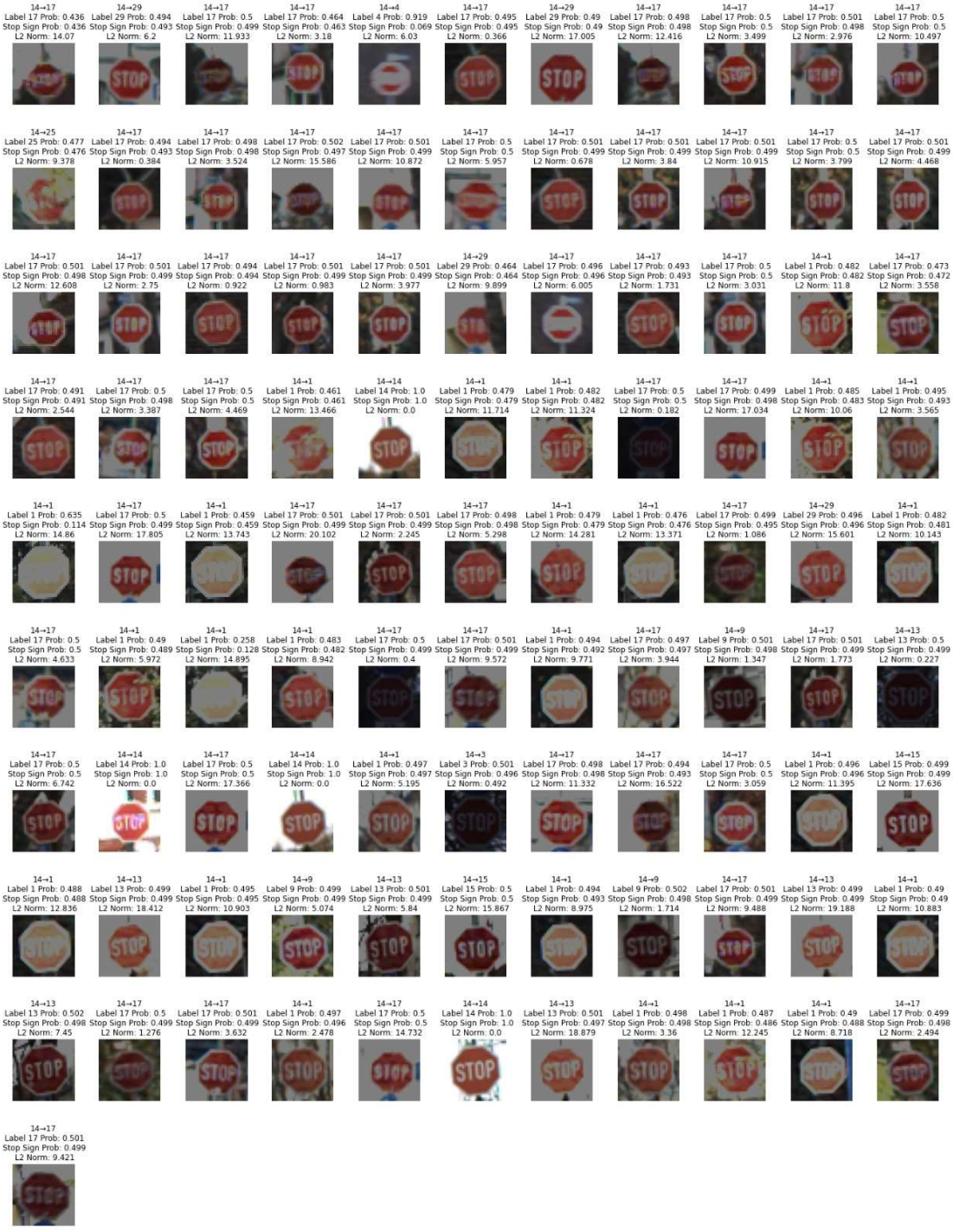
RESULTS ZOO UNTARGETED						
Initial Const	Confidence	Learning Rate	Max Iterations	Tasso Successo	Media Norme	*attack performance*
0.001	0.0	0.001	1500	97%	7.762	0.343
0.1	0.0	0.0005	1500	79%	5.414	0.338
0.001	0.0	0.0005	2000	90%	6.835	0.330
0.01	0.0	0.0005	1500	76%	5.183	0.328
0.7644	0.223	0.0250	1500	77%	5.423	0.318
0.001	0.0	0.0005	2500	97%	7.934	0.300
0.001	0.0	0.001	2000	100%	8.443	0.296
0.582	0.389	0.032	1500	61%	3.828	0.291
0.291	0.106	0.024	1500	99%	8.394	0.290
0.0001	0.0	0.1	1500	100%	8.647	0.279
RESULTS ZOO TARGETED						
0.001	0.0	0.0005	10000	100%	8.547	0.287
0.001	0.0	0.0005	6000	100%	8.588	0.284
0.001	0.0	0.0005	4000	98%	8.394	0.2805
10	0.0	0.0005	4000	98%	8.396	0.2803
1.262	0.641	0.049	4000	91%	7.678	0.270
5.246	0.0004	0.049	4000	100%	8.813	0.265
8.303	0.034	0.054	4000	100%	8.901	0.258
0.837	0.087	0.015	4000	70%	5.330	0.255
0.536	0.198	0.041	4000	82%	6.821	0.251
6.168	0.377	0.057	4000	71%	6.116	0.200

On the following pages are the images of the adversary examples generated by the best untargeted and targeted attack, with the relative class predicted by the model (expressed after the →), the probability related to the predicted class and the "stop sign" class (denoted by "Stop Sign Prob"), and the norm of the difference between the adversary image and the original image (denoted by "L2 Norm"):

Attack untargeted ZOO:



Attacco targeted ZOO: