Circumventing 7 defenses with black-box attacking.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Defense | Distortion \* | # of queries | S or F | Average distortion |
| Baseline | 0.999 | 172484 | S | 0.934(10) |
| Lid | 0.986 | 10687 | S | 0.797(100) |
| Thermometer | 0.968 | 167330 | S |  |
| Randomization | 5.444e-06 | 209 | S |  |
| Inputtransformation | 1.236e-06(jpeg) | 167824 | S |  |
|  | 2.431e-06( bitdepth) | 122375 | S |  |
|  | (quit) |  |  |  |
|  | 0.1023(crop) | 63096 | F |  |
|  | Too slow for each query(tv) |  | S |  |
| Sap | 1.675 | 8380 | S |  |
| Pixeldefend |  |  |  |  |
| DefendGAN |  |  |  |  |

\*: stop when distortion is less than 1.

The distortion C&W attack achieved is 0.005(l2), since we scale the image by dividing 255 before attacking, the equivalent distortion with C&W’s would be 1.275.