Supplementary Material for Simple Black-box Adversarial Attacks

S1 Experiment on CIFAR-10

In this section, we evaluate SimBA and SimBA-DCT on a ResNet-50 model trained on CIFAR-10. Both attacks remain very efficient on this new dataset without any hyperparameter tuning.

Figure S1 shows the distribution of queries required for a successful targeted attack to a random target label. In contrast to the experiment on ImageNet, the use of low frequency DCT basis is less effective due to the reduced image dimensionality. Both SimBA and SimBA-DCT perform similarly, with SimBA-DCT having a slightly heavier tail.

Table S1 shows aggregate statistics for the attack on CIFAR-10. Both methods achieve a success rate of 100% when limited to a maximum of $10,000$ queries. The actual required queries is much fewer, with both methods averaging to approximately 300 queries, matching the median. SimBA-DCT has a slightly worse performance compared to SimBA due its query distribution having a slightly heavier tail. Nevertheless, the average query count is in line with state-of-the-art attacks on CIFAR-10. For instance, AutoZOOM achieves a mean query count of 259 with an average L_2-norm of 3.53.

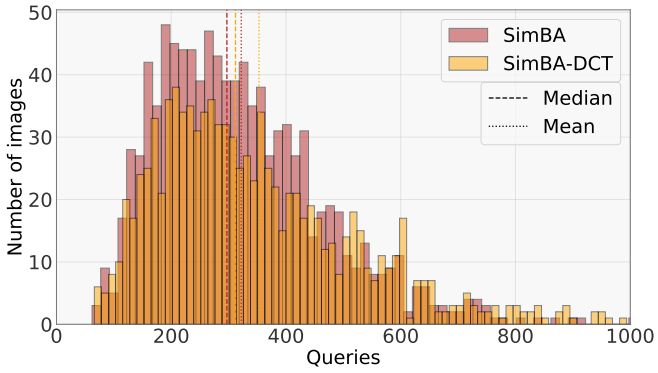


Figure S1: Histogram of number of queries required until a successful targeted attack on CIFAR-10 (over 1000 target images).

Attack	Average queries	Median queries	Average L_2	Success rate
SimBA	322	297	2.04	100%
SimBA-DCT	353	312	2.21	100%

Table S1: Average query count for SimBA and SimBA-DCT on CIFAR-10.

S2 Additional image samples for attack on Google Cloud Vision

To demonstrate the generality of our evaluation of the Google Cloud Vision attack, we show 10 additional random images before and after perturbation by SimBA. In all cases, we successfully remove the top 3 original labels.

origin_60.BMP

Dog Breed	93%
Dog Like Mammal	91%
Dog	90%
Lakeland Terrier	85%
Wire Hair Fox Terrier	84%
Terrier	83%
Snout	62%
Carnivoran	62%
Companion Dog	60%

after_60.BMP

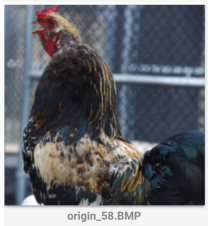
Grass	65%
Snout	63%
Terrier	60%

origin_14.BMP

Bird	96%
Fauna	89%
Branch	84%
Tree	82%
Beak	80%
Perching Bird	56%
Wildlife	55%
Twig	52%

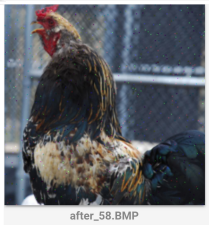
after_14.BMP

Tree	73%
Water	52%



origin_58.BMP

Chicken	98%
Beak	92%
Galliformes	91%
Rooster	89%
Feather	81%
Fowl	79%
Poultry	74%
Bird	70%
Phasianidae	58%



after_58.BMP

Snow	77%
Winter	69%

origin_59.BMP

Lighthouse	98%
Tower	96%
Landmark	89%
Beacon	81%
Sky	78%
Promontory	77%
Coast	59%
Sea	58%
Inlet	56%

after_59.BMP

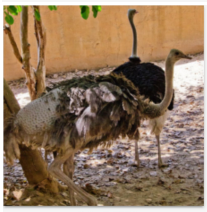
Sky	78%
Sea	73%
Computer Wallpaper	56%

origin_25.BMP

Marine Mammal	95%
Fauna	94%
Mammal	93%
Wildlife	82%
Terrestrial Animal	80%
Organism	77%
Snout	74%
Mouth	71%
Whales Dolphins And Porpoises	69%

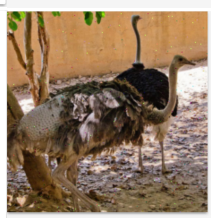
after_25.BMP

Fish	94%
Ecosystem	91%
Fish	86%
Shark	75%
Mouth	69%
Cartilaginous Fish	67%
Organism	67%
Marine Biology	63%
Jaw	53%



origin_2.BMP

Ostrich	99%
Ratite	96%
Fauna	93%
Flightless Bird	87%
Beak	81%
Wildlife	76%
Bird	75%
Emu	72%
Terrestrial Animal	62%



after_2.BMP

Wildlife	51%

origin_6.BMP

Cycle Sport	96%
Vehicle	91%
Mode Of Transport	88%
Unicycle	85%
Bicycle	77%
Flatland Bmx	76%
Sports Equipment	75%
Freestyle Bmx	75%
Bmx Bike	73%

after_6.BMP

Vehicle	81%
Cycle Sport	80%
Sports Equipment	72%
Unicycle	71%
Flatland Bmx	68%
Bicycle	68%
Bmx Bike	64%
Freestyle Bmx	64%
Recreation	58%

origin_19.BMP

Sky	87%
Energy	84%
Solar Power	78%
Solar Energy	72%
Technology	66%
Daylighting	63%
Roof	62%
Solar Panel	60%
Facade	56%

after_19.BMP

Roof	54%
Daylighting	53%

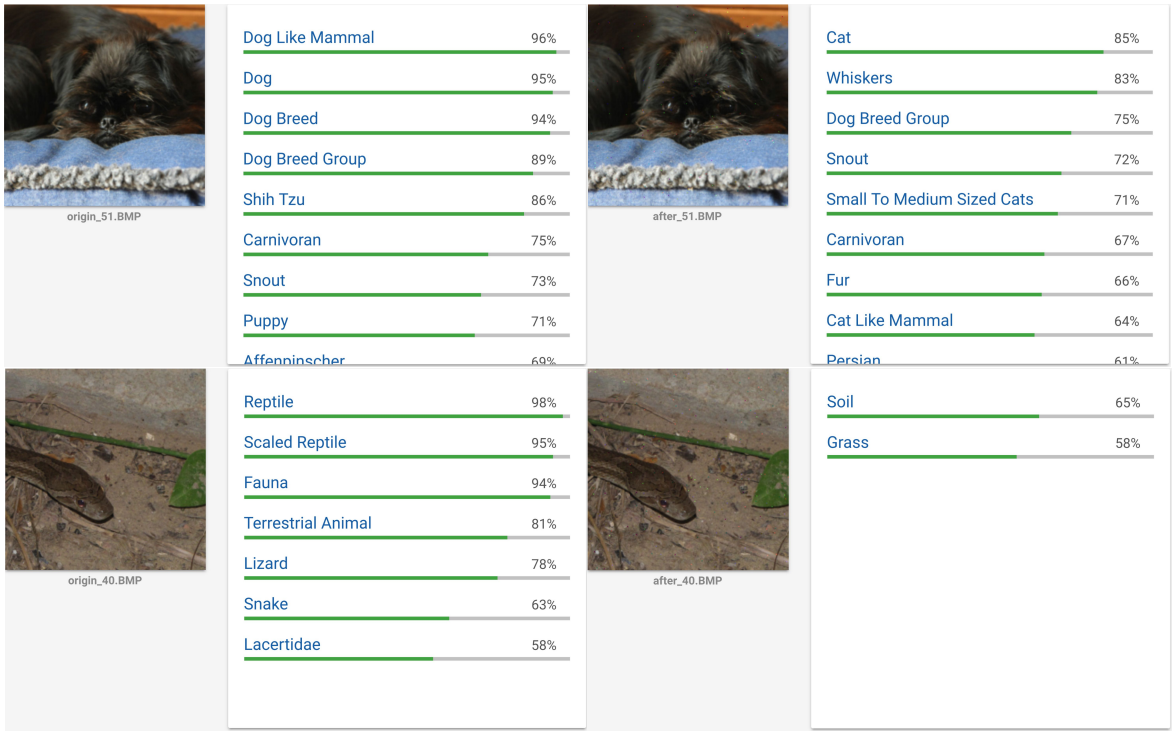


Figure S2: Additional adversarial images on Google Cloud Vision.