OMG-ATTACK: Self-Supervised On-Manifold Generation of Transferable Evasion Attacks: Supplementary

Ofir Bar Tal
Tel Aviv University
ofirbartal@mail.tau.ac.il

Adi Haviv
Tel Aviv University
adi.haviv@cs.tau.ac.il

Amit H. Bermano
Tel Aviv University
amit.bermano@gmail.com

A. Experimental Setting Details

Tables 1, 2 , 3 shows the evaluated models' hyperparameters per dataset.

Module	**# Parameters**
Generator	1.8M
Discriminator	1.6M
MNIST-CNN	694K
MNIST-TR1	629K
MNIST-TR2	893K
Resnet18	11.1M
Hyperparameter	**Value**
Batch Size	256
Max Optimization Steps	15,000
Encoder Loss Update Frequency	2
Embedding Dimension	128
Temperature	0.1
Contrastive Loss Weight	1
Budget	0.3
Generator Loss Update Frequency	1
Generator Learning Rate	0.0001
Generator Optimizer	Adam
Generator Weight Decay	0
On Manifold Loss Weight	10
Contrastive Loss Weight	2
Discriminator Learning Rate	0.0001

Table 1. Hyperparameters and the number of parameters per module used in training the OMG-ATTACK on the MNIST dataset.

Module	**# Parameters**
Generator	1.8M
Discriminator	1.6M
STN-CNN	855K
Resnet50	23.6M
Hyperparameter	**Value**
Batch Size	128
Max Optimization Steps	100,000
Encoder Loss Update Frequency	1
Embedding Dimension	350
Temperature	0.1
Contrastive Loss Weight	1
Budget	0.015
Generator Loss Update Frequency	1
Generator Learning Rate	0.0001
Generator Optimizer	Adam
Generator Weight Decay	0
On Manifold Loss Weight	5
Contrastive Loss Weight	5
Discriminator Learning Rate	0.0001

Table 2. Hyperparameters and the number of parameters per module used in training the OMG-ATTACK on the GTSRB dataset.

Module Name	**# Parameters**
Generator	1.8M
Discriminator	2.8M
Resnet18	11.3M
Resnet50	23.9M
Resnet50W	67.2M
Hyperparameter	**Value**
Batch Size	24
Max Optimization Steps	60,000
Encoder Loss Update Frequency	1
Embedding Dimension	2,048
Temperature	0.1
Contrastive Loss Weight	1
Budget	0.025
Generator Loss Update Frequency	2
Generator Learning Rate	0.0001
Generator Optimizer	Adam
Generator Weight Decay	0
On Manifold Loss Weight	10
Contrastive Loss Weight	2
Discriminator Learning Rate	0.0001

Table 3. Hyperparameters and the number of parameters per module used in training the OMG-ATTACK on the CUB-200 dataset.

B. Qualitative Results

We showcase adversarial examples generated by the OMG-ATTACK model for the various datasets.

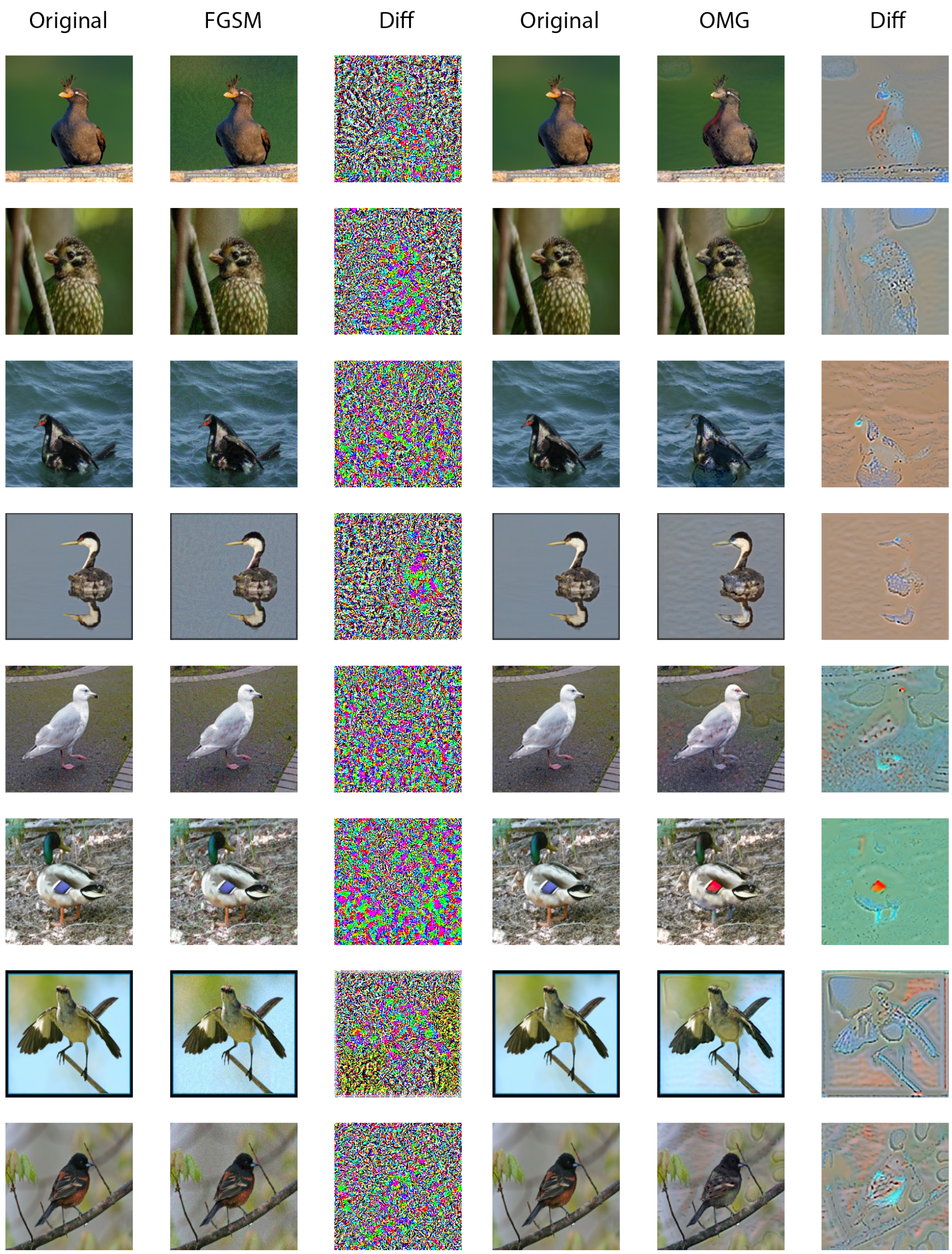

Figure 1. Evasion Attacks on CUB-200 dataset representative. On the left we have the original image, then the EAs using FGSM, then the diff. on the right side, we have the same paradigm for EAs generated using OMG-ATTACK model.

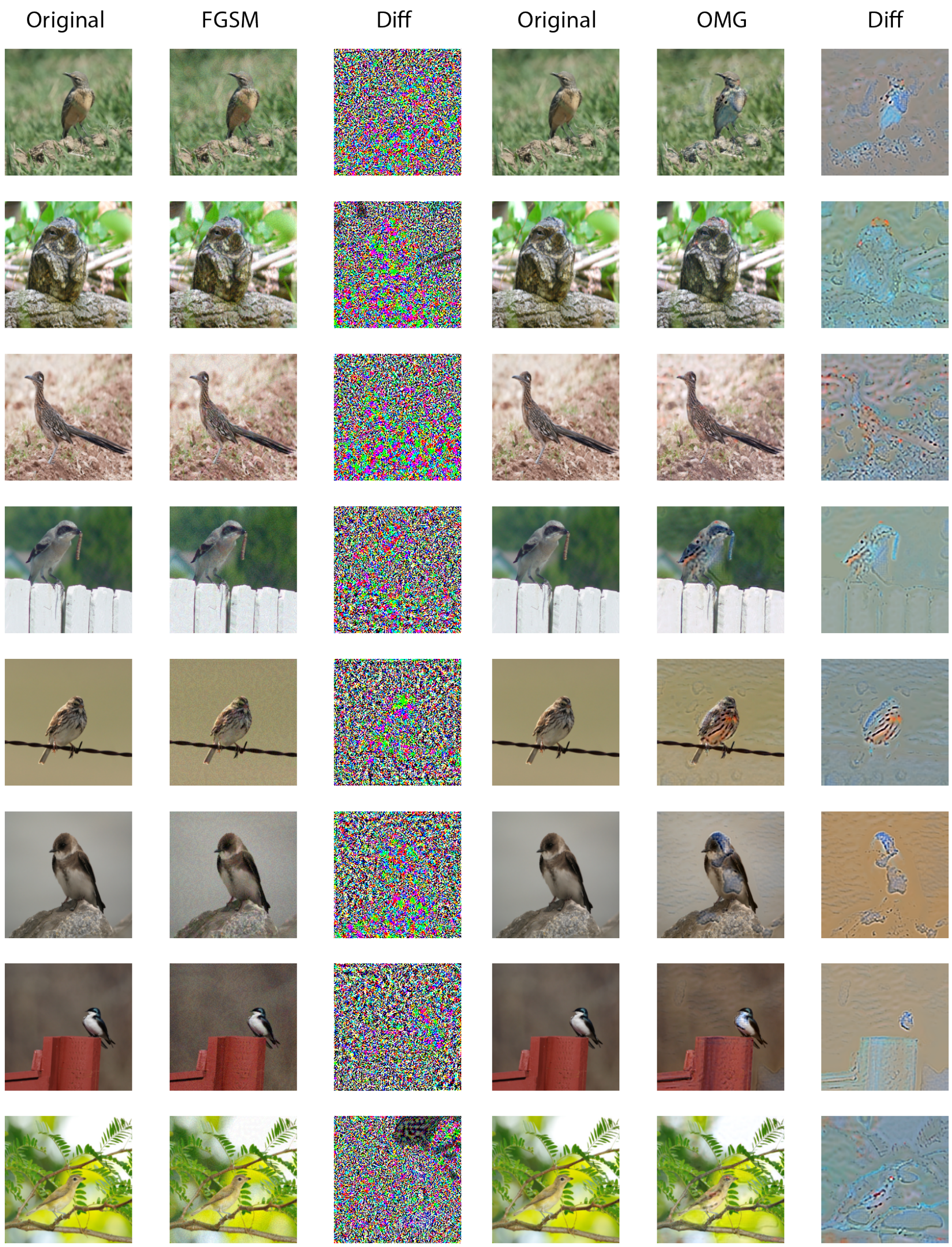

Figure 2. Evasion Attacks on CUB-200 dataset representative, Part 2. On the left we have the original image, then the EAs using FGSM, then the diff. on the right side, we have the same paradigm for EAs generated using OMG-ATTACK model.

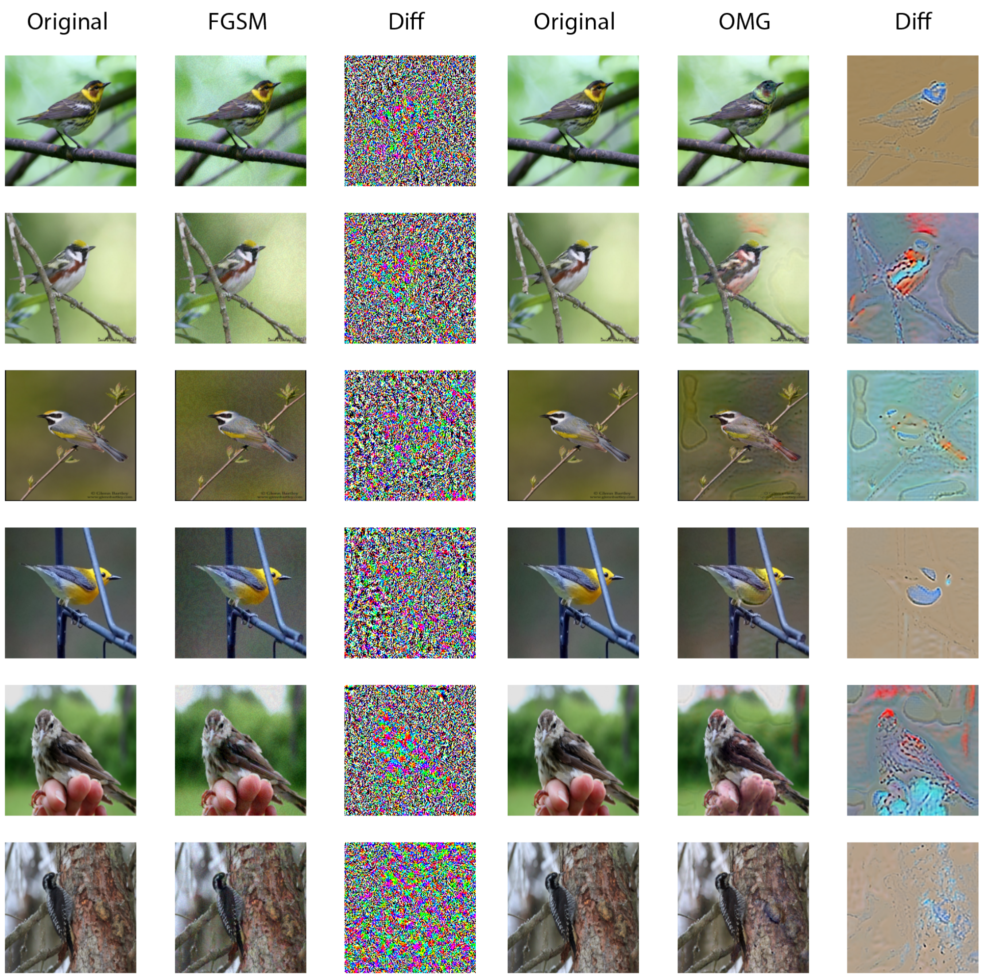

Figure 3. Evasion Attacks on CUB-200 dataset representative, Part 3. On the left we have the original image, then the EAs using FGSM, then the diff. on the right side, we have the same paradigm for EAs generated using OMG-ATTACK model.

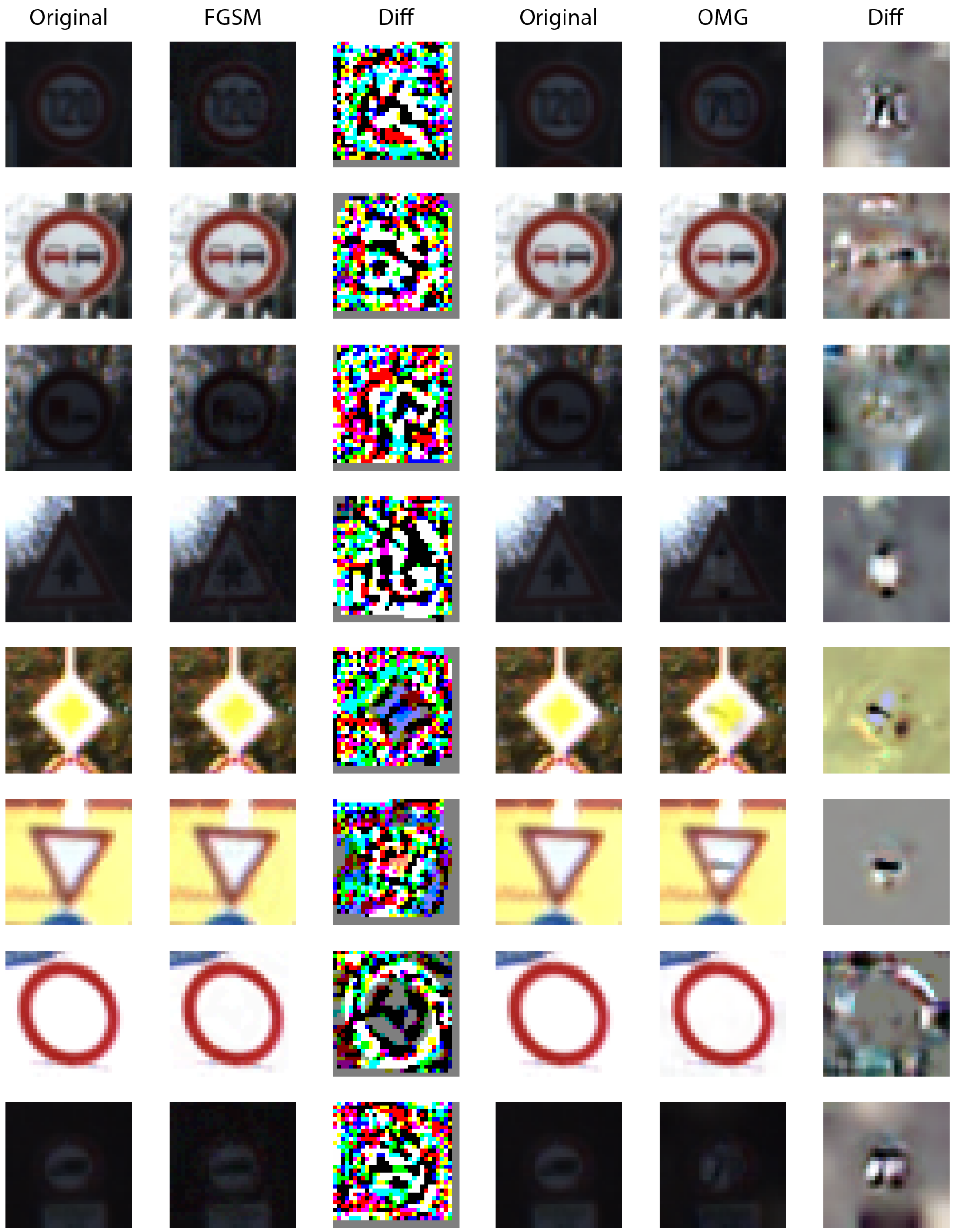

Figure 4. Evasion Attacks on GTSRB dataset representative. On the left we have the original image, then the EAs using FGSM, then the diff. on the right side, we have the same paradigm for EAs generated using OMG-ATTACK model.

Figure 5. Evasion Attacks on GTSRB dataset representative, Part 2. on the left we have the original image, then the EAs using FGSM, then the diff. on the right side, we have the same paradigm for EAs generated using OMG-ATTACK model.

Figure 6. Evasion Attacks on GTSRB dataset representative, Part 3. on the left we have the original image, then the EAs using FGSM, then the diff. on the right side, we have the same paradigm for EAs generated using OMG-ATTACK model.