FGSM Experiments

Description: Preliminary Experiment ran on Inception_V3 model. Tests basic FGSM adversarial attacks on images and detects saliency map of each using SmoothGrad.

Results: Better results when tested on images with multiple objects (Indoor/Outdoor images). Classification changes significantly and allows opportunity to create techniques that are robust against those types of images.
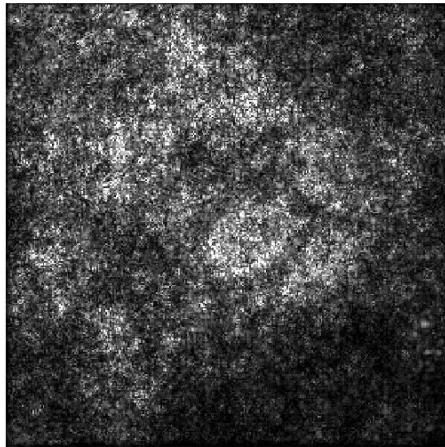
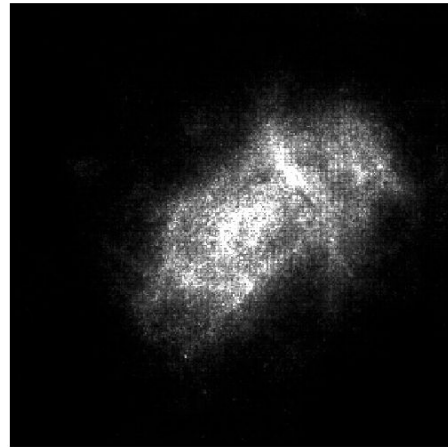**Beetle** (left- orig Rhinocerous Beetle, right - fgsm Longhorned Beetle)

Beetle Gradient (orig top, fgsm bottom)
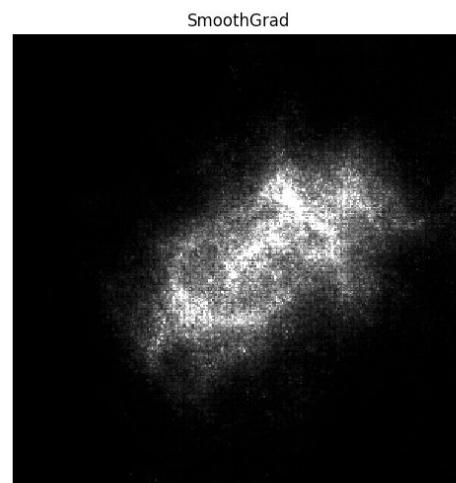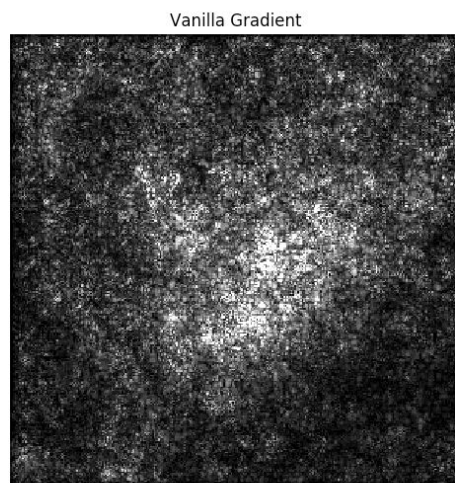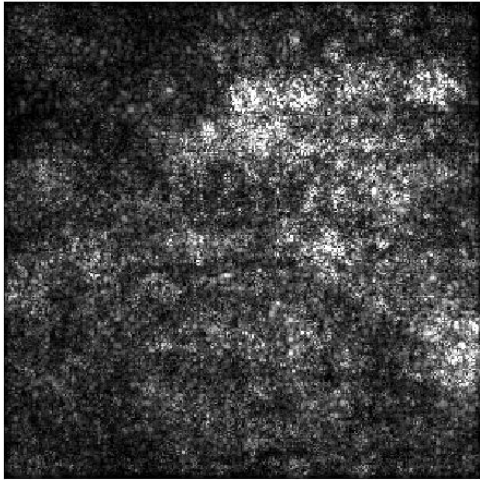
Vanilla Gradient   SmoothGrad

Outdoor (<u>left</u>- orig Patio/Terrace, <u>right</u> - fgsm Park Bench)
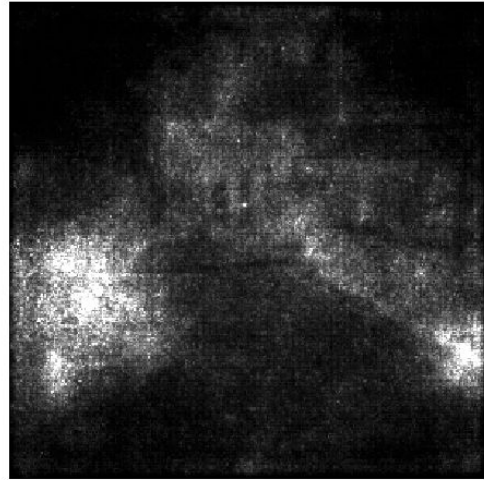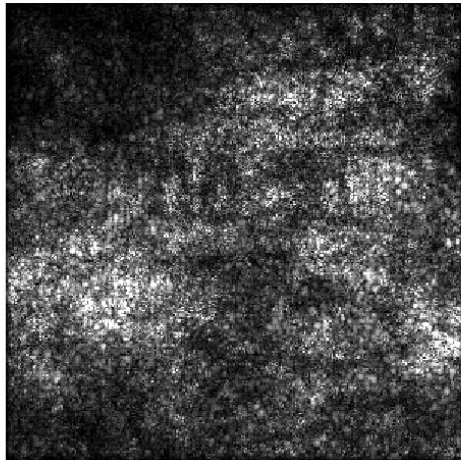
Outdoor Gradient (orig top, fgsm bottom)

Vanilla Gradient

SmoothGrad

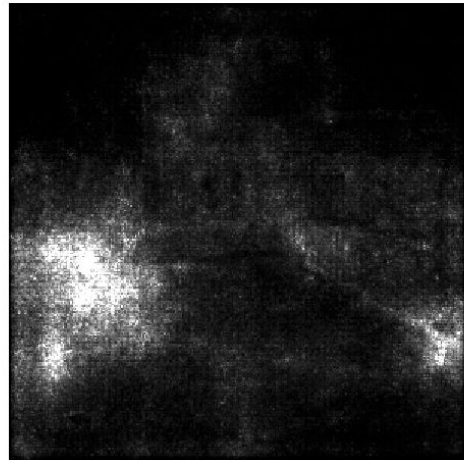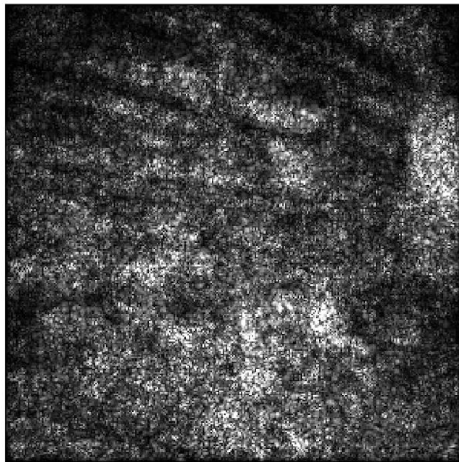## Vanilla Gradient

SmoothGrad

Indoor (<u>left</u>- orig Eating Place/Restaurant, <u>right</u> - fgsm Street Car/Trolley Car)
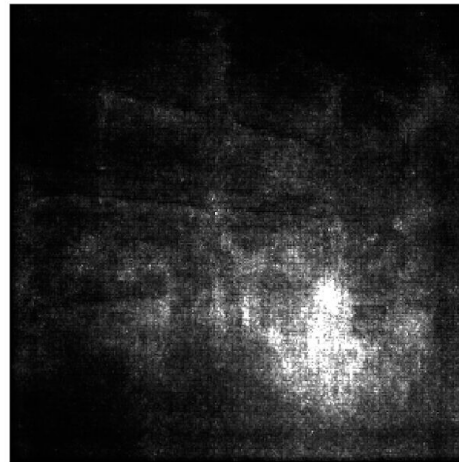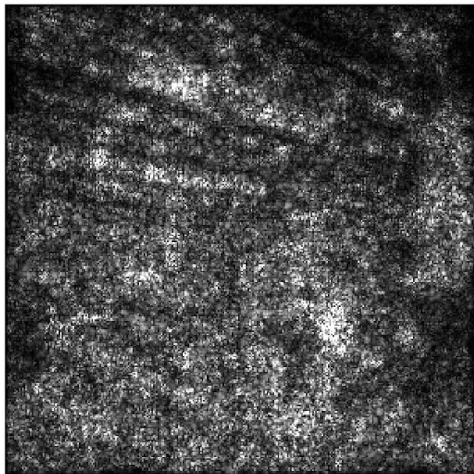
Indoor Gradient (orig top, fgsm bottom)

Vanilla Gradient

SmoothGrad

Vanilla Gradient

SmoothGrad