World Academy of Science, Engineering and Technology
International Journal of Electronics and Communication Engineering
Vol:13, No:01, 2019

Non-Targeted Adversarial Object Detection Attack: Fast Gradient Sign Method

Authors : Bandar Alahmadi, Manohar Mareboyana, Lethia Jackson

Abstract : Today, there are many applications that are using computer vision models, such as face recognition, image classification, and object detection. The accuracy of these models is very important for the performance of these applications. One challenge that facing the computer vision models is the adversarial examples attack. In computer vision, the adversarial example is an image that is intentionally designed to cause the machine learning model to misclassify it. One of very well-known method that is used to attack the Convolution Neural Network (CNN) is Fast Gradient Sign Method (FGSM). The goal of this method is to find the perturbation that can fool the CNN using the gradient of the cost function of CNN. In this paper, we introduce a novel model that can attack Regional-Convolution Neural Network (R-CNN) that use FGSM. We first extract the regions that are detected by R-CNN, and then we resize these regions into the size of regular images. Then, we find the best perturbation of the regions that can fool CNN using FGSM. Next, we add the resulted perturbation to the attacked region to get a new region image that looks similar to the original image to human eyes. Finally, we placed the regions back to the original image and test the R-CNN with the attacked images. Our model could drop the accuracy of the R-CNN when we tested with Pascal VOC 2012 dataset.

Keywords : adversarial examples, attack, computer vision, image processing
Conference Title : ICCVISP 2019 : International Conference on Computer Vision, Image and Signal Processing
Conference Location : Tokyo, Japan
Conference Dates : January 07-08, 2019

Open Science Index, Electronics and Communication Engineering Vol:13, No:01, 2019 publications.waset.org/abstracts/103308.pdf