CALIFORNIA SCIENCE & ENGINEERING FAIR
2019 PROJECT SUMMARY

Name(s)

Daniel Liu

Project Number

S1407

Project Title

Extending Adversarial Attacks and Defenses to Deep 3D Point Cloud Classifiers

Abstract

Objectives

Testing 3D point cloud neural networks to see if they are susceptible to adversarial attacks like previous results on 2D images, proposing new algorithms for both attacking and defending 3D point clouds, and explaining the effectiveness of the attacks and defenses to better understand the nature of 3D point cloud classifiers.

Methods

Two well-known networks were evaluated: PointNet and PointNet++. They were trained and tested with the ModelNet-40 and smaller ModelNet-Unique dataset. Previously proposed adversarial attacks and defenses were implemented and tested. Also, new algorithms for attacking networks using methods that make use of intrinsic properties of 3D space and the 3D network architectures are proposed.

Results

3D point cloud neural networks were found to be similarly susceptible against adversarial attacks compared to 2D image classifiers, and human-imperceptible perturbations can be generated for 3D point clouds. The defense algorithms that I proposed were also effective. 3D point cloud networks were found to exhibit a new "gradient hiding" phenomenon, which allows the correct shape of objects to be hidden from adversarial attacks that require gradients.

Conclusions

Though 3D point cloud classifying neural networks are weak against adversarial attacks, gradient hiding and other intrinsic properties of 3D point clouds and the network architectures allow 3D point cloud classifiers to be more easily defensible than 2D image classifiers. This means that 3D point cloud classifiers should be favored over 2D image classifiers in tasks that require visual data, like autonomous driving. Also, the gradient hiding theory leads to a better understanding of how 3D neural networks behave against adversarial perturbations.

Summary Statement

Examining the behavior of 3D point cloud classifiers against adversarial attacks and attempting to defend them.

Help Received

Ronald Yu and Hao Su (both from Department of CSE, UCSD) provided the topic for me to work on, and GPUs for running experiments. I also conversed with Ronald Yu on the general designs of the project. All code were written by me. Almost all algorithm ideas were designed by me.