Project Proposal

# Members

* Landon Moon
* Parker Steach
* Gilbert Lavin

# Brief description

We have heard about what is called an ‘adversarial attack on image classifiers’. The attack targets convolutional neural networks in an optimized fashion such that an image that is classified one way can be slightly changed to be classified another way. An image can look like a banana but be classified as a car with this attack.

<https://pyimagesearch.com/2020/10/19/adversarial-images-and-attacks-with-keras-and-tensorflow/> is a good example of what we are interested in, but there are a plethora of papers and articles on the subject.

We would like to first demonstrate some of the existing work on the topic and understand how it works. Afterwards we will try to find the boundaries of this kind of attack, and possibly what can be done to protect against it.