Project Proposal

Varshini Reddy
vreddy@mit.edu

March 31, 2022

1 Problem Statement

One of the grand challenges of neuroscience lies in understanding how the brain recognizes objects in the visual world. Simple neural networks and CNNs are a result of an effort at interpreting how humans learn and visualize. The aim of this project is to incorporate visual functions that uniquely enable humans (as well as other mammals) to view, interpret and distinguish images into vision models. Given that humans are immune to adversarial attacks, an additional aim is to help understand whether human motivated computer vision architectures can help in mitigating adversarial attacks.

2 Related Work

In [1], the authors discuss the effects of visual plasticity and how it positively impacts human vision. Authors in [2], use depth map of an image, instead of the actual image, as the input to CNN for identifying various human actions. Both these works talk about important features that are utilized by human vision namely, acuity and depth information. The work [3] helps give an overall understanding of human versus computer vision.

3 Dataset and Methodology

The aim of this project is to incorporate traits in human vision to machine learning models and test for improving resistance against adversarial attacks. This work mainly focuses on classification for which the mini Imagenet dataset is going to be employed. For testing against adversarial attacks, dataset from [4] will be used.

In this work test 3 main traits of human vision:

1. Visual acuity - Ability to gradually improve vision starting from blurry image with local narrow receptive fields to high resolution images with wide perception. This will be incorporated by blurring the input image to various degrees and input these processed images at various stages of the architecture. The kernel size is altered to gradually increase to ensure focus on local patterns first.
2. Rods and cones - The rods and cones in the retina aids humans to view the world, both at day and night. For this, input the original image and a gray-scaled edge drawing of that image to the network.
3. Depth information - One main advantage of human vision is that we have a 3D perspective of the world. Depth plays a vital role in both object detection and classification in humans. To this extent, input the original image along with some depth information such as depth map of that image to the network.

Further, it would be great to explore combination of these traits to check for improvement in image classification and adversarial resistance.

In this work, a custom CNN architecture will be used for the classification task. State of the art architectures such as AlexNet and VGG trained on Imagenet will be used as the baseline models. At the moment, to my knowledge there is no work that brings together such feature and architectural changes. However, work to incorporate human vision features into computer vision is being studied extensively and hence, there is a chance that these traits are independently considered.

4 Evaluation Metrics

Given that this is a classification task, accuracy and F1-Score (to support unbalanced classes) will be predominantly used for model design and development. For testing adversarial examples, uncertainty measured as a function of change in probability of class label will be used. Statistical p-test will be employed to support the validity of the results.

References

[1] Kalia A, Lesmes LA, Dorr M, Gandhi T, Chatterjee G, Ganesh S, Bex PJ, Sinha P. *Development of pattern vision following early and extended blindness* (Proc Natl Acad Sci U S A. 2014 Feb 4;111(5):2035-9. doi: 10.1073/pnas.1311041111. Epub 2014 Jan 21. PMID: 24449865; PMCID: PMC3918801).

[2] Pichao Wang, Wanqing Li, Zhimin Gao, Jing Zhang, Chang Tang, Philip Ogunbona *Deep Convolutional Neural Networks for Action Recognition Using Depth Map Sequences* (arXiv:1501.04686).

[3] Ali Borji, Laurent Itti *Human vs. Computer in Scene and Object Recognition* (Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 113-120).

[4] Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, Dawn Song *Natural Adversarial Examples* (arXiv:1907.07174).