Tim Tadros

20 July 2016

Undergraduate Research Funding Proposal

Advisor: Professor Emily Cooper

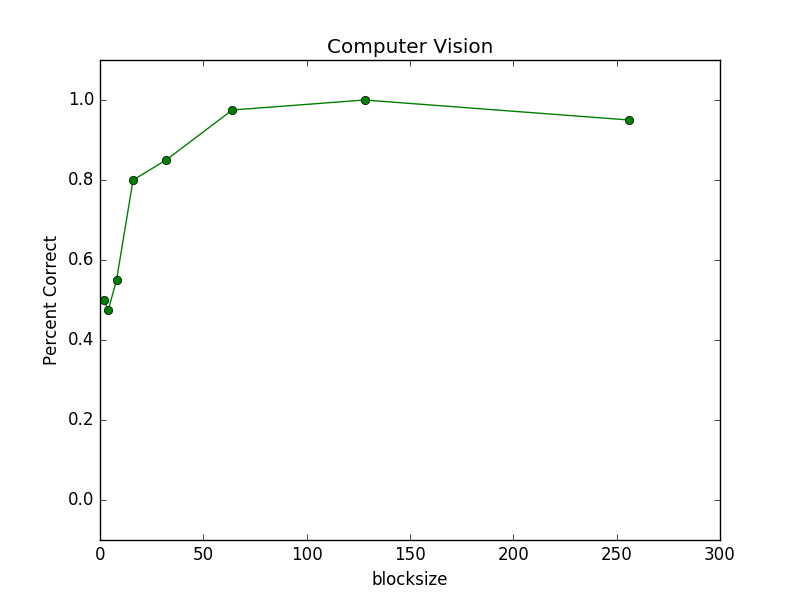
1. Project Statement

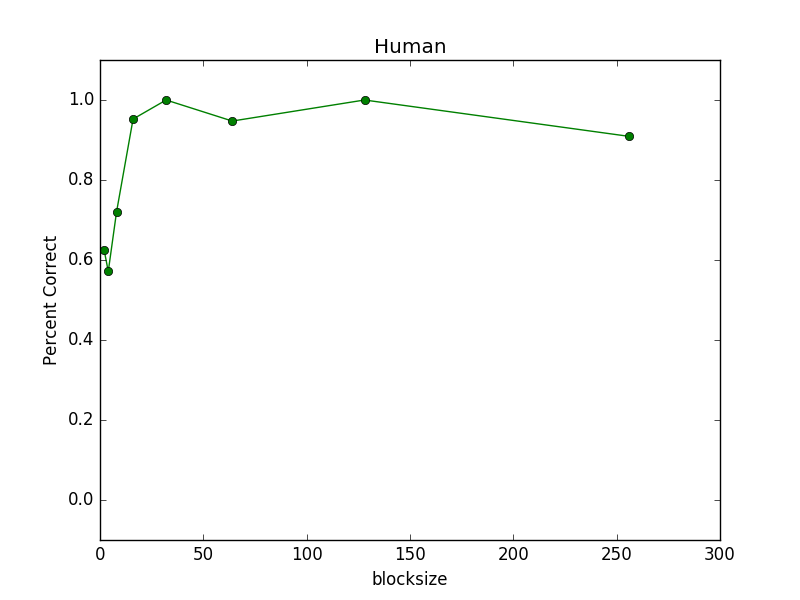
Human beings can capture the context of a scene with just a brief glance. This upcoming fall, I will be conducting research on human and computer performance on the task of image and scene recognition. Scene recognition involves looking at an image and determining what kind of scene that image depicts (for example, you may want to determine if an image is of an indoor scene or an outdoor scene). Recognizing the information in a given scene is important because it may help set up the context for how the human brain processes further information, such as the objects or the events occurring in the scene. We want to analyze how humans and computers perform on this task when the images are noisy. For example, if the images are jumbled up and rearranged, then information in the images is lost. Whereas a normal, unscrambled image contains local information about color and texture as well as global information about objects, a jumbled up image only contains the local information.

Two previous studies (Parikh 2011, Vogel 2007) have utilized jumbled images to compare human and computer vision performance at scene recognition. These papers, however, did not use neural networks for their computer vision algorithms. In the past few years, neural networks have revolutionized the field of computer vision and quickly overtaken other methods to become the current state-of-the-art for many difficult visual tasks. Because neural networks are very successful in image processing, we hope that by using neural networks to process scrambled images, we will be able to learn more about how humans process information or how the human visual process system may be lacking (if the computer algorithm performs better than human beings).

The first stage in this project will be to find and process images. We want to pick a representative set of indoor and outdoor images that captures most of the information and variance inherent in the two categories. Once we choose these images, we will process these images by chopping them up into blocks and randomly scrambling those blocks, as shown below (scrambled image on the right).

After we choose a set of images to use, we will then conduct a survey on the scrambled images to collect information about human beings’ performance on the task of categorizing scrambled images as indoor or outdoor scenes. We will show these images to people and ask them to select if they think the image depicts an indoor or an outdoor scene. We will measure how scrambling these images affects people’s ability to perform this task. At the same time, we will be using a state-of-the-art neural net trained to be highly accurate at scene recognition that allows computers to perform the same task (determine if an image is an indoor or outdoor scene) as the survey. By comparing the accuracies of human and computer performance in this manner, we hope to gain knowledge about both how humans process images and how well deep convolutional neural nets can replicate human perception. We may be able to achieve comparable levels of accuracy for human beings and computers, which may tell us that the computer algorithm used mirrors how the human visual system processes visual information.

I will be well-prepared to work on this project because I have already started working on it and I have a good understanding of the computer vision and machine learning topics involved in building a computer algorithm to perform scene recognition. As part of the Presidential Scholar’s program, I worked with Professor Cooper and we conducted a miniature version of this project in order to collect preliminary data and to make sure that the project was feasible. We ran a few people through the survey of images (we have already gotten approval from the Center for Protection of Human Subjects) and ran those same images through the computer vision algorithm and collected data. Our preliminary results are shown in the images below. The x-axis shows the size in pixels of the blocks that the images were scrambled into. The y-axis shows the percent of the images that were labeled correctly. These preliminary results suggests that the neural net performs as well as human beings do in scene recognition. 



On top of having a good understanding of the process involved in working on this project, I have also been learning about the concepts involved in building a neural network that performs the scene recognition task. I took a machine learning class this past spring which covered a lot of the basics when it comes to recognizing patterns in data (in this project, the data will be images), and learning how to use those patterns to perform a specific task (e.g. determining if an image is an indoor or outdoor image). I have also been reading about neural networks, which are one of the most successful techniques for learning from data and will allow me to make a unique contribution to this specific problem.

1. Timeline

Weeks 1-2: Collect and process (scramble images)

Week 3: Create survey on Amazon Mechanical Turk

Weeks 4-6: Collect data from Mechanical Turk and from the neural net

Weeks 7-8: Analyze data and perform statistical analyses

Weeks 9-10: Code up previous generation computer vision approaches to compare with the results achieved by the neural net, and make plan for training my own neural net for my senior thesis

1. Budget

We estimate surveys will cost around $2400. If we have 100 images in our set and each image has ~30 versions (such as scrambled, with a grid on top of it, all the different block sizes etc.), then to ensure that we are getting reliable responses, we will want to run each image version around 10 times. That's a total of 100\*30\*10 = 30,000 trials. If we split these trials into short surveys of 25 questions each, that's 1200 surveys. If we pay participants $2 per survey, that is $2400.

1. Relation to Senior Thesis

I will be continuing this project in the winter and spring of 2017 as part of my computer science honors thesis. I will apply for approval in the fall, but the funding I am asking for will not go for the terms where I am taking the Thesis Research class from the computer science department. For my senior thesis I envision using the results from the data collected in the fall as well as trying to train a neural net of my own that can identify the scenes in very noisy and scrambled images. Funding is only needed in running the human study which we hope to have done in the fall.

References

Parikh, Devi. "Recognizing jumbled images: The role of local and global information in image classification." *2011 International Conference on Computer Vision*. IEEE, 2011.

Vogel, Julia, et al. "Categorization of natural scenes: Local versus global information and the role of color." *ACM Transactions on Applied Perception (TAP)* 4.3 (2007): 19.