Research Assessment #8

Date: 5 November 2021

Subject: Methods for Improving Object Detection Through Computer Vision

MLA Citation:

Dewi, Christine, et al. “Evaluation of Robust Spatial Pyramid Pooling Based on Convolutional Neural Network for Traffic Sign Recognition System.” *MDPI*, Multidisciplinary Digital Publishing Institute, 27 May 2020, https://www.mdpi.com/2079-9292/9/6/889/htm.

Assessment:

Up until this point, I have researched a couple of methods for object detection through computer vision. However, I recognized that there were elements that could be added to these methods to make them more or less effective. One of these such methods was Spatial Pyramid Pooling (SPP).

The single biggest takeaway from this research on SPP was that it helped increase the performance of every single method that it was applied upon. For this reason, I understand that it is very important to grasp the inner workings of both these methods and added elements to know what the effect will be on a model if an element is added or taken away. Especially for my original work, in which I will propose my own method for computer vision, this will be very important to make educated conclusions and predictions.

At the same time, SPP's improvement on performance came at the expense of speed. This is yet another example of the balance between accuracy and speed, which is an important consideration to keep in mind when creating any proposal or application for a computer vision model. In addition, from my understanding, SPP's success largely rose from its simplicity. It was able to beat out many complex methods because it just split the image up into a grid and concatenated feature pools, meaning that features could all be pooled together without worrying about extra computation or maneuvering (Dewi et. al 3).

Through this research, I was also introduced to the Resnet algorithm, which was very deep but still strong in efficiency. This was very surprising because I have only seen a decrease in efficiency and speed with depth of the network, so I would like to further research about how this algorithm takes the best of both worlds. Furthermore, I was wondering, what is the bag of words approach and what is convergence? These seemed to be two prevalent concepts that could help determine what algorithms to use in a given situation and could lead me to more elements that could be added to these methods. With that said, however, it is important to keep in mind that increased elements are not always helpful and that more is now always better. The SPP method worked so well because of its simplicity, so adding extra elements may only complicate things and cause an algorithm to not run smoothly.

In addition, understanding the various problems that may occur commonly in computer vision applications will also be helpful. Knowing this will help me understand how the different elements of each method interact with each other so that I will be able to "mix-and-match" them as necessary.

More metrics that I found for comparing various methods, in addition to the mean absolute precision and detection time, are storage size, BFLOPS, and number of layers. All of these metrics will be vital to determining the ability of lower-quality machinery to run these algorithms. Since SPP is just one convolutional network, the depth of the method (in SPP's case) mostly depends on the structure of the method that SPP is implemented in. For this reason, YOLO seems to be the best choice because it has a lower storage size than other methods in the study.

The SPP is widely applicable to many methods, and may become a key part of my future research. I now have a strong direction to move into for my further research, and because of this, I'm able to see my efforts begin to come together and create a whole picture for me to evaluate in my research paper.