Research Assessment #6

Date: 21 October 2021

Subject: Computer Vision Object Detection Methods

MLA Citation:

Ahmad, Tanvir, et al. “Object Detection Through Modified YOLO Neural Network.” *Gale Academic OneFile*, MackinVIA, 6 June 2020, https://go.gale.com/ps/i.do?p=AONE&u=j043905001&id=GALE|A632058089&v=2.1&it=r&sid=sru. Accessed 21 Oct. 2021.

Assessment:

In order to compare different methods for object detection in my original work, I need to know more about them. For this reason, I chose to learn about the You Only Look Once (YOLO) Method.

Since this method proposes bounding boxes and classifies objects in one stage instead of two, I would feel much more confident in using this method or something similar to it in my final product. That is because it is much faster than a standard R-CNN, which takes extra time because it splits region proposal and classification into two separate stages. Because my focus is moving object detection, I will need a method that is as close to real-time as possible in order for it to make a decision that will react quickly enough to the object that is moving. If I chose to use something slower, such as an R-CNN, it would be too late to do anything by the time the R-CNN finished classifying objects in the image.

With this, however, comes a tradeoff. Increased speed means that the calculations themselves need to be much, much faster, leaving more room for error in the classification or feature extraction of the image. However, this should be acceptable as long as the margin of error is below a certain threshold. For small objects this may become a problem though, since even small margins of error can have a big impact on successful classification. Since smaller objects represent less space in the image, any amount of error has a greater impact on their classification than for a larger object, which takes up more pixels.

Since the YOLO algorithm does most of the work in one stage, it has trouble detecting small objects in its field of vision. This may be a problem for my final product because a moving object will likely get bigger and smaller as it moves different distances away from the camera. The deformable CNN that I learned about in my last research assessment *was* good at this, so it may be worth spending some more time to see if I can switch between the two models depending on the distance of the object.

Another consideration is the equipment used in the study. The researchers used an NVIDIA GeForce GTX 1060 graphics card, which, although priced at $300, is lower priced than many other graphics cards. As a result, it can be assumed that this is a mid-range graphics card that isn't a signifier of excessively expensive or high-tech equipment, which is an important consideration for my original work. Would a graphics card like this be able to handle switching between the two different models? Although it is relatively inexpensive, I know that it is probably the most money that I could spend on a graphics card for my final project, so knowing the extent that these cards can handle will be important to think about as well.

The next steps will be to research other types of models, such as R-CNNs, and find studies that are comparable to this one. By doing this, I will be able to compare the different methods of computer vision object detection to write my research paper for my original work.