Research Assessment #5

Date: 21 October 2021

Subject: Computer Vision Object Detection Methods

MLA Citation:

Cao, Danyang, et al. “An Improved Object Detection Algorithm Based on Multi-Scaled and Deformable Convolutional Neural Networks.” *Human-Centric Computing and Information Sciences*, Human-Centric Computing and Information Sciences, 11 Apr. 2020, https://scholar.archive.org/work/cf5wg6fm25e2rjlwz6h3rucz3e.

Assessment:

Since I am doing a research paper for my original work that compares different methods of detecting objects through computer vision, I needed to learn more about the methods themselves before I could compare them.

Through this article, I was able to gain a better understanding of the methods themselves, such as R-CNNs, YOLO and a vague understanding of SSD. As a result, I now understand that when creating my own final product, I may run into some of these big problems, such as having trouble predicting bounding boxes in a quick enough manner. If this becomes an issue, I could use a YOLO algorithm to help predict bounding boxes and classify objects inside of them in one go, which could really help processing time without sacrificing accuracy.

However, the biggest problems that I learned about were geometric transformations and the difficulty of classifying small objects. Since I am interested in

the detection and analysis of moving objects, it is conceivable to think that the objects might rotate while they are being observed by the computer vision system. This would make it difficult for most programs to classify the object, and as the article says, the previously successful YOLO algorithm “detected zero objects at this angle” when the object was turned in a certain direction (Danyang et. al 17). The same problem would occur with small objects, since any object starting from afar and moving towards the camera would appear as a small object to the computer while it is still a good distance away. However, this doesn’t mean that the computer can just ignore the object until it gets closer, since that may very well defeat the purpose of the application.

As a result, it may be useful to look more closely into deformable CNNs, which accounted for these shifts in the article. These algorithms don’t do well with larger objects though, so is there a way to switch the algorithm being used based on the size of the object at a given time?

I also noticed that there were a lot of equations used in the article that I could understand some of, but wouldn’t be able to create on my own. I am not sure that doing so would be necessary, but I do think that it would be beneficial for me to gain a greater understanding of the mathematical concepts involved with these algorithms. In fact, that was one of the biggest takeaways from my most recent interview, meaning that it is definitely worth looking into.

Finally, I more clearly define what metrics I will be comparing each algorithm with. In order to do that, I need to actually understand which metrics are frequently used in conjunction with computer vision models. After reading the article, I’ve learned that the

MAP and FPS can be very useful metrics, but are there any others that would be important to consider? Further research on this topic will help me define the scope of my research more clearly. I will also need to do more research on the PASCAL VOC library, which has hundreds of thousands of annotated training sets for machine learning models to use. Since this can be the most time consuming and tedious part of creating an ML model, the PASCAL VOC library could be a great resource for me to utilize. Overall, further research is required for me to gain a clearer picture of the options available to me, but I am now on more stable footing and know where to direct my research going forward.