Review of: "Comparing YOLOv8 and Mask RCNN for object segmentation in complex orchard environments"

Praveen Kumar Sekharamantry[1]

1 Gandhi Institute of Technology and Managment

Potential competing interests: No potential competing interests to declare.

1. The center-based approach might be less effective in fine-grained object classification and less accurate in detecting small objects. It might also have trouble with small object detection because of the higher sensitivity of the IOU rate to bounding box offsets for small objects, which can affect the bounding box's overall position. How do you deal with it?

2. The angle-loss function in YOLOv8 might penalize small targets more heavily during model training because deviations result in larger angles in the small target boxes.

3. As multiple models must be trained and maintained with different configurations, the pseudo-supervision method may require more computational resources and training time. Moreover, the complexity of managing and optimizing multiple models for various predictions may present difficulties for model deployment and maintenance.

4. What is the range at which the Intel Real-sense camera operates as at an ideal range of 7 cm to 50 cm with minimum object detection down to <1 mm at 7 cm only the camera operates generally. This variation would result in differences in accuracy.

5. In the experiment of segmentation of tree trunks and branches with multi class and single class how did you overcame the background tree trunks and branches, what method is employed to identify the foreground objects.

6. One of the major issues with YOLOv8 is it will produce false positive detection's, as it will identify objects incorrectly that are not in the image like leaves. As the present dataset is pertaining to green apples fruit it's obvious that sunlight or low light on leaves would result as apples detected. To mitigate this problem, what method or technique are employed to classify the objects of interest and objects that do not have objects of attention? This way, the model will be able to distinguish between objects it should detect and ones it shouldn't.

7. Are there no noise or stream distortion when using YOLOv8 for real-time detection on videos. If so how did you handle it?

8. If the model is adapted with the mere YOLOv8 model, the occlusions, overlapping and branching problems would arise with apples detection. Is there any particular measure considered in the work.