

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

Steven Frank

Potential competing interests: No potential competing interests to declare.

This paper compares the performance of YOLOv8 with Mask-RCNN in single-class and multi-class segmentation tasks in apple orchards. The paper is well-written and well-organized, providing a thorough discussion of the object-detection systems tested, the instance segmentation tasks evaluated, and the results obtained. The references are timely and comprehensive. The authors' conclusion is consistent with my own work, which found YOLO to be superior to Mask-RCNN and Faster-RCNN for detecting breast masses in mammograms. (In my case, YOLOv5 outperformed YOLOv8, even with SAHI enhancement.)

I have a couple of suggestions. First, there is some apparent inconsistency in 3.3, which first mentions a learning rate of 0.001 and a batch size of 32, then in the same paragraph refers to an initial learning of 0.01 using weight decay and a batch size of 16. This should be corrected or explained.

More substantively, the paper argues persuasively for the superiority of YOLOv8 over Mask-RCNN for the tasks explored, but the authors may overstate the case in considering practical utility in the real world. For example, the authors state, "Both YOLOv8 and Mask R-CNN models can provide [[practically → practical]] application segmentation results for apple tree canopy images acquired in dormant and early growing seasons." This conclusion cannot be drawn solely from performance on the validation sets; the authors need to test their best-performing models on an out-of-sample set to judge true performance on the instance segmentation tasks. The reported results may reflect overfitting to the validation sets – not a problem for comparing models, but insufficient to assess real-world performance.

One minor note – Eq. 5 should be corrected so that the parentheses match up.