

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

Huanxue Zhang<sup>1</sup>

<sup>1</sup> Shandong Normal University

**Potential competing interests:** No potential competing interests to declare.

## Reviewer comments:

### Major points

The article entitled "Comparing YOLOv8 and Mask RCNN for object segmentation in complex orchard environments" compared the one-stage YOLOv8 and the two-stage Mask R-CNN machine learning models for instance segmentation under varying orchard conditions across two datasets. The topic is of interest to readers. The authors state that they (1) compared the performances of YOLOv8 and Mask R-CNN models in segmenting single-class objects, specifically green apples (fruitlets), in images collected from variable orchard environments in the early growing season; and (2) evaluated the capabilities of these two models in segmenting multi-class objects, specifically primary branches and tree trunks of apple trees in images collected from a model apple orchard during the dormant season.

This paper shows that the authors have done a lot of work. However, the architectures of the YOLOv8 and Mask R-CNN were proposed earlier. Thus, the paper lacks novelty, and its methodological contribution is limited in terms of new neural network architectures for instance segmentation.

Several issues remain unclear and should be addressed before the paper can be accepted for publication. Detailed comments and suggestions are provided below to the authors to improve the manuscript.

### Several other points

#### 1. The "Abstract" section

The inspirational conclusion is not concise, and the significance of the research is unclear.

#### 2. The "Introduction" section

1. The reason for choosing "complex orchard environments" as the instance segmentation scene or the characteristics of them is not clearly stated.
2. The first five paragraphs repeatedly emphasize the advantages and applications of deep learning and instance segmentation, with unclear logical organization.
3. The sixth paragraph, "Unlike Mask R-CNN, a two-stage model suitable for segmentation tasks, YOLO optimizes the overall processing ensuring speed and efficiency crucial for real-time applications in agriculture such as robotic

pruning, thinning, and pesticide application," has already introduced the efficiency issues of the two models, weakening the significance of this research.

4. The existing problems in this area of research or the substantive significance of conducting this research need to be addressed in this section.

### 3. The “Background” section

The two-stage nature of Mask R-CNN is not reflected.

### 4. The “Materials and Methods” section

1. The paper involves instance segmentation with different models, under different environmental (lighting) conditions, and with different recognition objects. Section 3.1's introduction to the orchard does not sufficiently reflect the complexity of the orchard.
2. There may be issues with the "Trunk" sample annotation in Figure 4.
3. Needs to add images showing the advantage of Mask R-CNN in recognizing objects in obscured scenes and explain why this model is still the preferred choice.
4. Paragraph 4 of Section 4.2 is repetitive and consistent with the previous content.
5. The content in paragraph 5 of Section 4.2 is compared with other researchers' studies, which is not highly comparable. This may be because model accuracy must consider many factors such as sample quantity, sample quality, parameter settings, etc. The experiments of others were not conducted in the same or similar environment, making the comparison not strong. It is suggested to add local comparative experiments to enhance persuasiveness.

### 5. The “Conclusion” section

The final paragraph does not specify the different orchard conditions.

### 6. Other points

1. How is the model's transferability? Need to add validation experiments of other orchard scenes.
2. Need to read through the whole article, organize logic and language, and avoid unclear and inappropriate expressions as well as repetitive issues.
3. Attention should be paid to the consistency in the expression of key terms throughout the text, such as "Mask R-CNN".

Need to pay attention to formatting issues.