# Computer Vision Tasks with YOLOv8 Model

**Antonio Sánchez Salgado**
Universidad Carlos III de Madrid
100428745@alumnos.uc3m.es

**Alexia Durán Vizcaíno**
Universidad Carlos III de Madrid
100429771@alumnos.uc3m.es

## Abstract

Object detection and semantic segmentation are fundamental tasks in computer vision, with applications spanning numerous fields. This paper presents a comparative study between Faster R-CNN and the latest YOLOv8 model using the PASCAL VOC dataset. The results highlight YOLOv8's superior performance in both object detection and semantic segmentation, demonstrating significant improvements in precision, recall, and F1 scores across various classes. These findings suggest that YOLOv8 is a robust and highly effective model for real-time object detection.

## 1 Introduction

Object detection is a fundamental task in computer vision with widespread applications. Traditional methods like Faster R-CNN have established benchmarks for accuracy by combining classification and precise localization of objects. However, the need for models that can achieve high accuracy while maintaining real-time performance has driven the development of new approaches. The Ultralytics YOLO (You Only Look Once) series of models has been at the forefront of this evolution, emphasizing both speed and accuracy. This paper compares both methods in terms of object detection by terms of various performance metrics.

### 1.1 Faster R-CNN

Faster R-CNN, introduced by Ren et al. [1], is a robust deep learning model for object detection that integrates object classification with precise object localization. It incorporates a Region Proposal Network (RPN) to generate refined bounding boxes and classify objects. Utilizing a pre-trained ResNet-50 for feature extraction and a Feature Pyramid Network (FPN) to handle objects of various scales, Faster R-CNN achieves high precision in object detection and localization.

### 1.2 YOLOv8

YOLOv8 represents the latest stable version in the YOLO series. Known for its real-time performance, YOLOv8 excels in both speed and accuracy in predicting bounding boxes and class probabilities. The model's capabilities extend beyond object detection to include classification, tracking, segmentation, and pose estimation [2]. YOLOv8 has been fine-tuned for both object detection and semantic segmentation tasks.

## 2 Methods

We used the PASCAL VOC 2012 dataset [3] from the previous lab session, adapting the information into the YOLO format. This format requires a unique text file for each image, with different rows containing the class label and (1) the normalized coordinates for the bounding boxes, for object detection; or (2) the polygon, the normalized bounding coordinates around the mask area, for semantic segmentation. We then fine-tuned the YOLOv8 pre-trained model with our adapted dataset, experimenting with various hyperparameters. Optimal performance was attained with 50 epochs, a batch size of 18, and using the SGD optimizer.

## 3 Results and Discussion

### 3.1 Object Detection

#### Precision and Recall Trade-off

The performance metrics for Faster R-CNN and YOLOv8 provide insights into their precision and recall trade-offs, see Table 2. Faster R-CNN's Objectness-RPN metric evaluates the ability of its RPN to suggest regions containing objects, while the Global Classification metric assesses overall classification accuracy. Meanwhile, YOLOv8's metrics focus on precision and recall for detected bounding boxes. Additionally, we calculated the F1 score for YOLOv8 using the formula:

$$F1 = 2 \cdot \frac{P \cdot R}{P + R}$$

where $P$ is the precision and $R$ is the recall.

| | Faster R-CNN | | | YOLOv8 | | |
|---|---|---|---|---|---|---|
| **Class** | **Recall** | **Precision** | **F1** | **Recall** | **Precision** | **F1** |
| Bottle | 0.722 | 0.742 | 0.732 | 0.704 | 0.927 | 0.799 |
| Chair | 0.653 | 0.365 | 0.468 | 0.827 | 0.708 | 0.763 |
| Dining Table | 0.600 | 0.391 | 0.473 | 0.733 | 0.638 | 0.682 |
| Sofa | 0.592 | 0.592 | 0.592 | 0.852 | 0.921 | 0.885 |

Table 1: Faster R-CNN and YOLOv8 class-wise performance metrics.

YOLOv8 significantly outperforms Faster R-CNN in F1 score, achieving 0.788 compared to Faster R-CNN's 0.576. Precision indicates the proportion of correctly detected objects, while recall measures how many actual objects are detected. This indicates YOLOv8's superior ability to accurately classify objects overall.

| Metric | Objectness-RPN | Global Classification |
|---|---|---|
| Precision | 0.568 | 0.523 |
| Recall | 0.794 | 0.642 |
| F1 | 0.662 | 0.576 |

Table 2: Faster R-CNN performance metrics.

| Metric | Box |
|---|---|
| Precision | 0.798 |
| Recall | 0.779 |
| F1 (calculated) | 0.788 |

Table 3: YOLOv8 performance metrics.

### Class-wise Performance

A class-wise comparison using precision, recall, and F1-score performance metrics is shown in Table 1 for both models. For Faster R-CNN, the F1 scores vary significantly across different classes, with "bottle" achieving the highest value of 0.732, indicating robust detection. However, "chair" and "dining_table" show lower F1 scores, reflecting the model's difficulty in these classes. In contrast, YOLOv8 shows high precision across all classes, especially for "bottle" and "sofa". Its recall values are also higher, indicating that it effectively detects most instances of the objects present.

### Confusion Matrix Analysis

The normalized confusion matrices for YOLOv8 and Faster R-CNN, shown in Figure 1 and Figure 2 respectively, highlight the performance of both models across various classes. YOLOv8 achieves high precision, particularly for classes like "bottle" and "sofa", reflecting its ability to minimize false positives. The recall values for YOLOv8 are also generally high, indicating successful detection of most instances of each class.

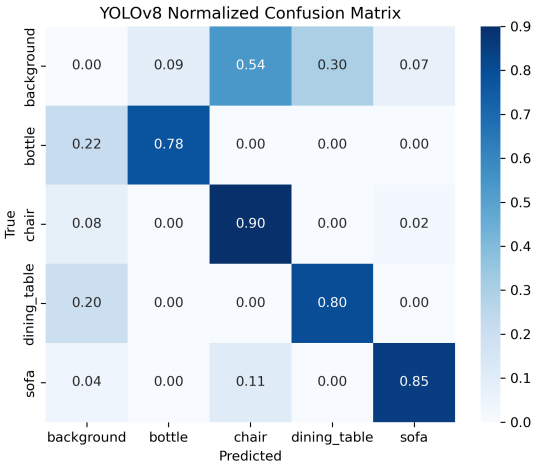


Figure 1: YOLOv8 norm. confusion matrix.

On the other hand, Faster R-CNN also shows strong performance in the "bottle" and "sofa" classes but has higher misclassification rates for other classes such as "chair" and "dining_table". This suggests that while Faster R-CNN can effectively detect certain objects, it struggles more with others, leading to more false positives and negatives compared to YOLOv8.

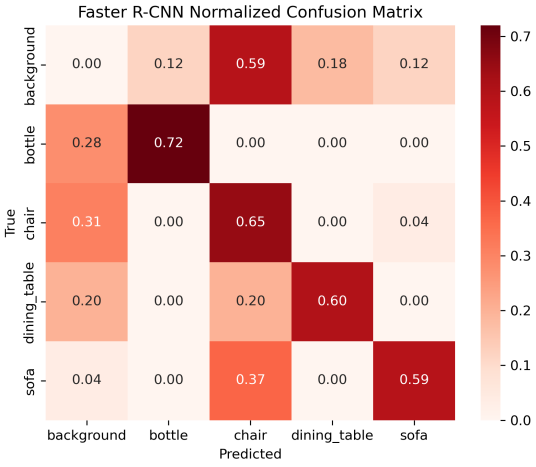


Figure 2: Faster R-CNN norm. confusion matrix.

### 3.2 Semantic Segmentation

In addition to object detection, we fine-tuned the YOLOv8 model on our dataset to perform segmentation. The evaluation metrics shown in Table 4 indicate that YOLOv8-seg achieved an F1 score of 0.693 for all classes on the bounding boxes and an F1 score of 0.683 for segmentation masks. These results demonstrates that the model is not only capable of detecting objects accurately but also can identify the precise boundaries of objects within the images.

| Metric | Box | Mask |
|---|---|---|
| Precision | 0.681 | 0.670 |
| Recall | 0.706 | 0.696 |
| F1 (calculated) | 0.693 | 0.683 |

Table 4: YOLOv8-seg performance metrics.

Table 5 shows the results across different classes, indicating that the YOLOv8 model segmented the "bottle" and "sofa" accurately, while achieving comparatively lower performance on the "chair" class.

| Class | Precision | Recall | F1 |
|---|---|---|---|
| Bottle | 0.782 | 0.797 | 0.789 |
| Chair | 0.488 | 0.582 | 0.530 |
| Dining Table | 0.692 | 0.632 | 0.661 |
| Sofa | 0.719 | 0.771 | 0.744 |

Table 5: YOLOv8-seg class-wise metrics.

Figure 3 shows the results on an unseen data example, where a picture of a living room was used to evaluate the performance. In this example, we can see that YOLOv8-seg accurately detected and segmented multiple objects within the image, including dining tables, sofas, and bottles, as well as classified them. This demonstrates the model's robustness and accuracy in both detection and segmentation tasks.

## 4 Conclusions

In conclusion, our study highlights the clear advantages of the YOLOv8 model over Faster R-CNN in object detection tasks. By consistently achieving higher precision, recall, and F1 scores across different object classes, YOLOv8 demonstrates its superiority in accurately identifying objects within images. Moreover, we have also confirmed the robustness of YOLOv8 in handling segmentation tasks, further underscoring its versatility and effectiveness. Overall, our exploration of this state-of-the-art model suggests its potential for advancing research and development in fields such as autonomous driving, or medical imaging.

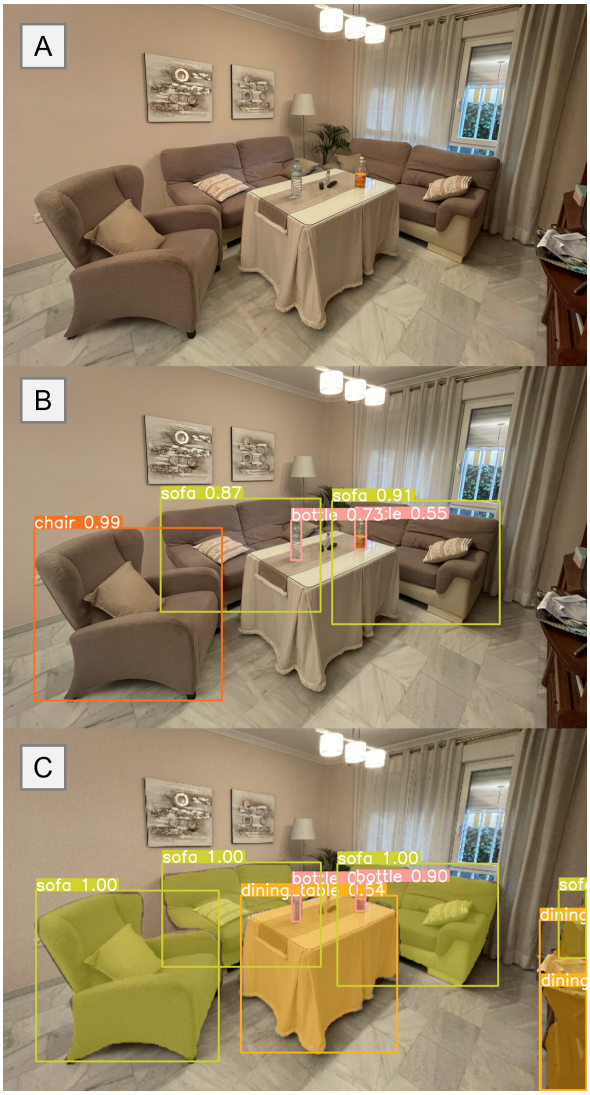


Figure 3: Example on unseen data. (A) Homemade photo. (B) YOLOv8 results for object detection. (C) YOLOv8-seg results for segmentation.

## References


[1] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," *arXiv preprint arXiv:1506.01497*, January 2015. Extended tech report.

[2] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 779–788, 2016.

[3] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2012 (VOC2012) Results."