

Comparison of YOLO algorithms

Enhar Örmeci

INTRODUCTION

This report presents the results of an evaluation of the performance of YOLOv3 , YOLOv5 ,YOLOv6 , YOLOv7 and YOLOv8 object detection models. The models were compared in terms of their ability to detect objects in different images.

DATASET

- Used Dataset = Traffic Signs
- Classes = Prohibitory , danger , mandatory , other
- Training Data Count : 675
- Test Data Count : 65

Performance Metrics

Below are the key metrics used to evaluate the performance of the YOLO models ;

1. Precision:

- YOLOv3: 0.98
- YOLOv5: 0.65
- YOLOv6: 0.22
- YOLOv7: 0.64
- YOLOv8: 0.88

2. Recall:

- YOLOv3: 0.95
- YOLOv5: 0.19
- YOLOv6: 0.43
- YOLOv7: 0.08
- YOLOv8: 0.77

3. F1 Score:

- YOLOv3: 0.96
- YOLOv5: 0.29
- YOLOv6: 0.29
- YOLOv7: 0.14
- YOLOv8: 0.82

4. Average Precision (AP):

- YOLOv3: 0.74
- YOLOv5: 0.11
- YOLOv6: 0.24
- YOLOv7: 0.04
- YOLOv8: 0.63

5. Processing Speed (Hours):

- YOLOv3: 0.40
- YOLOv5: 0.33
- YOLOv6: 0.36
- YOLOv7: 0.61
- YOLOv8: 0.46

YOLOv3 Results







YOLOv6 Results







YOLOv8 Results







Results

Precision:

Precision measures the accuracy of positive predictions made by the model. It's the ratio of true positive predictions to all positive predictions made.

- YOLOv3: Precision of 0.98 indicates that it has a high accuracy in identifying and correctly predicting positive cases.
- YOLOv5: Precision of 0.65 shows that it's somewhat accurate in positive predictions, but not as high as YOLOv3.
- YOLOv6: Precision of 0.22 suggests that it has a lower accuracy in positive predictions compared to the previous versions.
- YOLOv7: Precision of 0.64 indicates moderate accuracy in positive predictions.
- YOLOv8: Precision of 0.88 shows a relatively high accuracy in positive predictions.

Recall:

Recall measures the ability of the model to correctly identify all actual positive cases.

- YOLOv3: Recall of 0.95 indicates that it's good at identifying most of the actual positive cases.
- YOLOv5: Recall of 0.19 suggests that it's not very effective in capturing a significant portion of positive cases.
- YOLOv6: Recall of 0.43 indicates a moderate ability to identify actual positives.
- YOLOv7: Recall of 0.08 shows that it struggles to identify most of the actual positive cases.
- YOLOv8: Recall of 0.77 indicates a relatively good ability to capture actual positives.

F1 Score:

F1 Score is the harmonic mean of precision and recall, providing a balanced measure of a model's performance.

- YOLOv3: F1 Score of 0.96 indicates a good balance between precision and recall.
- YOLOv5: F1 Score of 0.29 suggests a trade-off between precision and recall, likely due to the low recall value.
- YOLOv6: F1 Score of 0.29 also shows a similar trade-off between precision and recall.
- YOLOv7: F1 Score of 0.14 indicates a significant imbalance, likely due to both low precision and low recall.
- YOLOv8: F1 Score of 0.82 suggests a good balance between precision and recall.

Average Precision (AP):

Average Precision measures the area under the precision-recall curve, indicating the overall performance of the model across different levels of thresholding.

- YOLOv3: AP of 0.74 indicates relatively good performance across different thresholds.
- YOLOv5: AP of 0.11 suggests poor performance across thresholds.
- YOLOv6: AP of 0.24 indicates a moderate performance across thresholds.
- YOLOv7: AP of 0.04 shows poor performance across thresholds.
- YOLOv8: AP of 0.63 indicates relatively good performance across thresholds.

Processing Speed (Hours):

Processing speed measures the time it takes for each model to perform inference on the data.

- YOLOv3: Processing speed of 0.40 hours suggests it's relatively fast.
- YOLOv5: Processing speed of 0.33 hours indicates a similar level of speed to YOLOv3.

- YOLOv6: Processing speed of 0.36 hours also suggests a similar level of speed to YOLOv3.
- YOLOv7: Processing speed of 0.61 hours indicates a slower performance.
- YOLOv8: Processing speed of 0.46 hours suggests a moderate processing speed.

In summary, YOLOv3 generally stands out as having a good balance between precision, recall, F1 score, and average precision. YOLOv8 also performs well across these metrics. YOLOv5 and YOLOv6 seem to have lower performance, particularly in recall and average precision. YOLOv7 appears to be the weakest performer among the options provided. Keep in mind that these analyses are based solely on the given values and do not take into account other potential factors affecting model selection.