Chapter 4

Results & Discussion

Performance Analysis of Object Detection Models

for Visually Impaired Assistance

Evaluation Date: November 29, 2025

Dataset: COCO val2017 (5000 images)

Models Evaluated: 14

4.1 Evaluation Overview

This chapter presents the comprehensive evaluation results of 14 object detection models evaluated on the COCO val2017 dataset subset. The models evaluated include variants from the YOLO family (YOLOv8 and YOLOv10), transformer-based RT-DETR, and traditional architectures (SSD, RetinaNet, Faster R-CNN). Each model was evaluated using standardized metrics including mean Average Precision (mAP), recall, precision, and inference speed (FPS).

The evaluation was conducted with a specific focus on assistive technology applications for visually impaired persons, where high recall (detecting all relevant objects) and real-time performance are critical requirements. Models must balance accuracy with speed to be deployable on mobile devices and edge computing platforms.

4.1.1 Models Evaluated

Family	Variants	Source	Architecture
YOLOv8	n, s, m, l, x	Ultralytics	Single-stage, anchor-free
YOLOv10	n, s, m, l, x	Ultralytics	NMS-free, efficient
RT-DETR	l	Baidu	Transformer-based
SSD	300	Liu et al.	Single-stage, multi-scale
RetinaNet	base	Facebook AI	Focal loss, FPN
Faster R-CNN	base	Microsoft	Two-stage, RPN

4.2 Main Results

Table 4.1 presents the comprehensive evaluation results for all 14 models, sorted by mAP@0.5:0.95 in descending order. The results demonstrate clear trade-offs between model complexity, accuracy, and inference speed.

Table 4.1: Comprehensive Model Comparison

Model	mAP@0.5	mAP@0.5:0.95	Recall	Precision	FPS	Latency(ms)
YOLOv10x	70.7	54.3	63.6	74.8	159.6	6.3
YOLOv8x	70.4	53.9	64.5	73.7	139.3	7.2
YOLOv10l	69.4	52.9	63.7	72.7	209.1	4.8
YOLOv8l	69.2	52.8	63.1	74.0	190.1	5.3
RT-DETR-l	69.9	51.9	65.2	74.2	147.7	6.8
YOLOv10m	67.4	50.8	60.2	73.3	270.6	3.7
YOLOv8m	66.5	50.0	60.7	71.6	253.0	4.0
Faster R-CNN	67.5	46.7	N/A	N/A	48.4	20.6
YOLOv10s	62.3	46.0	56.1	70.0	384.5	2.6
YOLOv8s	61.1	44.6	56.0	68.2	352.2	2.8
RetinaNet	61.7	41.5	58.8	N/A	55.0	18.2
YOLOv10n	53.1	38.2	48.7	64.1	462.0	2.2
YOLOv8n	51.9	37.0	47.3	63.2	410.9	2.4
SSD300	41.3	25.0	35.0	N/A	82.8	12.1

4.2.1 Key Findings

Best Accuracy (mAP@0.5:0.95): YOLOv10x achieved the highest accuracy with 54.3%, demonstrating excellent detection capability across various IoU thresholds.

Best Recall: RT-DETR-l achieved the highest recall at 65.2%, which is critical for assistive technology applications where missing detections could impact user safety.

Fastest Model: YOLOv10n achieved 462.0 FPS, making it highly suitable for real-time mobile deployment while maintaining reasonable detection accuracy.

4.3 YOLO Family Analysis

The YOLO (You Only Look Once) family represents state-of-the-art single-stage object detectors. We evaluated five variants each of YOLOv8 and YOLOv10, ranging from nano (n) to extra-large (x), to understand the accuracy-speed trade-offs.

4.3.1 YOLOv8 Results

Model	mAP@0.5	mAP@0.5:0.95	Recall (%)	FPS
YOLOv8n	51.9	37.0	47.3	410.9
YOLOv8s	61.1	44.6	56.0	352.2
YOLOv8m	66.5	50.0	60.7	253.0
YOLOv8l	69.2	52.8	63.1	190.1
YOLOv8x	70.4	53.9	64.5	139.3

Figure 4.1 shows the scaling behavior of YOLOv8 variants. As model size increases from nano to extra-large, accuracy improves significantly (37.0% to 53.9% mAP@0.5:0.95) while speed decreases (411 to 139 FPS).

4.3.2 YOLOv10 Results

Model	mAP@0.5	mAP@0.5:0.95	Recall (%)	FPS
YOLOv10n	53.1	38.2	48.7	462.0
YOLOv10s	62.3	46.0	56.1	384.5
YOLOv10m	67.4	50.8	60.2	270.6
YOLOv10l	69.4	52.9	63.7	209.1
YOLOv10x	70.7	54.3	63.6	159.6

YOLOv10 introduces NMS-free detection, resulting in faster inference compared to YOLOv8 at equivalent model sizes. YOLOv10x achieves the best overall accuracy (54.3% mAP@0.5:0.95) among all YOLO variants.

4.3.3 YOLOv8 vs YOLOv10 Comparison

Direct comparison between YOLOv8 and YOLOv10 variants of the same size reveals interesting trade-offs. Table 4.4 presents a side-by-side comparison.

Size	v8 mAP	v10 mAP	v8 Recall	v10 Recall	v8 FPS	v10 FPS
N	37.0	38.2	47.3	48.7	411	462
S	44.6	46.0	56.0	56.1	352	385
M	50.0	50.8	60.7	60.2	253	271
L	52.8	52.9	63.1	63.7	190	209
X	53.9	54.3	64.5	63.6	139	160

Key observations:

- YOLOv10 consistently achieves higher FPS due to NMS-free design
- YOLOv10 shows marginally better mAP at most model sizes
- Both families show similar recall patterns, with larger models achieving ~64% recall
- For real-time mobile applications, YOLOv10n offers the best speed-accuracy trade-off

4.4 Alternative Architectures

4.4.1 RT-DETR (Transformer-based)

RT-DETR-I achieves impressive results with 51.9% mAP@0.5:0.95 and notably the highest recall (65.2%) among all evaluated models. At 147.7 FPS, it offers real-time performance suitable for assistive applications where high recall is critical.

4.4.2 SSD300

SSD300 achieves 25.0% mAP@0.5:0.95 with 82.8 FPS. While faster than two-stage detectors, its accuracy is significantly lower than modern YOLO variants, making it less suitable for safety-critical applications.

4.4.3 RetinaNet

RetinaNet with focal loss achieves 41.5% mAP@0.5:0.95 and 58.8% recall. Its 55.0 FPS is suitable for near-real-time applications but falls short for mobile deployment.

4.4.4 Faster R-CNN

Faster R-CNN, as a two-stage detector, achieves 46.7% mAP@0.5:0.95 with 48.4 FPS. While offering competitive accuracy, its slower inference speed limits real-time mobile deployment potential.

4.5 Speed and Latency Analysis

For assistive technology applications, inference speed is crucial. The system must process frames in real-time (>30 FPS) to provide timely feedback to users. Table 4.5 presents the speed analysis with mobile feasibility assessment.

Model	FPS	Latency (ms)	Real-time*	Mobile**
YOLOv10n	462.0	2.2	Yes	Yes
YOLOv8n	410.9	2.4	Yes	Yes
YOLOv10s	384.5	2.6	Yes	Yes
YOLOv8s	352.2	2.8	Yes	Yes
YOLOv10m	270.6	3.7	Yes	Yes
YOLOv8m	253.0	4.0	Yes	Yes
YOLOv10l	209.1	4.8	Yes	Yes
YOLOv8l	190.1	5.3	Yes	Yes
YOLOv10x	159.6	6.3	Yes	Yes
RT-DETR-l	147.7	6.8	Yes	Yes
YOLOv8x	139.3	7.2	Yes	Yes
SSD300	82.8	12.1	Yes	Maybe
RetinaNet	55.0	18.2	Yes	Maybe
Faster R-CNN	48.4	20.6	Yes	No

* Real-time: >=30 FPS; ** Mobile Feasible: Estimated based on GPU-to-mobile scaling factors

All YOLO variants and RT-DETR achieve real-time performance on the evaluation hardware (NVIDIA RTX 4090). For mobile deployment, nano and small variants are recommended, with expected 20-35 FPS on high-end mobile devices after optimization.

4.6 Discussion

4.6.1 Accuracy vs Speed Trade-off

The evaluation reveals a clear accuracy-speed trade-off across model families. The Pareto frontier analysis shows that YOLOv10 variants dominate the efficiency curve, offering the best accuracy at any given speed constraint. For applications requiring >100 FPS, YOLOv10n provides 38.2% mAP with 462 FPS. For maximum accuracy without strict speed constraints, YOLOv10x achieves 54.3% mAP at 160 FPS.

4.6.2 Recall Considerations for Assistive Technology

For visually impaired assistance, recall is arguably more important than precision - failing to detect an obstacle (false negative) is more dangerous than a false alarm. RT-DETR-l achieves the highest recall (65.2%), followed by YOLOv8x (64.5%) and YOLOv10l (63.7%). The nano variants achieve approximately 48% recall, which may be insufficient for safety-critical applications.

4.6.3 Model Recommendations

Based on the evaluation results, we recommend the following models for different deployment scenarios:

- High-end Mobile Devices: YOLOv10s (46.0% mAP, ~25 FPS expected)
- Edge Devices (Jetson): YOLOv10m (50.8% mAP, ~20 FPS expected)
- Maximum Accuracy: YOLOv10x or RT-DETR-l (>51% mAP)
- Maximum Speed: YOLOv10n (38.2% mAP, highest FPS)

4.6.4 Limitations

Several limitations should be considered:

(1) Evaluation was performed on COCO dataset, which may not fully represent real-world assistive technology scenarios.
(2) Speed measurements were taken on a high-end GPU; actual mobile performance will be significantly lower.
(3) The evaluation focuses on general object detection; domain-specific fine-tuning may improve results for assistive applications.
(4) Small object detection remains challenging across all models.

4.7 Summary

This chapter presented a comprehensive evaluation of 14 object detection models for potential use in assistive technology for visually impaired persons. The key findings can be summarized as follows:

(1) YOLOv10 family offers the best accuracy-speed trade-off, with YOLOv10x achieving the highest mAP (54.3%) and YOLOv10n the highest speed (462 FPS).

(2) RT-DETR-l achieves the highest recall (65.2%), making it suitable for safety-critical applications where missing detections must be minimized.

(3) For mobile deployment, YOLOv10s or YOLOv10m offer the best balance between accuracy (46-51% mAP), speed, and model size.

(4) Traditional architectures (SSD, Faster R-CNN) are outperformed by modern YOLO variants in both accuracy and speed.

(5) All top-performing models achieve real-time performance (>30 FPS) on GPU hardware, with expected viable performance on modern mobile devices after optimization.