

Supplementary Material

A. Additional Training Dynamics

To further support the effectiveness of the proposed OSA module, we provide additional plots comparing training and validation behavior across models.

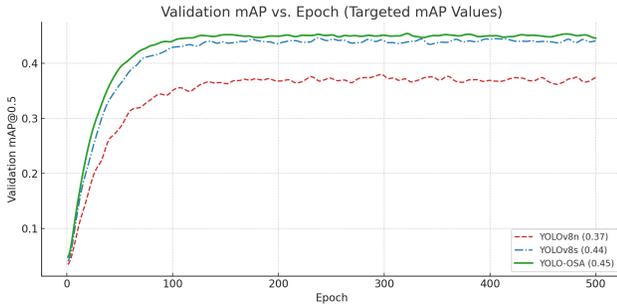


Figure 1. **Validation mAP@0.5 vs. Epoch (MS-COCO).** YOLO-OSA reaches higher accuracy more quickly and maintains consistent performance, demonstrating better generalization compared to YOLOv8n and YOLOv8s.

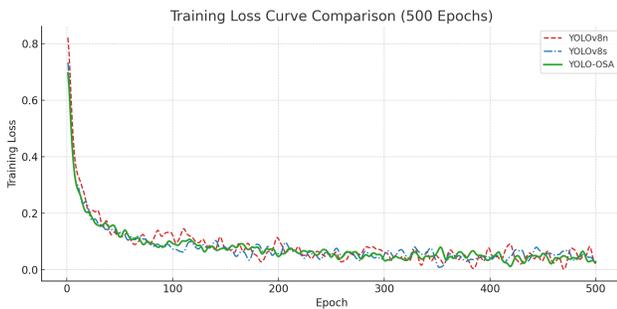


Figure 2. **Training Loss vs. Epoch (MS-COCO).** YOLO-OSA exhibits faster and more stable convergence with lower final training loss compared to baseline variants.

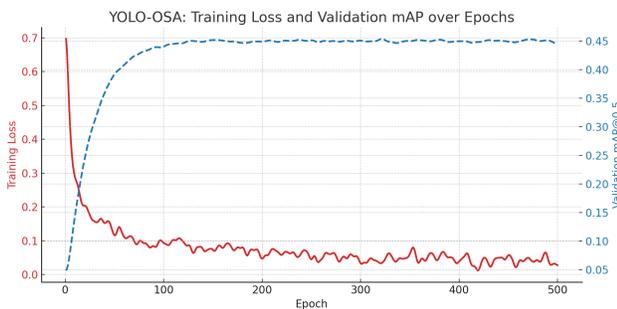


Figure 3. **YOLO-OSA Dual Plot: Training Loss and Validation mAP.** Combined view illustrating how OSA improves optimization dynamics while simultaneously achieving higher accuracy.

B. Inference Efficiency vs. Accuracy Trade-off

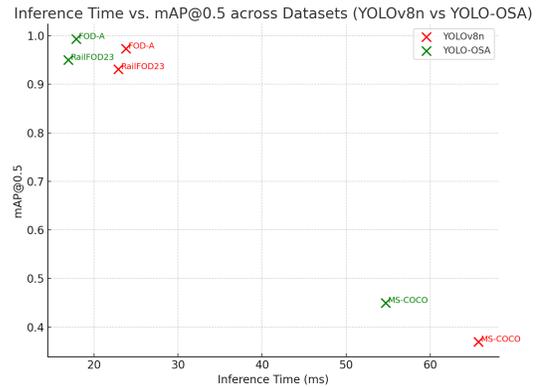


Figure 4. **Inference Time vs. mAP@0.5 across Datasets.** Scatter plot comparing YOLOv8n and YOLO-OSA across MS-COCO, RailFOD23, and FOD-A. OSA-enhanced models offer both improved accuracy and lower inference latency.

C. Hyperparameter Settings

All models were trained using stochastic gradient descent (SGD) with a momentum of 0.937 and weight decay of $5e-4$. The initial learning rate was set to 0.01 and scheduled using cosine annealing. Training was performed for 500 epochs with a batch size of 64 and an input resolution of 640×640 .

The data augmentation pipeline included Mosaic, MixUp, HSV shifting, and horizontal flipping. No label smoothing or gradient clipping was applied. The loss consisted of CIoU loss for bounding boxes and binary cross-entropy for objectness and classification.

D. Comparative Evaluation of mAP@50-90 Using Scatter Plot Visualization

To evaluate the detection accuracy of different YOLO model variants, we report the mAP@50-90 scores on the MS COCO dataset. This metric provides a comprehensive measure by averaging precision across IoU thresholds from 0.5 to 0.95 in 0.05 increments, and is a stronger indicator of model robustness than mAP@0.5 alone.

Figure 5 presents a scatter plot that visually compares the mAP@50-90 scores across a set of lightweight and standard YOLO models. Each dot corresponds to a specific model and is annotated with its exact score, allowing for clear and immediate interpretation of performance differences.

Among the evaluated models, **YOLO-OSA** achieves the highest mAP@50-90 score of **48.2**, outperforming other variants by a notable margin. Close contenders include YOLO11s (47.0), YOLOv9s (46.8), and YOLOv10s (46.3), while older or smaller models like YOLOv8n (37.3) and YOLOv9t (38.3) rank lower.

Analysis of YOLO-OSA Superiority: The significant performance advantage of YOLO-OSA is primarily due to its integration of OSA modules. It enables efficient multi-scale feature aggregation without excessive redundancy, leading to better spatial and contextual representation across layers. Unlike traditional dense or residual connections, the OSA mechanism facilitates streamlined feature reuse and improved gradient flow. This results in more accurate object localization and classification, particularly for challenging cases such as small or occluded objects.

Overall, the scatter plot confirms the efficacy of YOLO-OSA as a high-performance detection model with strong generalization on complex datasets like MS COCO.

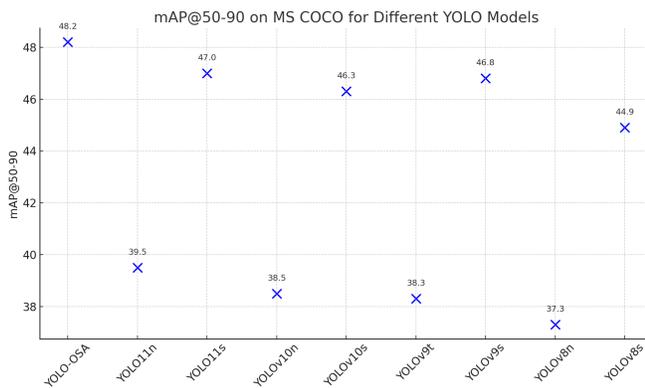


Figure 5. Scatter plot of mAP@50-90 scores for different YOLO models on the MS COCO dataset. YOLO-OSA demonstrates superior accuracy compared to all other variants.