

Comparative Analysis of Seven YOLO Architectures Applied for the Detection of Objects of Different Scale at Different Scene Densities

Kazimierz Choroś and Farid Kazimov | Wrocław University of Science and Technology | ACIIDS 2026

Why this study?

Problem: lightweight YOLO models are widely used, but their behavior changes with object scale and crowding.

Gap: most comparisons use one dataset or focus only on numerical metrics.

Contribution: a unified benchmark of seven nano/tiny YOLO variants across three challenging regimes with both quantitative and qualitative analysis.

Experimental setup

- Seven models: YOLOv5n, YOLOv6n, YOLOv8n, YOLOv9t, YOLOv10n, YOLO11n, YOLO12n
- Unified Ultralytics pipeline; all models trained from scratch
- Same settings for all experiments: 50 epochs, batch 16, learning rate 0.01, momentum 0.937, weight decay 0.0005, image size 640×640
- Evaluation: precision, recall, mAP@0.5, mAP@0.5:0.95, plus visual stress-test inspection

Datasets

Bee

small objects
1,500 images

Sheep

medium scale
1,200 images

Human Head

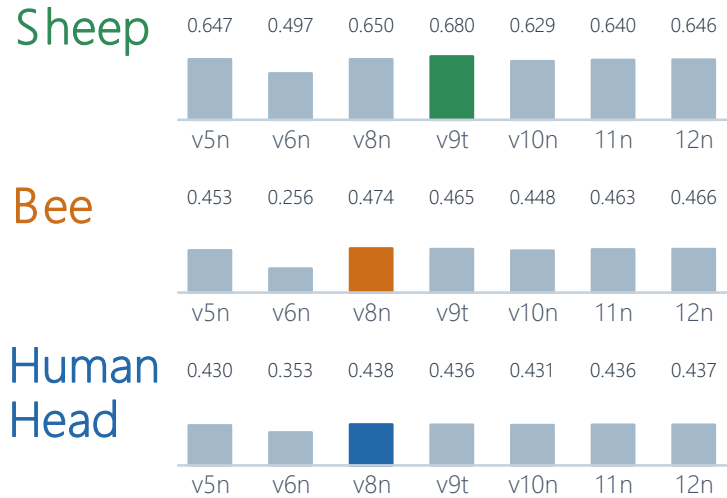
dense / occluded
1,500 images

Benchmark pipeline



Goal: isolate architectural effects and connect numerical scores with real visual behavior

Quantitative results (mAP@0.5:0.95)



Best per regime: YOLOv9t for medium-scale sheep, YOLOv8n for bees and dense heads (with very small margins in the dense case)

Qualitative insights

Bee scene:

YOLOv5n showed very high sensitivity; YOLOv8n looked more balanced; YOLOv10n was too conservative.

Sheep scene:

YOLOv6n looked best visually in one multi-scale case, despite ranking last numerically.

Dense crowd scene:

YOLO11n and YOLO12n detected more visible heads, suggesting stronger recall in crowded scenes.

Interpretation

The numerically best model is not always the most useful one in practice.

Aggregate metrics can hide false positives, missed detections, and behavior changes across scale and occlusion.

mAP alone is insufficient for practical model selection.

Takeaways

- There is no universally best lightweight YOLO model.
- Choose the detector according to target scene characteristics.