

Object Detection using YOLO

Prepared by: Subash Sah

What is Object Detection?

- ▶ A computer vision task to identify and locate objects in images or videos.
- ▶ Outputs:
 - ▶ Bounding boxes around objects
 - ▶ Class labels for each detected object
 - ▶ Confidence scores (probability of correct detection)
- ▶ Applications: Surveillance, autonomous vehicles, image retrieval, robotics.

Traditional Object Detection Models

- ▶ R-CNN (Regions with CNN features)
 - ▶ Extracts region proposals and classifies each with CNN.
 - ▶ Pros: High accuracy
 - ▶ Cons: Slow due to separate region proposal and CNN classification.
- ▶ Fast R-CNN
 - ▶ Processes the whole image once and pools features for regions.
 - ▶ Faster than R-CNN.
- ▶ Faster R-CNN
 - ▶ Introduces Region Proposal Network (RPN) for real-time region proposals.
 - ▶ Balances speed and accuracy

YOLO (You Only Look Once)

- ▶ Object detection as a single regression problem:
 - ▶ Input image - predicts bounding boxes + class probabilities in one pass
- ▶ Key features:
 - ▶ Real-time detection
 - ▶ Lightweight and fast
 - ▶ End-to-end training
- ▶ Used in applications requiring speed and accuracy.

YOLO Architecture Overview

- ▶ Backbone: Extracts features from the image.
- ▶ Neck: Combines multi-scale features.
- ▶ Head: Predicts bounding boxes, classes, and confidence scores.
- ▶ Works on grid-based prediction to detect multiple objects simultaneously.

Advantages of YOLO

- ▶ High speed: Real-time detection possible.
- ▶ Single-stage detection: No separate region proposal step.
- ▶ End-to-end training for bounding boxes and classes.
- ▶ Scalable to small and large models: YOLOv8n (nano), YOLOv8s (small), YOLOv8m (medium), YOLOv8l (large).

Conclusion

- ▶ Object detection identifies and localizes objects in images/videos.
- ▶ Traditional methods (R-CNN, Fast/Faster R-CNN) are accurate but slower.
- ▶ YOLO provides a fast, end-to-end solution for real-time detection.