VIT®
Vellore Institute of Technology
(Deemed to be University under section 3 of UGC Act, 1956)
CHENNAI

You Only Look Once (YOLO)

Dr. David Raj M

Assistant Professor
Department of Mathematics
School of Advanced Sciences
Vellore Institute of Technology, Chennai

Feb 13, 2026

This lecture is delivered at the FDP in Dayananda Sagar Engineering College.

Contents

You Only Look Once (YOLO)

The Object Detection Challenge

THE PROBLEM WE FACE:

- Find + Classify + Localize ALL objects in image
- Real-time speed required (30+ FPS)
- Multiple objects, varying sizes, overlapping

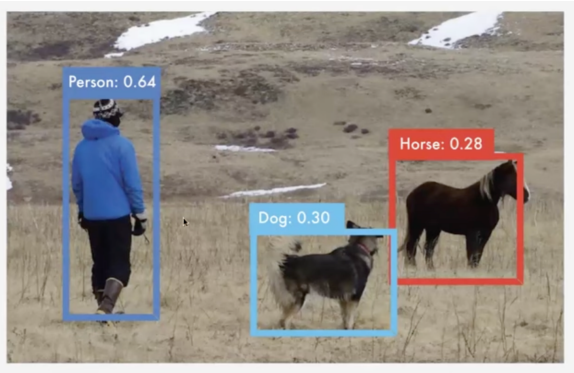
Traditional Pipeline (R-CNN family):

1. Region Proposals $\rightarrow$ 2000+ boxes
2. CNN Feature Extraction $\rightarrow$ Per box
3. Classification $\rightarrow$ Per box
4. Bounding Box Regression $\rightarrow$ Per box

2-5 SECONDS per image!

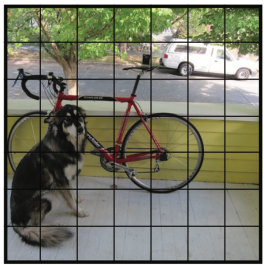
Object Detection

Object detection is a task that involves identifying the location and class of objects in an image or video stream.



- The output of an object detector is a set of bounding boxes that enclose the objects in the image, along with class labels and confidence scores for each box.
- Object detection is a good choice when you need to identify objects of interest in a scene, but don't need to know exactly where the object is or its exact shape.

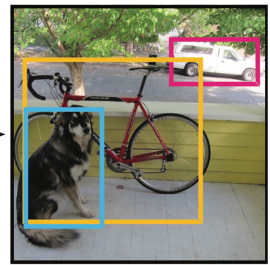
Real-Time Detectors	Train	mAP	FPS
100Hz DPM [31]	2007	16.0	100
30Hz DPM [31]	2007	26.1	30
Fast YOLO	2007+2012	52.7	**155**
YOLO	2007+2012	**63.4**	45
Less Than Real-Time			
Fastest DPM [38]	2007	30.4	15
R-CNN Minus R [20]	2007	53.5	6
Fast R-CNN [14]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[28]	2007+2012	73.2	7
Faster R-CNN ZF [28]	2007+2012	62.1	18
YOLO VGG-16	2007+2012	66.4	21



S × S grid on input

Bounding boxes + confidence

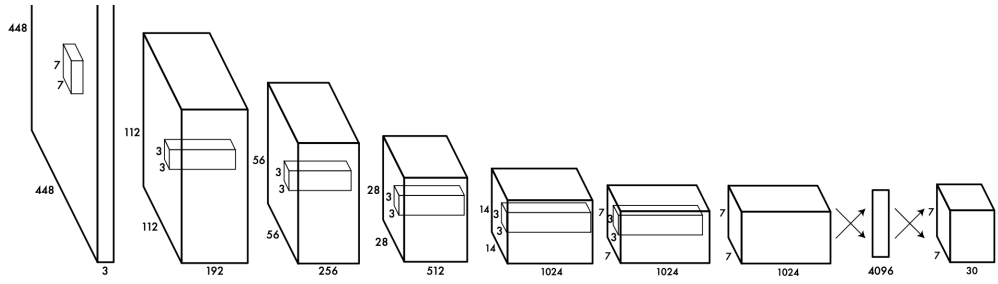
Class probability map



Final detections

YOLO v1 Architecture

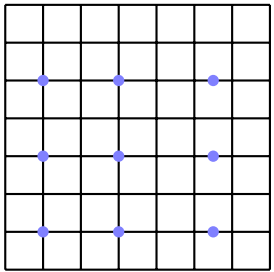
24 Conv Layers + 2 FC Layers



Input Image (448×448)

↓

7×7 Grid



Each cell predicts 2 boxes + 20 classes

↓

NMS (Non-Max Suppression)

↓

Final Detections with Boxes!

YOLO vs Traditional Methods

Metric	**R-CNN**	**Fast R-CNN**	**YOLO v1**
mAP (Accuracy)	66.0%	70.0%	63.4%
Speed (FPS)	0.5-15	17	**45**
Pipeline Steps	4 stages	3 stages	**1 stage**

YOLO TRADES slight accuracy for 3-10x SPEED!

Generalizations

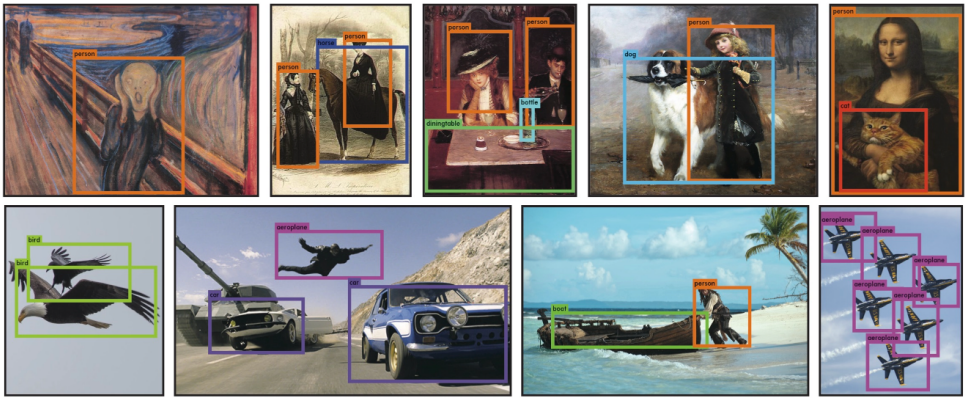


Figure 6: Qualitative Results. YOLO running on sample artwork and natural images from the internet. It is mostly accurate although it does think one person is an airplane.

Why YOLO Wins

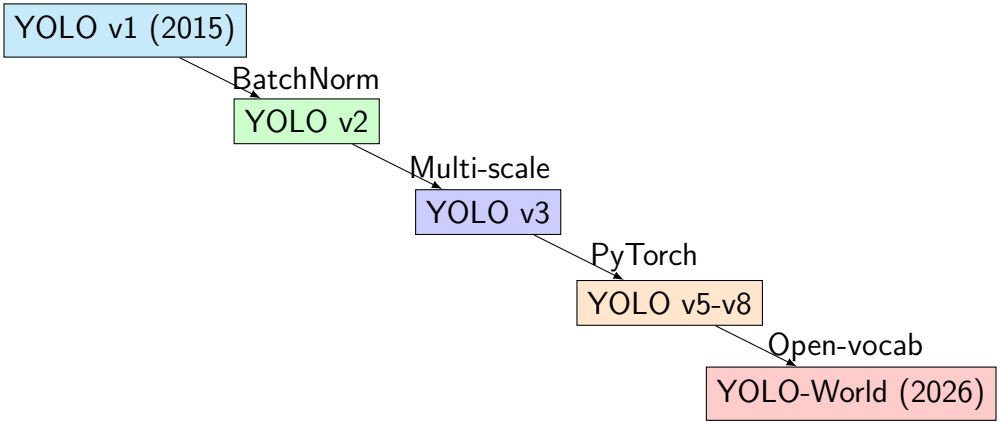
- **SINGLE** neural network (end-to-end)
- Real-time inference **(45+ FPS)**
- Global context (sees entire image)
- Simple to train/deploy
- Scales to video

Tradeoffs:

- Struggles with small objects
- Fewer bounding boxes per cell

PROVEN: Speed > Perfection for real-world!

YOLO Evolution Path



45 FPS → Production Ready

Ready for LIVE Demo?

Next: YOLOv8 in Action!

1. `ultralytics/yolov8n.pt` (2MB model)
2. Webcam $\rightarrow$ Real-time detection
3. Custom training on YOUR data

Questions before we code?

References I

[1] Christopher M Bishop. *Neural networks for pattern recognition*. Oxford university press, 1995.

[2] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[3] Joseph Redmon et al. "You Only Look Once: Unified, Real-Time Object Detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2016, pp. 779–788. URL: http://pjreddie.com/media/files/papers/YOLO_1.pdf.