

Object detection is a crucial task in computer vision that involves identifying and localizing objects within an image. Over the years, various architectures have been developed, improving accuracy, speed, and robustness. The object detection family can be categorized into two primary types:

1. Two-Stage Object Detectors (Region Proposal-Based)
2. One-Stage Object Detectors (Regression-Based)
3. Transformer-Based Object Detectors (Recent Advances)

Classification + Localization

2 Stage Detectors

1) RCNN 2014

Region Proposal Network (Selective Search)

2) FAST RCNN

2015 Girshick

ROI Pooling

(Selective Search)

3) Faster RCNN

2015

RPN → Region Proposal Network

Not Real Time

4) MASK RCNN

2017

Classification + Localization + Segmentation

One Stage Detectors

1) YOLO 2016 Redmon

V1 — V12

2) SSD (Single Shot Multibox) 2016

Cons: less accurate

3) RetinaNet 2017

FPN (Feature Pyramid Network)

High Accuracy with Real Time Speeds

Computationally Expensive

Transformer Based Detectors

1) DETR (Detection Transformer) 2020

Cons: Computationally Very Expensive

Eliminates (NMS)

Advantages :- E2E OD without Anchors

2) Deformable DETR 2021

Deformable Attention to focus only on relevant objects

3) DINO (DETR with improved Noisy optimization)

Improves training convergence and detection accuracy.

Comparison of Object Detection Architectures				
Architecture	Type	Speed	Accuracy	Notable Features
R-CNN	Two-Stage	Slow	High	Uses Selective Search
Fast R-CNN	Two-Stage	Faster	High	RoI Pooling
Faster R-CNN	Two-Stage	Faster	Very High	RPN for region proposals
Mask R-CNN	Two-Stage	Slower	Very High	Adds segmentation masks
YOLOv3	One-Stage	Very Fast	High	Multi-scale detection
YOLOv5	One-Stage	Real-time	High	Lightweight, efficient
SSD	One-Stage	Fast	Medium	Multi-scale feature maps
RetinaNet	One-Stage	Moderate	Very High	Focal Loss for imbalance
DETR	Transformer	Slow	Very High	End-to-end detection
Deformable DETR	Transformer	Faster	Very High	Focused attention mechanism