Problems With RCNN

- Computational cost: R-CNN is computationally expensive, with a slow inference time due to its two-stage architecture.

- Poor recall: R-CNN has poor recall performance, meaning it can miss object detections, particularly for smaller or occluded objects.

- Inefficient proposal generation: R-CNN uses selective search to generate object proposals, which is slow and does not scale well with the size of the input image.

- Limited versatility: R-CNN is designed to handle a limited set of object categories and may not perform well on new or unseen categories.

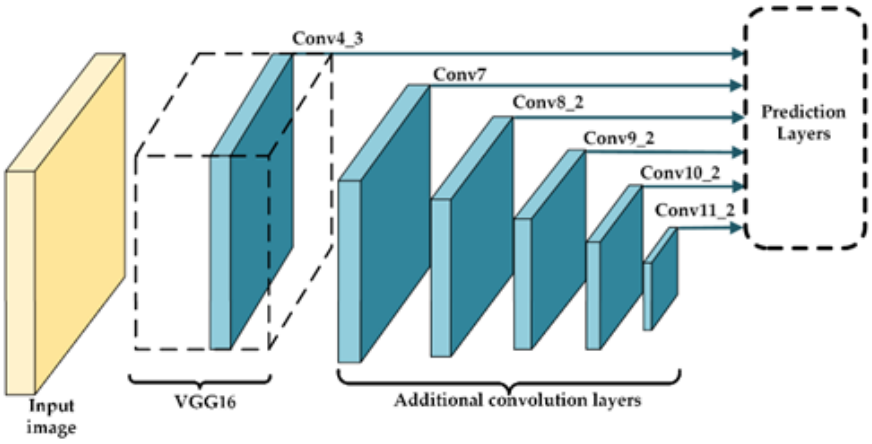
- Difficult to fine-tune: R-CNN can be difficult to fine-tune for specific tasks, particularly when the training data is limited or difficult to obtain.

- Limited scalability: Scaling up the size of an R-CNN model to handle a large number of object categories can be challenging, leading to slower inference speed and increased memory usage.

Single Shot Detector (SSD)

Single Shot MultiBox Detector (SSD) is a one-stage object detection method that predicts class labels and bounding boxes for objects within an image.

It's an efficient alternative to two-stage object detectors, such as R-CNN, and is designed to handle a large number of object categories in real time.



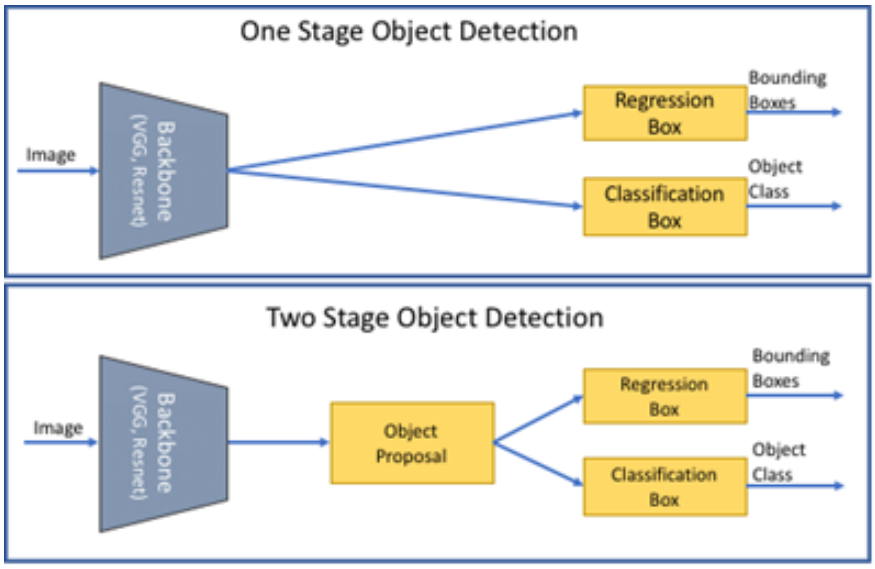
How SSD work?

- Single Shot MultiBox Detector (SSD) is an object detection method that uses multiple feature maps from different layers of a convolutional neural network (CNN) with a predictive head to predict class labels and bounding boxes for objects within an image.

- It uses anchor boxes, or prior boxes, and a multi-task loss function to handle objects at different scales, shapes, and locations in a single forward pass.

- This allows for fast and efficient object detection.

Single Stage VS Two-Stage Methods

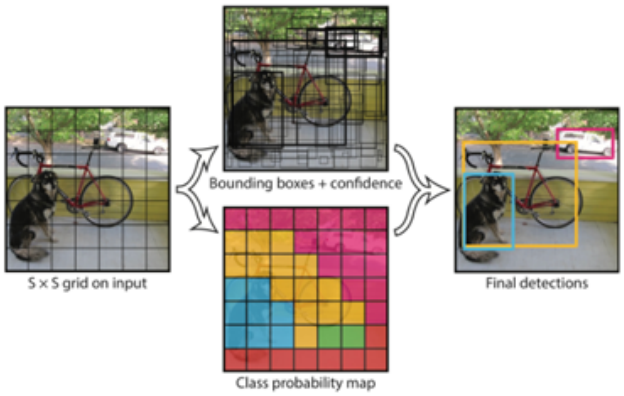


You Only Look Once (YOLO)

- You Only Look Once (YOLO) is a one-stage object detection method that predicts class labels and bounding boxes for objects within an image.

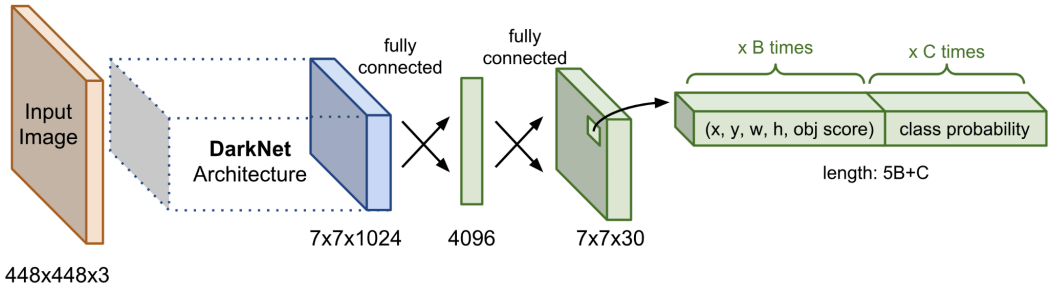
- It's designed to handle a large number of object categories in real time and is known for its fast inference time and efficient use of computational resources.

How YOLO work?



- The bounding box contains four values: x, y, w, h, (x, y) represents the center of the box. (W, h) defines the width and height of the box.
- Confidence indicates the probability of containing objects in this prediction box, which is the IoU value between the prediction box and the actual box.
- The class probability indicates the class probability of the object, and the YOLOv3 uses a two-class method

Simple YOLO Architecture



Drawbacks of YOLO

- Both one-stage detection methods, like SSD and YOLO, evaluate almost 10^4 to 10^5 candidate locations per image.
- But only a few locations contain objects (i.e. Foreground) and rest are just background objects.
- This leads to the class **imbalance problem**.
- Small objects and close-by objects may be missed by YOLO-like algorithms
- YOLO can only detect a limited number of objects per image, making it less suitable for applications where a large number of objects need to be detected.

RetinaNet

- RetinaNet is considered to be better than single-stage object detection methods for several reasons:
- **Better accuracy:** RetinaNet was designed to address the problem of class imbalance in object detection, where a large number of negative examples (i.e., background) can overwhelm the network and lead to poor performance.

- RetinaNet addresses this by using a two-stage approach that uses a class-specific confidence score to filter out false positive detections.

- **Faster inference:** Single-stage object detectors such as YOLO and SSD use anchor boxes to detect objects and can generate multiple detections per object.

- RetinaNet, on the other hand, generates only one detection per object, which results in faster inference times.