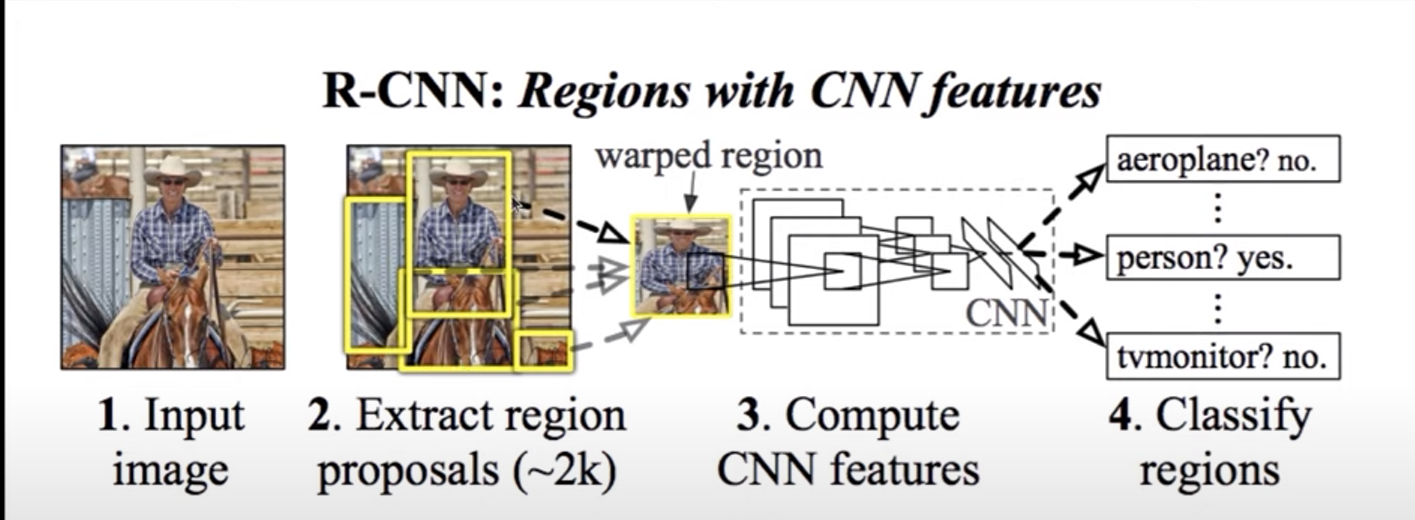
**OBJECT DETECTION**Object detection involves classification and localisation of an object which involves locating the object as well as identifying their class.There are many different algorithms build for detecting the object. It includes the sliding window algorithm. Like if we want to detect a dog, we have to consider a fixed window size , the dog will occupy most of the window. this is essentially a sub image that we would like to classify as a dog, the other sub image would be classified as background. Each image that does not contain the image of dog will be considered as a background class. The sliding window algorithm is a more systematic approach. We start from one region of image, classify that sub image and then shift to the next sub image. We repeat the process. When we reach the horizontal border , we move a few pixels down in the vertical direction and repeat the process . When object occupies most of the window it will be classified as a dog. But there are many problems . Object detectors often output overlapping detections. We also have the issue of object sizes where the same object can come in different sizes. One way is to reshape the image. Bounding box is another method. It is a rectangular box that can be determined with the lower right corner of the rectangle with coordinates y0 and x0 and the width and the height. It can also be determined by the upper left corner ymin and xmin and the lower right corner ymax and xmax. These are not the labels . Just like classification , we have the dataset of class x and y and the bounding box. Similar to classification, we use the dataset to train the model and include the box coordinates. The result is an object detector with updated learning parameters. We input the image with the objects and the detecObject detection involves classification and localisation of an object which involves locating the object as well as identifying their class.There are many different algorithms build for detecting the object. It includes the sliding window algorithm. Like if we want to detect a dog, we have to consider a fixed window size , the dog will occupy most of the window. this is essentially a sub image that we would like to classify as a dog, the other sub image would be classified as background. Each image that does not contain the image of dog will be considered as a background class. The sliding window algorithm is a more systematic approach. We start from one region of image, classify that sub image and then shift to the next sub image. We repeat the process. When we reach the horizontal border , we move a few pixels down in the vertical direction and repeat the process . When object occupies most of the window it will be classified as a dog. But there are many problems . Object detectors often output overlapping detections. We also have the issue of object sizes where the same object can come in different sizes. One way is to reshape the image. Bounding box is another method. It is a rectangular box that can be determined with the lower right corner of the rectangle with coordinates y0 and x0 and the width and the height. It can also be determined by the upper left corner ymin and xmin and the lower right corner ymax and xmax. These are not the labels . Just like classification , we have the dataset of class x and y and the bounding box. Similar to classification, we use the dataset to train the model and include the box coordinates. The result is an object detector with updated learning parameters. We input the image with the objects and the detector predicts the class and the box coordinates. Many object algorithms provide a score that kets us know how confident the detectors prediction is. It is in the range 0 to 1.tor predicts the class and the box coordinates. Many object algorithms provide a score that kets us know how confident the detectors prediction is. It is in the range 0 to 1.

****The approach that quickly took over was the Regional based network. In this we take an input image and we have an algorithm to take some region proposals and this wasn't a neural network. They used some deterministic algorithm that extracted potential bounding boxes for the image. They extract about 2000 region proposals. So if there was an object with a very high probability then they are going to extract that region and then they resize the image to a fixed shape of 224×224 and then it is send to cnn model and then get the classes for that image. So what it solved is we have a fixed no of 2000 iamges which we are sending in the xnn which is a lot less than the sliding windows.   
And we don't have to worry about the size of the bounding box due to the selective search algorithms. There are many other release like fact rcnn and then there is faster rcnn . Then they changed the region proposal to become a neural network. Potential problems with this approach is that it is still slow ( far away from real time object detection), unnecessary two step solution where it can be a single step from end to end and this is where yolo ( you only look once)comes.

n this the image is split into S×S grid and each cell is responsible for predicting if there is a bounding box or not also will determine the class probability.  The cell that is responsible for determining the bounding box of the object is the one that contains the center pont of the object.

Yolo Metrics

1. Intersection over union.

A way to quantify how good our prediction is.

IoU = Area of intersection between prediction and the ground truth bbox/ area of union

Iou score of 1 indicates perfect prediction and iou score of 0 indicates no intersection between predicted and the ground truth box. Iou score of 0.5 is taken as threshold(in computer vision origin is top left corner). We discard the bbox with iou less than the threshold.

2. Non max supression (NMS)

It is basically cleaning of bbox when we get multiple bbox. In nms, we take out the box with highest class probability and then we calculate the iou between the chosen box and the other boxes. If the iou turns out to be higher than the threshold, then the ither boundary box is discarded. But if it is less than the threshold, then that bbox is not removed. But if we have multiple classes, we do the nms seperate for each class which means we cant compare iou across classes.

3. mAP

It is based on precision recall curve . The precision recall curve is based on the IoU and detection confidence curve. Recall measures how effectively we can find objects and precision tells us how well we perform once we find the object. Higher is better.  
Precision= no. Of true positives/ ( no. Of true positives + no. Of false positives)  
Recall = no. of true positive/ total no of ground truths.  
The area under the prevcion recall curve gives usvthe mean average precision.

4. Anchor boxes

These are predefined boxes of different sizes and aspect ratios that serve as reference for predicting bounding boxes. They are crucial in handling objects of various scales and shapes.

In this image we have two dogs , we have to detect the two dogs . The idea behind yolo algorithm is that well split it into s×s grid and each cell will output a prediction with a corresponding bounding box. Each dog is in multiple cell. We want one bbox for each object. The idea is well find one cell that is responsible for the bounding box which is the cell that contaisn the center point of the object. Each output and label will be relative to the cell . Each bbox will have [x, y, w, h] . X and y are object midpoint in cell and w and h is the width and the height relative to cell. X and y are between 0 and 1, w and h can be greater than 1 if the object is wider and taller than the cell. Labels will look like .. [ c1, c2, c3.......c20, pc, x, y, w, h]. Predictions will look very similar , but each cel will output two bounding boxes. A cell can detect only one object.

Predictions will look like......

[c1,c2,....c20, pc1, x, y, w, h , pc2, x, y , w, h]

Target shape for one image (S,S,25)

Prediction shape for one image (S,S,30)

APPLICATIONS OF OBJECT DETECTION

1. It helps in finding rusted parts in an iron tower.
2. It helps in locating tumors inside the person to determine whether it is cancerous or not.
3. In autonomous vehicles, it helps in detecting pedestrians, vehicles and obstacles.