# Real-Time Object Detection using YOLO: A review

[Resource link](https://www.researchgate.net/publication/351411017_Real-Time_Object_Detection_using_YOLO_A_review)

## Introduction:

Although the human eye is capable of instantly and precisely identifying a given visual, including its content, location, and visuals close by interacting with it, the human made, computer vision-enabled systems are relatively low in accuracy and speed. Any advancements leading to improvements in efficiency and performance in this field could pave paths to creating more intelligent systems, much like humans. These advancements, in turn, would ease human life through systems such as assistive technologies that allow humans to complete tasks with little to no conscious thought.

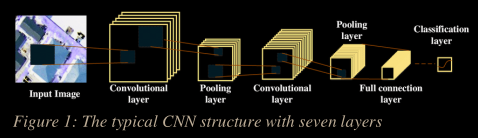
For instance, driving a car equipped with a computer vision-enabled assistive technology could predict and notify a driving crash prior to the incident, even if the driver is not conscious of their actions. Therefore, real-time object detection has become a highly required subject in continuing the automation or replacement of human tasks. Computer vision and object detection are prominent fields under machine learning and are eventually expected to aid unlocking the potential general-responsive robotic systems.

Most human lives revolved around mainstream pc and mobile phones. The expansion of information and images available on the internet/cloud has become to the point of millions per day.

Different strategies have been proposed to solve the problem of object identification throughout the years. These techniques focus on the solution through multiple stages. Namely, these core stages include recognition, classification, localization, and object detection.

## Structure of CNN:

A typical CNN is structured with multiple layers: an input layer, a convolutional layer, an active layer, a pooling layer, a fully connected layer and finally, an output layer.



### Types of object detection algorithms:

Classification based algorithms

Regression based algorithms

## YOLO:

YOLO is a novel approach to detect multiple objects present in an image in real-time while drawing bounding boxes around them. It passes the image through the CNN algorithm only once to get the output, thus the name. YOLO practically runs a lot faster than Faster R-CNN because of its simpler architecture. Unlike Faster R-CNN, YOLO can classify and perform bounding box regression at the same time.

YOLO algorithm extends GoogLeNet equations to be used

as their base forwarding transport computation, With 24 convolutional

layers, two fully connected layers, 1x1 reduction layers and

3x3 convolutional layers,

YOLO algorithm is an algorithm based on regression. It predicts class probabilities of the object and bounding boxes the object’s location, for the entire image. The bounding boxes of the object are described as: bx, by, the x, y coordinates represent the center of the box relative to the bounds of the grid cell. The bw, bh as the width and height are predicted relative to the whole image and the value c is representing the class of the object. YOLO takes the image as input and divides it into S x S grids (3 x 3). Then, image classification and object localization techniques are applied to each grid of the image and each grid is given a label. The YOLO algorithm then checks every grid for an object and identifies its label and bounding boxes. The label of a grid that does not have an object is indicated as zero. Every labeled grid is defined as S.S having 8 values. The 8 values namely are pc, bx, by, bw, bh, c1, c2, c3. Pc shows if a particular grid has an object or not. If an object is available,the pc is assigned 1 else 0. bx, by, bh, bw are bounding box parameters of a grid and are only defined if a proper object is available in that grid. c1, c2, c3 are classes.

## Strength and weakness of YOLO:

It maintains a good balance between speed and accuracy. YOLO achieved a mean Average Precision (mAP) rate of 76.8 at 67 Frames per Second (FPS) and 78.6 mAP rate at 76 FPS. Another great strength in YOLO is its global reasoning skills that encode the contextual information about the whole image rather than a specific region.

A weakness of YOLO is its spatial constraints on bounding boxes.