# A Probabilistic Object Detection in Computer Vision Using Deep Learning Approach

## Problem Statement

Humans can easily identify and detect the objects present in an image. Humans are very fast and accurate at identifying multiple objects and performing multiple tasks. So, humans can easily train the computers to classify and detect multiple objects within an image with high accuracy by having large amounts of data, faster GPUs, and appropriate algorithms. Humans train the computer for three tasks. They are image classification, object localization and object detection. Image classification is nothing but assigning class labels (person, animal, things) to image. It classifies the image into different class labels. Object localization separates the objects by drawing bounding box. Object detection combines the above two tasks.

But it is quite hard to distinguish the difference between object localization and detection. Because all these three come under object recognition. To safely operate in the real world, robots need to evaluate how confident they are about what they see around. A new challenge in computer vision algorithms to not just detect and localize objects, but also report how certain they are. Object detection is often an important part of the perception system of robots or autonomous systems such as driverless cars. It provides crucial information about the robot's surroundings and has significant influence on the performance of the robot in its environment. For example, driverless cars need object detection to be aware of other cars, pedestrians, cyclists and other obstacles on the road. Future domestic service robots and robots in healthcare will have to be able to detect a large range of household objects in order to properly fulfil their tasks.

## Background

This section provides the literature survey of the work done in the object detection field in the past few decades. Developers had started working on it since early 1958, but due to very poor processing speed and a deficient amount of storage space for large datasets it takes a long gap until most powerful Viola-Jones Algorithm comes in 2001, which uses Haar-Like features, Cascading and Ada-boost to detect faces. The modern steps of object detection go along with the improvement of Convolutional Nets (CN) which began in 2012 when Alex-Net won the 2012 image net large-scale visual recognition competition (ILSVRC). CN is utilized as image feature mapping. Alex-net based on old Le-Net along with data augmentation, ReLU, and GPU implementation. Girshick Ross introduced Region Based ConvNet (RCNN) which is a natural combination of heuristic region proposal method and CN feature extractor. The Alex-Net and support vector machine (SVM) model is then trained to classify the object. The overfeat method is introduced by Sermanet Pierre which uses Alex-Net to extract features at multiple evenly-spaced square windows in the image over multiple scales of an input image.

ZFNet is the ILSVRC 2013 winner, which is basically an Alex-Net with minor modification.

Spatial pyramid pooling net (SPPNet) is an enhanced version of RCNN by introducing two important concepts: adaptive-sized pooling and computing feature volume only once. For scalable and high-quality object detection Multi-box method is introduced which is not an object recognition but a CN based object proposal method. GoogLeNet (inception) is the winner of ILSVRC 2014 in which instead of using traditional conv and max-pooling layers, it stacks up inception modules. Fast RCNN is SPPNet with trainable feature extraction network and region of interest pooling in replacement of the SPP layer. Significant changes take place in the field of object detection when you only look once (YOLO) is introduced which is a direct development of the multi-box method. It turns multi-box from object proposal solution to an object recognition method by adding a soft-max layer.

Faster RCNN is the modification to Fast RCNN in which heuristic region proposal is replaced by the region proposal network (RPN) inspired by multi-box. Single Shot Multibox Detector was introduced in 2016 which can do everything in a single shot, it just has to look at the image once it does not have to go back to the same image. It does not have to do the object proposal and does not have to run many convolutional layers which reduces the time and computational cost.

## Methodology

**Step 1: Dataset Selection**

This will involve researching on various datasets and selecting the best one for our project.

**Step 2: Creation of Model Using Darknet Architecture**

In this step, a Darknet architecture is used to build a model for object detection. Various algorithms like imageai, RCNN, faster RCNN, SSD and YOLOv3 are tried for better accuracies and certainties of objects detected.

**Step 3: Selecting the best Algorithm & Training**

Training and testing is performed on model on the MS COCO datasets to do the prediction accurately.

## Experimental Design

**Dataset**

The dataset used for training the models is Microsoft Common Objects in Context (MS COCO).COCO dataset is an excellent object detection dataset that has 80 different classes, 80,000 training images and 40,000 validation images available.

**Evaluation Measures**

Evaluation is measured in terms of accuracy of probabilities mAP and FPS are used.

**Software and Hardware Requirements**

Python based Computer Vision and Deep Learning libraries will be exploited for the development and experimentation of the project. Tools such as Anaconda Python, and libraries such as OpenCV, Tensorflow, and Keras will be utilized for this process. Training will be conducted on NVIDIA GPUs/ CPU.