RESEARCH ARTICLE OPEN ACCESS

# YOLOv2 based Real Time Object Detection

Sakshi Gupta [1], Dr. T. Uma Devi [2]
Student [1], Department of Computer Science,
GITAM University, Visakhapatnam, Andhra Pradesh, India
Associate Professor [2], Department of Computer Science,
GITAM University, Visakhapatnam, Andhra Pradesh, India

## ABSTRACT

Object detection could be a primitive work to spot objects in an image and video processing. It's considered to be one among the difficult and challenging tasks in computer vision. There are many machine learning and deep learning models are proposed in the past like F-CNN, RNN, YOLO. Within the current scenario, requirement of detection algorithm is to figure end to finish and take minimum time to compute. Real-time object detection and classification from images and video provide the bottom for generating many sorts of scientific aspects as an example the majority of traffic signals during a particular district or total objects during a particular image. In process, the work usually encounters occurrence of errors or the slow processing of detection and classification due to the tiny and light-weight datasets to beat these problems, this paper proposes You Only Look Once version 2 (YOLOv2) based detection and classification approach. This model improves the time of computation and speed also as efficiently identify the objects in images and videos. Additionally, COCO-2017 dataset used for implementing YOLOv2due to the pretrained model of detection is already exist in it and it uses GPU to reinforce the speed and processes 40 frames per second.

***Keywords :***-- R-CNN, YOLOv2, Object classification, Object detection, F-CNN

## I. INTRODUCTION

Object detection is one among the classical problems in computer vision where the human employed to acknowledge what and where—specifically what objects are inside a given image and where are within the image. The real-time application is self-driving cars, ship detection, etc [1][3]. Object detection not only includes recognizing whether specific object is present or not, but also finds the precise position of that specific region where object is present. The matter of object detection is more complex than classification, which can also recognize objects but doesn't indicate where the thing is found within the image and also classification doesn't work on images containing quite one object. The aim of this paper is to detect multiple objects from an image and video. There are various techniques for object detection, it will often split into two categories, one is Classification based. Classification based categories like CNN, RNN and F-CNN pick out the interested regions from the image and classify them using convolutional neural network and this process called Selective Search [4]. CNN is incredibly slow because it predicts a selected region for each run. Subsequent category is predicated on Regressions. Sample of COCO dataset shown in figure consisting of sample objects like bottle, sofa, chair, motorbike, car presents the picture for detection and classification.

Sample of COCO Dataset

Neural networks have made the work very simple. Fast R-CNN neural networks to Faster R-CNN, all models have shared a crucial role within the field of computer vision. This paper focuses in classification and detection area from single class objects to multi class objects. Here, YOLO comes into picture where there is no need to select the regions in image. Instead, YOLO predicts the classes and bounding boxes of multiple objects in a complete image using a single neural network. YOLO could be a clever convolutional neural network for object detection in real-time. YOLO is extremely fast and process 40 frames per second. This algorithm makes localization errors but predicts less false positives within the background. YOLOv2 is that the extension of YOLO which works on framework called Darknet. YOLOv2 focuses on anchor boxes and use the features that are fine grained to vary smaller objects are often predicted better.

Darknet framework is employed to train neural networks, inspired by GoogleNet architecture which is written in C/CUDA. YOLOv2 is far faster than traditional approaches like R-CNN and produce minimum errors [4]. This model divides each image into grid boxes and every grid box makes prediction on bounding boxes related to confidence levels. Consistent with threshold values, most of the bounding boxes and grid boxes automatically removed if threshold value is extremely less.

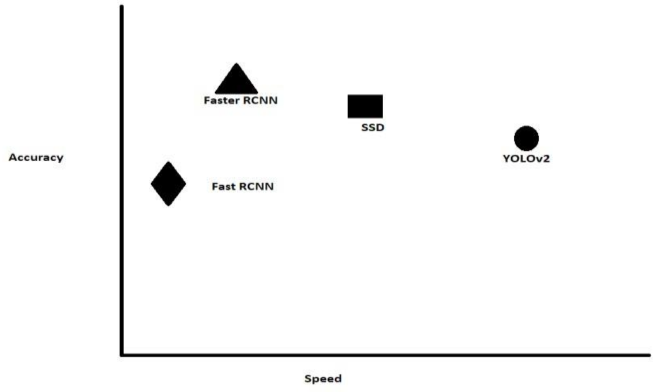


**Fig. 1. Speed v/s Accuracy of detecting algorithms.**

Fig. 1. shows the graph of speed v/s accuracy between several detecting algorithms [4]. This paper uses YOLOv2 algorithm on darknet framework and run it on the image dataset and video, which can predict the bounding boxes on the objects. OpenCV library is used for detection to recognize and classify face, objects on images using the features like shape, length, width, etc. [4]. There are two major problems within the traditional algorithms that motivates this work i.e. low accuracy rate and slow

speed of algorithm due to GPU (CNN, R-CNN, F-CNN doesn't have GPU support). YOLOv2 uses sliding window-based method for detecting objects in single shot detection framework.

## II. LITERATURE SURVEY

Real-Time Object Detection with YOLO, by Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam. In this paper, their work is to detect multiple objects from an image using YOLO approach [1]. You Only Look Once: Unified, Real-Time Object Detection, by Joseph Redmon. This paper explains object detection as regression problem and repurposes classifier using YOLO approach [2]. Object Detection and Recognition in Images, by Sandeep Kumar, Aman Balyan, Manvi Chawla. This paper used Easynet model to recognize images and detection of objects for instances of real objects like bicycles, fruits, animals and buildings in images [3]. Object Detection and Classification Algorithms using Deep Learning for video Surveillance Applications, by Mohana and H. V. Ravish Aradhya. This paper prior work is the classification of objects in images and video, have use YOLOv2 approach [4].

## III. WORKING OF YOLOV2 ALGORITHM

Step 1- An image is taken and divide it into a grid cell. Here, example has taken where the image splits into grids of 7x7 matrices. It will divide the image into any number of grids, looking on the complexity of the image.

Step 2- Once the image is split, classification and localization of the image is performed in each grid cell. If an object is detected then it represents the probability of each grid vector. The output of this is the dimension of bounding box and class.

Step 3- Now, thresholding is performed and based on the value grid cells with the highest probabilities are picked. This step produces the removal of bounding boxes which doesn't have object or the confidence score less than a threshold of 0.35.

Step 4- The YOLOv2 algorithm uses Anchor Boxes which detects the objects in single grid cell and gives the location of object. Finally, Non-max suppression uses Intersection over Union for final detection.

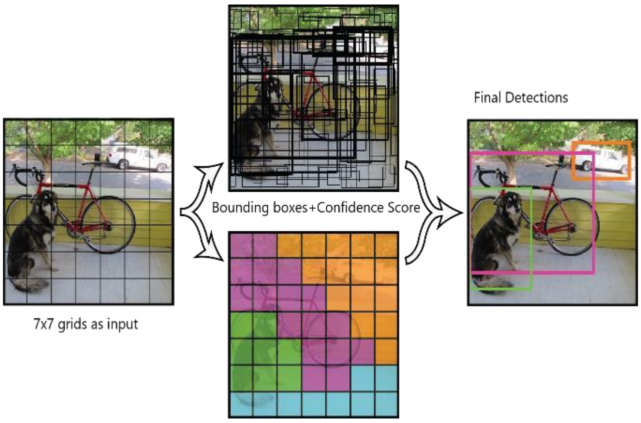
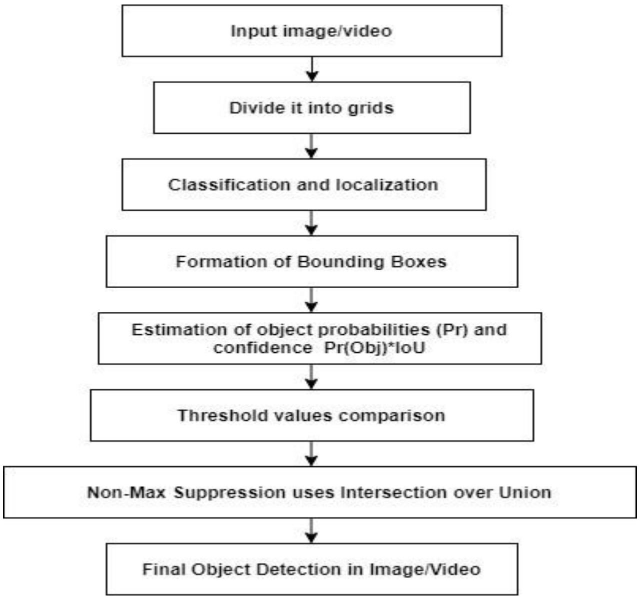


**Fig. 2. Working of YOLOv2**

In fig 2. Shows the working of the algorithm which divides the image into SxS grid cell and every grid cell predicts bounding boxes and class probabilities which is mapped in different colors [2].

### *A. Flow Diagram of YOLOv2 Model*


Input image/video
Divide it into grids
Classification and localization
Formation of Bounding Boxes
Estimation of object probabilities (Pr) and confidence Pr(Obj)*IoU
Threshold values comparison
Non-Max Suppression uses Intersection over Union
Final Object Detection in Image/Video

### *B. Bounding Boxes*

The YOLOv2 model is used to provide accurate prediction of boxes in the image as shown in fig 3. Here, image contains two cars and this image is divided into 3x3 grids to predict the bounding boxes in each grid. Later, it shows the class probabilities to predict the object in that specific grid. Label is allotted to each grid of the image in the predicted bounding boxes which is applied to both image classification and object localization. Finally, when object is detected then bounding box shown on the image and the grid which doesn't contain any object their class probability is shown as 0.

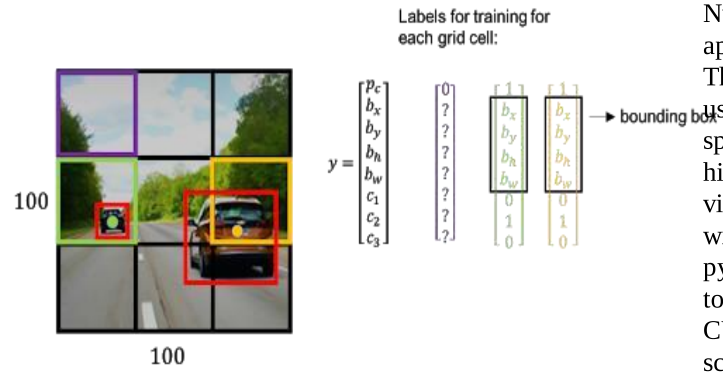


**Fig 3. Example of bounding box and class values**

### *C. Anchor Boxes*

Anchor boxes is a set of bounding boxes that is predefined with a specific height and width. The anchor boxes are used to solve the issues i.e. prediction of the localization of object. Here, algorithm divides the input image into any grids like PxP cells. Now these cells find the mid-point of the object and if an object found within the midpoint the localization task is completed. If this mid-point coincides with two objects then YOLOv2 picks one of the objects among them.

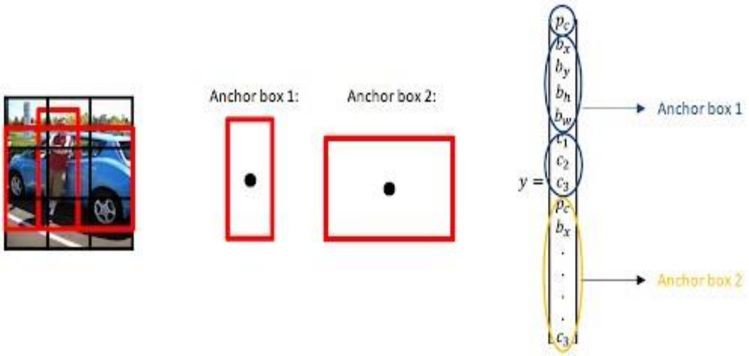


**Fig 4. Example of anchor boxes and prediction values**

In fig 4. shows the image is divided into 3x3 grids containing a car and a person. If the image classification and object localization algorithm is applied to classify three sets of categories, for example, mango, banana and orange, then the output vector "target variable" of the neural net are usually outlined as a matrix of eight possible outcomes.

### *D. CUDA (Compute Unified Device Architecture)*

CUDA is a parallel processing architecture developed by Nvidia to make use of GPU resources. It is used in a variety of applications like machine learning, parallel computing, etc. This paper uses Darknet framework with GPU support and to use this framework having CUDA is necessary. It boosts the speed of darknet in image processing because CPU may cause hindrance to productivity for any processing of images or video. CUDA only supports Nvidia hardware, it can be used with several different programming languages like C++, python. The host and therefore the device work hand in hand to enhance the workflow and computation speed. In YOLOv2, CUDA environment plays an important role. In current scenario, normalization of object is easier using CUDA as noise reduction and object redundancy is predicted easily and fastly. cuDNN is a library of CUDA which provides GPU support to boost the speed of system.

## IV. RESULTS & ANALYSIS

The paper proposes YOLOv2 to make reorganization layer. The reorganization layer uses alternate pixel and then creates a special channel. For instance, with 3x3 pixels in a single channel the reorganization layer reduces the size and creates pixels in different channels that is adjacent. YOLOv2 uses batch normalization in all convolutional layers which improves mAP. Here, fig 5. shows the network architecture of YOLOv2 which represents the convolutional layer starting with 3x3x32 and ends with 3x3x1024 and shows how this network process40 frames at a time.

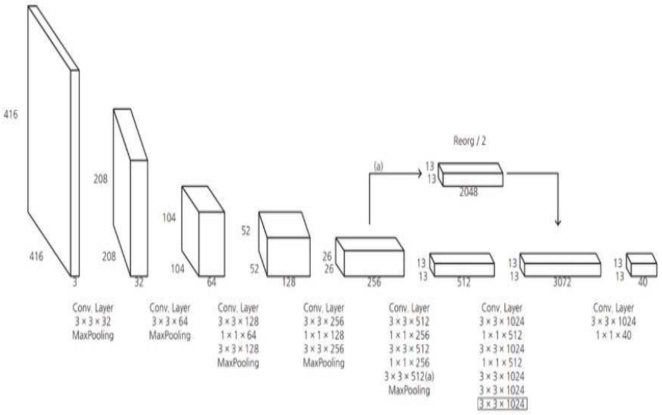


**Fig 5. YOLOv2 Network**

For using the model it's mandatory to install Microsoft Visual Studio 2015, NVIDIA GEFORCE with GPU and CUDA 9. To use pretrained model MS COCO dataset all these are useful to detect objects and classify objects. To use Darknet framework CUDA installation is necessary. Now, the results of detection of objects has shown below where single object image given as input and fig6.shows the detection and labelling of the objects with localization on that single image.

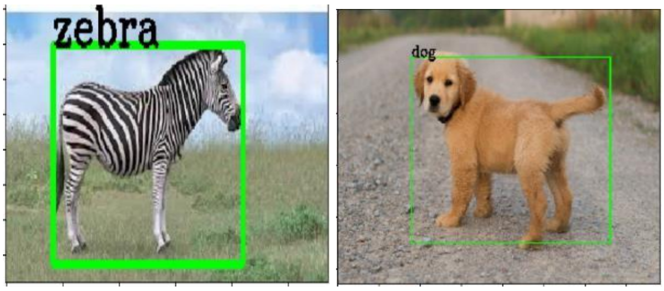


**Fig 6. Detection and labelling of single image**

Now, if further when objects count increases then GPU support doesn't lower the execution speed.Fig. 7 shows the detection of multiple objects in a single image. Both images show the detection of different objects like dog, person and horse and in another image bicycle, car, person, etc.

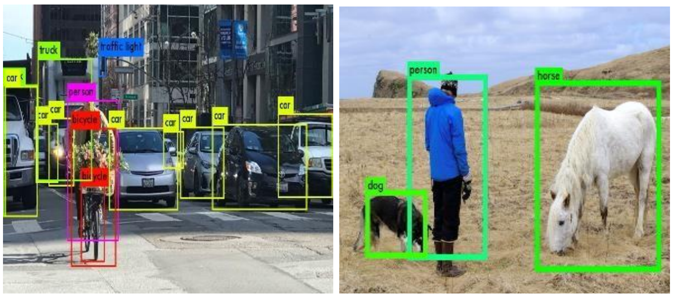


**Fig 7. Detection of multiple objects in single image**

When it comes to video records, it's totally different. In video records objects are moving and continuously changes in very high speed. Here, images are taken from the video record that has recorded in traffic signal. Fig. 8, shows how the algorithm works with accuracy and speed including the object detection. Objects in video are multiple like car, person, bag, signal, etc. but the detection using darknet framework and YOLOv2 model are accurate whether the objects are increasing continuously.

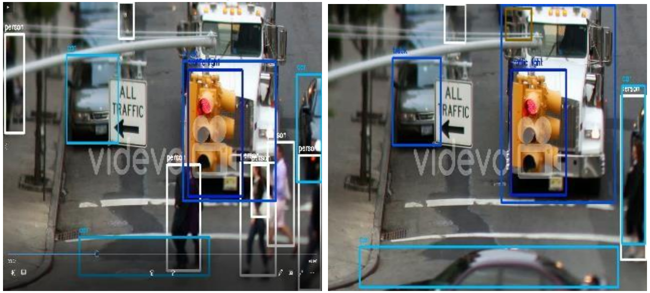


**Fig 8. Real time detection using video records in traffic signal**

## V. CONCLUSION

This paper proposes YOLOv2 algorithm for the detection of objects in images with localization and video records. The main aim of this paper is to detect the objects in real time i.e. live detection using webcam and also through video records. GPU version is extremely fast which helps the functionalities perform accurate using anchor boxes. The dataset used in this paper is COCO which consists 80 classes. Using the model YOLOv2 it is easy to detect objects with grids and boundaries prediction and also it helps in predicting with very small objects or objects which very far in the image. In video records detection of moving objects are easier using darknet and it produces .avi file with detections. In live detection system uses webcam to detect live objects. Pretrained datasets helped to detect in efficient way and classifying the objects in less time.

## ACKNOWLEDGMENT

I would like to show my gratitude to Dr. T. Uma Devi, Associate Professor, GITAM Institute of Science for sharing her ideas in this paper and helping to remove anomalies.

## REFERENCES

[1] Redmon Joseph, et al. "You only look once: Unified, real-time object detection." proceedings arXiv in May 2016.

[2] Geethapriya. S, et al. "Real-Time Object Detection with Yolo" proceedings of the International Journal of Engineering and Advanced Technology (IJEAT) inFeb 2019

[3] Swetha M S, et al. "Object Detection and Classification in Globally Inclusive Images Using Yolo" proceedings of the International Journal of Advance Research in Computer Science and Management Studies (IJARCM) in Dec 2018

[4] Keerthana T, et al. "A REAL TIME YOLO HUMAN DETECTION IN FLOOD AFFECTED AREAS BASED ON VIDEO CONTENT ANALYSIS" proceedings of the International Research Journal of Engineering and Technology (IRJET) in Jun 2019

[5] Sandeep Kumar, et al. "Object Detection and Recognition in Images" proceedings of the International Journal of Engineering Development and Research (IJEDR)in 2017

[6] M R Sunitha, etal. "A Survey on Moving Object Detection and Tracking Techniques" proceedings of the International Journal of Engineering and Computer Science (IJEAC) in 2016.

[7] Jifeng Dai, et al. "R-FCN: Object Detection via Region-based Fully Convolutional Networks", proceeding of the Advances in Neural Information Processing Systems 29 (NIPS) in 2016.