IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Object Identification using YOLO Algorithm

Binny S, Cina Mathew, Cini Joseph

Kristu Jyoti College of Management and Technology

ABSTRACT *- The aim is to identify object using the You Only Look Once (YOLO) method. Comparing this method to other identification algorithms, reveals that it offers limited benefits. In algorithms like Convolutional Neural Network(CNN), Fast-Convolutional Neural Network(FCNN) the algorithm does not take the picture totally, but using YOLO algorithm it takes the picture totally. The YOLO algorithm does the detection by using the bounding boxes. In these cases, the class probabilities make it easier to differentiate the image better when compared to other algorithms.*

Keyword: CNN, YOLO, Bounding Boxes, FCNN

1. INTRODUCTION

Object discovery is a way to find the meaningful parts of a class in digital images and recordings. In this paper our objective is to find out the various items from the image. For finding an item, we use Object localization. Object identification, can be done using different methods. First is the calculation using CNN and RNN. In this, we need to choose the intrigued areas from the picture and need to arrange them utilizing Convolutional Neural Network. The second method is calculations based on regressions. YOLO technique belongs to this class. In this, we cannot select the region of interest from the picture. The multiple objects can be detected using single neural network and we can predict the classes and bounding boxes of the entire picture in a single execution of the algorithm. YOLO method performs better than other classification algorithms.

II. RELATED STUDY

Joseph Redmon's “You Only Look Once: Unified, Real-Time Object Detection”. Their earlier research focused on using a regression approach to find objects. In this study, the YOLO algorithm was proposed to get better accuracy and predictions [1].

They generally discussed object detection families such CNN and R-CNN in this study, compared their efficacy, and presented the YOLO algorithm to improve it [2]. By Matthew B. Blaschko, "Learning to Localize Objects using Structured Output Regression." The objective of this paper is object localization. To get around the limitations of the sliding window approach, they adopted the bounding box method for object localization in this [3].

III. YOLO ALGORITHM - WORKING

The initial process is to apply YOLO calculation to the picture. In our model, the picture is separated as frameworks of 3x3 lattices. We can partition the picture into any number networks, contingent upon the intricacy of the picture. When the picture is separated, every network goes through arrangement and restriction of the item. Every network's objectness or confidence score is discovered. The grid bounding value is set to zero when the object is not found and it is set to 1 if the object is found. The bounding box prediction is discussed in the next section. In order to enhance the efficiency of an object detection, anchor boxes are also deployed, as is further described below.

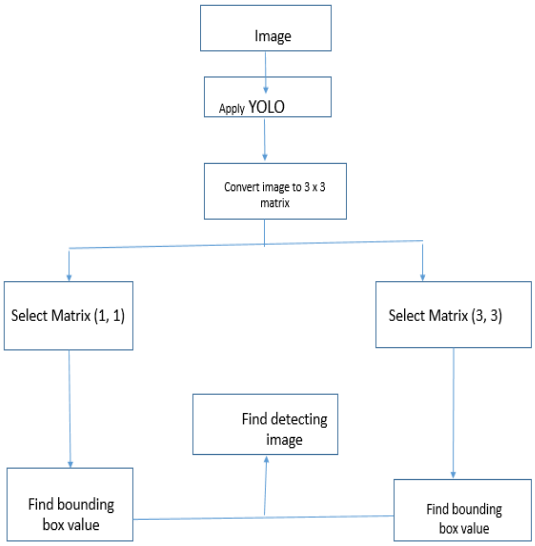

Fig 1: YOLO working

3.1. Bounding box predictions

In order to get better prediction of bounding boxes from the picture YOLO algorithm can be used. The picture partitions into S x S frameworks by anticipating the bouncing boxes for every matrix and class probabilities. Each grid is assigned with a label when both image classification and object localizations techniques are used. The algorithm checks each cell separately and marks the label which contains an object and also labels the bounding boxes.

The labels of the null matrix are represented as zero.

Fig 2: 3x3 matrix image

In the example above, an image is captured and split into 3 x 3 matrixes. Each matrix is named, and every cell is put through algorithms for object localisation and image classification. The label is considered to be Y. Y has eight different values.

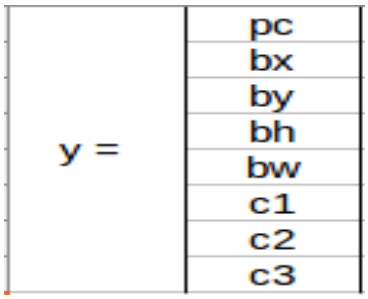

Fig 3: Label Y elements

Pc refers the presence or absence of an object. If an object is present in the cell Pc=1 else Pc = 0.

If present the bounding boxes of the objects are bx, by, bh and bw.
The classes are c1, c2 and c3.
(c1,c2,c3) = (0,1,0), then the object is a car.

In the example, the first grid does not contain any object. So it is represented as,

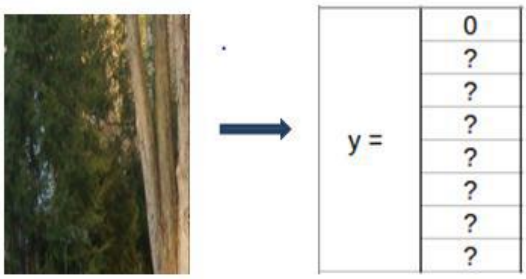

Fig 4: Values of Bounding box and Class of grid 1

In this lattice, there exists no object so the Pc value is set to 0.

Furthermore, rest of the qualities are doesn't matter since there exist no item. In this way, it is addressed as ?. Think about a network with the presence of an item. Both fifth and sixth network of the picture contains an item. Let' consider the sixth network, it is addressed as.

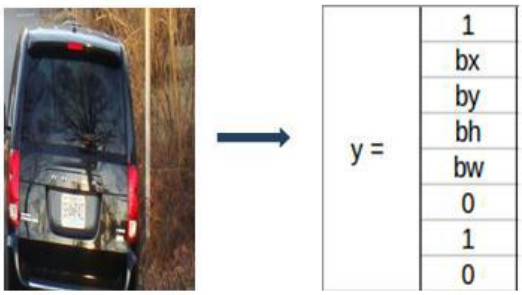

Fig 5: Values of Bounding box and Class values of grid 6.

In the above table, bx,by, bh and bw are the bounding box of the object in grid 6.The value 1 denotes the presence of the object. The class (0,1,0) represents a car. The matrix of Y is Y=3x3x8. For the fifth cell also have some similarity with other bounding boxes.

On the off chance that at least two lattices contain a similar item, the middle place of the item is found and the framework which has that point is taken. For this, to get the precise recognition of the item we can use to techniques. They are Intersection over Union and Non-Max Suppression. In InU, it will take the genuine and anticipated jumping box esteem and ascertains the InU of two boxes by utilizing the formulae,

InU = Area of Intersection / Area of Union.

A reliable prediction may be made if the InU value exceeds or is equal to our threshold value (0.5). The threshold value is only an assumption. To improve accuracy or to better predict the item, we may also use a higher threshold value.

The second method, known as non-max concealment, involves taking high probability boxes while suppressing instances with high InU. This should be repeated until a crate is determined; use that as the bouncing box for that thing.

3.2 ANCHOR BOX

Only a single object can be detected on a grid using bounding box. For recognizing more than one item Anchor box is used.

Fig 6: Anchor box example

Consider the image above, where the midpoints of the human and the car are both below the same matrix cell. We apply the anchor box approach in this case. Both of the articles' anchor boxes are in the red variety matrix cells. In a single image any number of anchor boxes can be used to detect multiple objects. In the above fig 6, two anchor boxes is used.

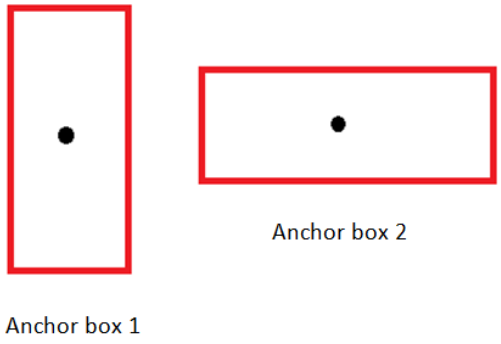

Fig 7: Anchor box 1 and 2

In fig 7, Anchor box 1 is used to represent a human and Anchor box 2 is used to represent vehicle.

$$y = \begin{bmatrix} p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \\ p_c \\ b_x \\ b_y \\ b_h \\ b_w \\ c_1 \\ c_2 \\ c_3 \end{bmatrix} = \begin{bmatrix} 1 \\ b_x \\ b_y \\ b_h \\ b_w \\ 1 \\ 0 \\ 0 \\ 1 \\ b_x \\ b_y \\ b_h \\ b_w \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

Anchor box 1
Human

Anchor box 2
Car

Fig 8: Prediction values of Anchor box

The presence of object in both the anchor boxes is represented by Pc. The corresponding bounding values in both the anchor boxes are
represented by bx, by, bh, bw
In the case of anchor box 1, the detected object is a human with class value (1,0,0) and for anchor box 2, the detected object is a car with class value (0,1,0).
Because of two anchor boxes (i.e., 2 x 8), the matrix form of Y = 3x3x2x8 or Y = 3x3x16.

IV. RESULTS AND DISCUSSION

Making CNN to predict a (7, 7, 30) tensor is the concept behind YOLO. It decreases the spatial dimension to 7x7 with 1024 output channels for each location using a convolutional neural network. It does

a linear regression to get a 7x7x2 bounding box prediction using two fully connected layers. Finally, a conclusion is formed by taking into consideration a box's high confidence score.

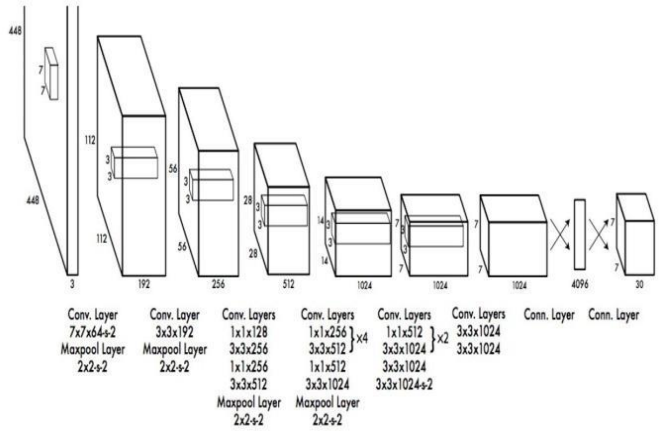

Fig 9: Network Design of CNN

4.1. Loss function

The computation forecast various bounding boxes for a single lattice. Here only one bounding box is used for finding loss function. The bounding box with high InU value is selected. The box with high InU will be liable for the article.

REFERENCES

1. Joseph Redmon, Santosh Divvala, Ross Girshick, "You Only Look Once: Unified, Real-Time ObjectDetection", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788.
2. YOLO Juan Du1," Understanding of Object Detection Based on CNN Family", New Research, and Development Center of Hisense, Qingdao 266071, China.
3. Matthew B. Blaschko Christoph H. Lampert, "Learning to Localize Objects with Structured Output Regression", Published in Computer Vision – ECCV 2008 pp 2-15.
4. Dumitru Erhan, Christian Szegedy, Alexander Toshey, "Scalable Object detection using deep Neural networks", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2147 – 2154.
5. Shaoqing Ren, Kaiming He, Ross Girshick Jain Sun, "Faster R-CNN : Towards Real – Time Object Detection with Region Proposal Networks", Published in Advances in Neural Information Processing Systems 28 (NIPS 2015)
6. Joseph Redmon, Ali Farhadi, "YOLO9000 : Better, faster, Stronger", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp.7263 - 7271