REAL TIME OBJECT DETECTION USING YOLO ALGORITHM

Bobburi Taralathasri[1]; Dammati Vidya Sri[2]; Gadidammalla Narendra Kumar[3]; Annam Subbarao[4]; Palli R Krishna Prasad[5]

Department of Computer Science and Engineering, Vasireddy Venkatadri Institute of Technology, JNTUK, Guntur, India
DOI: 10.47760/ijcsmc.2021.v10i07.009

Abstract- The major and wide range applications like Driverless cars, robots, Image surveillance has become famous in the Computer vision .Computer vision is the core in all those applications which is responsible for the image detection and it became more popular worldwide. Object Detection System using Deep Learning Technique" detects objects efficiently based on YOLO algorithm and applies the algorithm on image data to detect objects.
Keywords - Video Surveillance, Computer Vision, Image Classification, Neural Networks, Deep learning.

I.   INTRODUCTION

The main aim of Object Detection is to detect all the objects that are present in that particular image. The important thing in object detection is to make system detect the object by making it to learn how to do that which is nothing but training the dataset and allowing system to detect the object by itself. The searching or recognition process in a real time scenario is so tough. So far, no effective solution has been found. Though there is a lot of work going on, the methods developed till now aren't efficient, require long training time, are not suitable for real time application, and are not scalable to large numbers of classes. Object detection is relatively simpler if the machine is looking for detecting one particular object. However, recognizing all the objects inherently requires the skill to differentiate one object from the other, though they may be of the same type. Such a problem is very difficult for machines, if they don't know about the various possibilities of objects. When anyone goes to any hospital or to a Traffic controlling monitor or self driving cars all they do is to detect the objects for the best performances. For example in a hospital when the doctor sees the scan report they will be able to tell whether there is a virus in our body or not. So we decided to make a REAL TIME OBJECT DETECTION System. Sometime it's very tough to negotiate between different Object detection techniques. In image classification we assign the class label to an image, In Object localization we draw the bounding boxes around the objects. Detecting the Object is tough because we combine both the tasks where we include both the bounding box around the each object and then we assign the class label to an image, all this comes under the category of Image[1].

II.   RELATED WORK

At present there are Automated systems which uses the Object detection techniques to detect the object, in that object detecting techniques they have used the CNN older algorithms. But here we are using the Yolo algorithm which is new algorithm to detect the objects quickly with more accuracy so that detecting speed will be improved and action will be done quickly[2]. During the last years, it evolved very quickly and successful expansion on computer vision research. Being Success in this region is because of adapting the Object detecting Techniques, whereas others from the development of the representation and models or from developing the efficient algorithms. Object detection has attained great progress in computer vision. The current works gives how actually the object detection techniques works.

In Object detection it first detects the location and scale of an input objects instance if such things are present in an image. So the objective of the Object detector is to find any number of the object belongs to particular class irrespective of class type objects, location, size in the input image. Object detection is the first thing that is being executed in the Computer systems so that it will collect the additional information such as (i) recognize the particular instance (e.g: face of an human),(ii) tracking the object over the image sequence(e.g tracking the action), (iii)To get more details about object. Many application uses the Object Detection techniques, among some of the popular applications are: (i) siri or alexa(Human Computer interaction) (ii) robots (iii) smart phones, (iv) Data tracking, (v) search engines such as Google, Firefox and suggest ads (vi) Driverless cars. Every applications in the object detection has different features some considers the faces like that whereas some consists the previously searched data and updates automatically, normally in these techniques some of them will consider the single object in single view only and then it will give the output regarding that view only and whereas some systems will detect the single object in multiple views and some of them are like for multiple images they detect the multiple views but the output is based on the views we are given during training.

III.   THE PROPOSED FRAMEWORK

So what we proposed is to detect the image as fast as possible with the best accuracy of the object. Here we are using YOLO(You only look once)   Algorithm to make it work fast ,it detects and identifies the object.
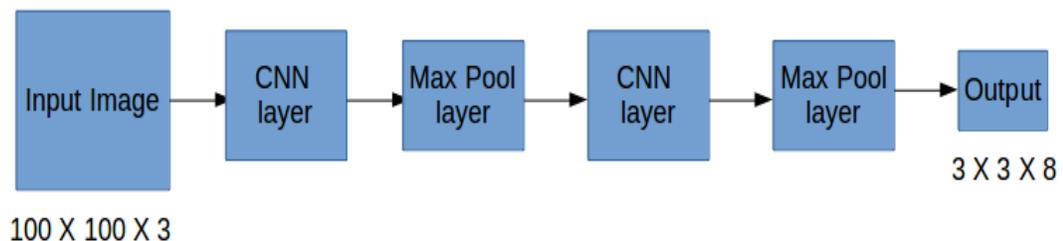

Fig1: Process Flow

Yolo Algorithm: YOLO Stands for (**you only look once**).Previous Object detection algorithms use the regions to identify the object but don't use the entire image , some parts may contain the objects . YOLO or You Only Look Once is an object detection algorithm much different from the region based algorithms which are seen above. In YOLO a single convolutional network predicts the bounding boxes and the class probabilities for these boxes.
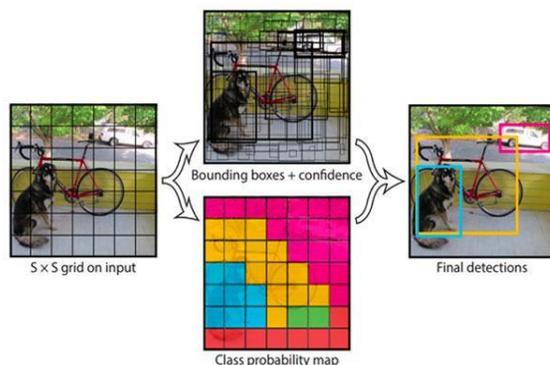

Fig 2 Yolo Algorithm Process

It's difficult for everyone to contain the resources for the DL research. So that's where this Yolo came into the picture. And lots of pre-trained models and datasets are available now.

Image processing in YOLO: Yolo algorithm splits the given image into the SXS grid and it forms the m bounding boxes where every grid has the m bounding box and every bounding box will be having the class output probability in addition to the offset values. The bounding boxes have the class probability along with the offset values. Threshold value is used to know if it is needed image or not. Yolo Object detection algorithm is faster than any other algorithm as it process the 45 frames per second. The only limitation of the Yolo algorithm is that it's not able to identify the small images that are in the image because of the spatial constraint in that algorithm for example flick of the birds.

The Yolo algorithm deals with object detection in a unique manner to other detection algorithms.

The Yolo algorithm takes the entire object into consideration and it takes the entire image in a single instance and forms the bounding boxes around the object. The main advantage of the Yolo is it is very fast and processes 45 frames per second.

It stores the data in Vector Form:- **Y=(pc,bx,by,bh,bw,c1,c2,c3),**
Where pc defines the probability and indicates if object is present or not,bx, by, bh, bw specifies whether there are objects for the classes c1,c2,c3. So if there is any object with respect to class c1 then it will have the value 1 otherwise 0.

If there are more bounding boxes for particular image then Yolo uses the non max suppression the bounding box with more accuracy is selected and remaining are ignored.

Formula for Non Max Suppression is :-

IoU = Area of the intersection / Area of the union

CNN in YOLO: From the last few decades Object detection has become more Powerful because it can handle large amounts of the data and it is evolving like anything .Nowadays everyone is using hidden layers rather than the

traditional methods. Researchers has struggled a lot in order to make system to understand the visual data since from 1950s.This field is used to be called Computer vision in the following years. CNN came into limelight when the group of researchers did a project on object detection using convolutional neural networks in 2012[3]. It made everyone in the object detection field look into that algorithm as it was the best algorithm of that time so far it contains the hidden layers which process the image recognition in it[4].
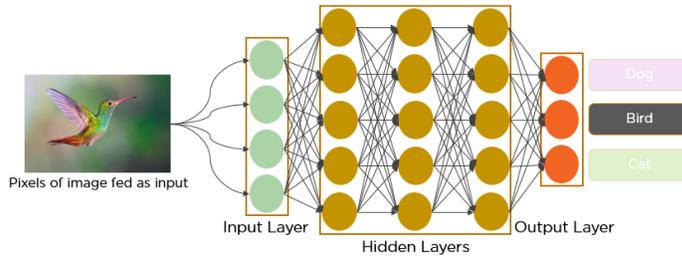
Fig 3 Layers in CNN

In 2012 AlexNet had won the first prize in the ImageNet Computer vision contest with the accuracy of 84% and it used the CNN for the Object detection and The runner up accuracy is 73%.

CNN has become important to any Computer vision techniques and it is heart for the AlexNet computer vision project which won in the ImageNet Compute vision competition. So let's look into the CNN Algorithm.

Working of CNN: Before we go deeper into the CNN let's first how the image is represented in it and what was that we can actually have is two kinds of images those are grayscale and RGB[3]. In RGB we will be having 3D matrix with values where in grey Scale image will be having only Single plane. Take a look into picture to understand it.
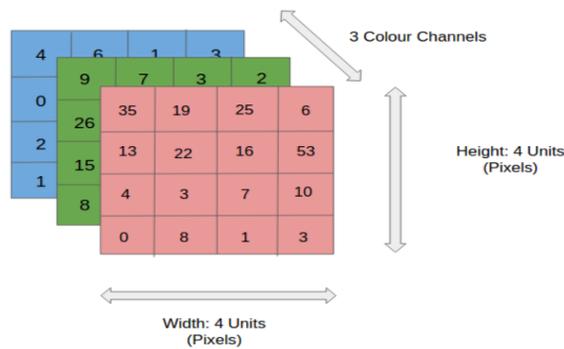
Fig 4 Matrix Calculation

Convolutional neural networks usually contain many artificial neural network layers which contains the artificial neurons. Artificial neurons are the imitation of the neurons in the human body which are called biological neurons, they actually are the mathematical functions which calculates the weighted sum of the inputs and gives the activation value as the output .When you pass the image to CovNet every layer generates the activated values which are being passed to the next Layer.

In CNN the first layer detects the features like horizontal and diagonal edges and that will be passed to the next level which will detect the corner and combinational edges features and the following layers will detect the more difficult features like face, birds etc.

IV. IMPLEMENTATION

Yolo Algorithm: Yolo takes the image as the input and then it divides it into SxS grid and every grid will have the m bounding boxes. Every Bounding box will contain the Output class probability and the offset value. Bounding box which has the class probability more than the threshold value will be considered to detect the object in the input image. Yolo algorithm is very speed compared to any detection algorithms in the Computer vision it will process the 45 frames per second. But the only limitation of the Yolo is it cannot detect or give the desired output for the small objects like flock of birds because of its spatial constraints[7].

The Yolo algorithm is different from the other object detection algorithms. It will take the entire object in a single instance and process the image and the main adage of the Yolo is, its Superb speed it processes the 45 frames per second and it forms the bounding boxes around the object and assigns the class probabilities and it can also understand the Object Generalization[8]

We will use both forward and backward propagation models for the training of the dataset. In testing phase we will be passing the image to it and we will perform the forward propagation until we get the output for that. In real world scenarios the grid size usually the large size it can be ven 20X20 based on the input we are providing.

Algorithm for Object Detection:-

INPUT: Trained Image dataset, Testing Trained image dataset

OUTPUT: Input image labeled with its class name along with the Average precession value with Rectangular box around every Object.

1: Pass the Input Image for detection

2: yolo Algorithm for Image

3: Yolo Algorithm process the image

4: Based on the trained dataset it assigns the class   names

5: Verify threshold value crosses 0.5 or less

6: If the threshold is less than 0.5 then it is original and provides the mean Procession value otherwise it simply ignores.

V. RESULTS

The image which we passed for input will be divided into the same number of the grids as in the training dataset. For every grid the output value will be 3x3x19.16 values will be the same as in the prediction model. The first value is the probability of the object corresponding to a particular class and the first 8 values will be the corresponding to the anchor box 1 .and the values from the 2 to 5 will be the values corresponding to the bounding boxes for the image and the last three tell us the classes that are present in that particular image. And the next 8 values will belong to the 2nd Anchor box which follows the same format and coordinates.
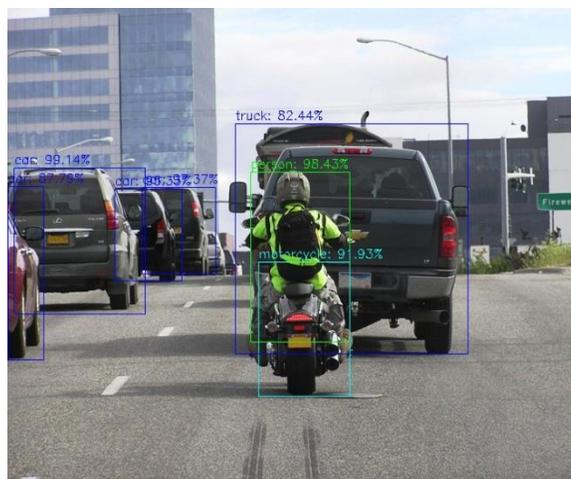
Fig 5 Input Image for detection

Fig 6 Output Image with Object names in   it.

And last we will apply the Non-Max suppression for every bounding box to get the single bounding box.

Training of the Yolo follows the below steps:

Input image of the shape(608, 608, 3)

Now we will pass this input image to the CNN which produces the (19,19,5,85) which will flattens the last 2 dimensions and combined will give (19,19,425).

Here,

1. Every cell in the grid 19 x 19 will returns the 425 numbers

2. 85*5= 425, 5 is the number of the anchor boxes that will be present in the grid.

3. 85 = 5 + 80, 5 are the center, width, height, x, y values of the bounding boxes and the 80 is the total classes that we have to detect.

4. Finally we apply the Non max suppression in order to get the single box around the detected object to avoid the overlapping we use the Non Max suppression.

VI. CONCLUSION AND FUTURE SCOPE

Now a days we are seeing that object detections applications are becoming popular in most of the fields , so based on that perception we've developed the console based application which takes image as a input and gives the same image with detecting object names on the top of the bounding boxes which are drawn around the image. Since this is the custom trained data set we've trained the dataset using Google Colab since this is the supervised learning because we are labeling the data using the LabelImg. So here we are using the YOLO to give faster results, we are using the latest version for accurate output.

REFERENCES

[1]. Tanvir Ahmad et al.,"Object Detection through Modified YOLO Neural Network", Hindawi.

[2]. P. F. Felzenszwalb et al., "Object detection with discriminatively trained part-based models," IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 9, pp. 1627–1645, Sep. 2010

[3]. P. N. Druzhkov and V. D. Kustikova, "A survey of deep learning methods and software tools for image classification and object detection," Pattern Recognit. Image Anal., vol. 26, no. 1, pp. 9–15, 2016.

[4]. A. Krizhevsky et al., "ImageNet classification with deep convolutional neural networks," in Proc. NIPS, 2012, pp. 1097–1105.

[5]. X. Zeng et al., "Gated bi-directional CNN for object detection," in Proc. ECCV, 2016, pp. 354–369.

[6]. S. Ren et al., "Object detection networks on convolutional feature maps," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 7, pp. 1476–1481, Jul. 2017.

[7]. A. Tiwari, A. Kumar, and G. M. Saraswat, "Feature extraction for object recognition and image classification," International Journal of Engineering Research & Technology (IJERT), vol. 2, pp. 2278–0181, 2013.

[8]. J. Yan, Z. Lei, L. Wen, and S. Z. Li, "'e fastest deformable part model for object detection," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2497–2504, New York, NY, USA, 2014.