REAL-TIME OBJECT DETECTION USING YOLO: A REVIEW

Shanka Subhra Debnath*, Rahul Banik*, Purba Das*

Department of Computer Science From Techno College Of Engineering Agartala

ABSTRACT-- With the vacuity of enormous quantities of data and the need to computerize visual- grounded systems, exploration on object discovery has been the focus for the once decade. This need has been accelerated with the adding computational power and Convolutional Neural Network(CNN) advancements since 2012. With colorful CNN network infrastructures available, the You Only Look formerly(YOLO) network is popular due to its numerous reasons, substantially its speed of identification applicable in real- time object identification. Followed by a general preface of the background and CNN, this paper wishes to review the innovative, yet comparatively simple approach YOLO takes at object discovery.

Keywords—YOLO, CNN, object detection, image classification

I. INTRODUCTION

Object Discovery is a technology that detects the semantic objects of a class in digital images and vids. One of its real-time operations is tone- driving buses . In this, our task is to descry multiple objects from an image. The most common object to descry in this operation is the auto, motorcycle, and rambler. For locating the objects in the image we use Object Localization and have to detect further than one object in real- time systems. There are colorful ways for object discovery, they can be resolve up into two orders, first is the algorithms grounded on groups. CNN and RNN come under this order. In this, we've to elect the interested regions from the image and have to classify them using Convolutional Neural Network. This system is veritably slow because we've to run a vaticination for every named region. The alternate order is the algorithms grounded on Retrogressions. YOLO system comes under this order. In this, we will not elect the interested regions from the image. rather, we prognosticate the classes and bounding boxes of the whole image at a single run of the algorithm and descry multiple objects using a single neural network. YOLO algorithm is fast as compared to other bracket algorithms. In real time our algorithm process 45 frames per alternate. YOLO algorithm makes localization crimes but predicts lower false cons in the background.

ii. WORKING OF YOLO ALGORITHM

First, an image is taken and YOLO algorithm is applied. In our illustration, the image is divided as grids of 3x3 matrixes. We can divide the image into any number grids, depending on the complexity of the image. Once the image is divided, each grid undergoes bracket and localization of the object. The objectness or the confidence score of each grid isfound.However, also the objectness and bounding box value of the grid will be zero or if there set up an object in the grid also the objectness will be 1 and the bounding box value will be its corresponding bounding values of the set up object, If there's no proper object set up in the grid. The bounding box vaticination is explained as follows. Also, Anchor boxes are used to increase the delicacy of object discovery which also explained below in detail.

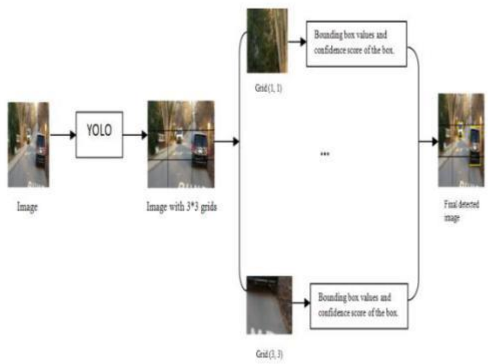

Fig 1: Working of YOLO

1. *Bounding box predictions*

YOLO algorithm is used for prognosticating the accurate bounding boxes from the image. The image divides into S x S grids by prognosticating the bounding boxes for each grid and class chances. Both image bracket and object localization ways are applied for each grid of the image and each grid is assigned with a marker. also the algorithm checks each grid independently and marks the marker which has an object in it and also marks its bounding boxes. The markers of the slur without object are marked as zero.

iii. METHODS OF IMPLEMENTATION

A. *Object detection frame differencin*

At regular intervals, the camera records new frames. From the following frames, the difference is estimated. Visual Flow The optical flow field is estimated and calculated using this method with an optical flow algorithm. Then, to improve it, a local mean algorithm is employed. A self-adaptive algorithm is used to filter the noise. It is useful in avoiding time-consuming and difficult preprocessing methods and offers a wide range of adaption to the quantity and size of the items. Background Exclusion A quick method for locating moving objects in a video taken by a stationary camera is background subtraction (BS). An elaborate vision system has this as its first stage. This kind of image processing divides the background from the foreground object sequentially.

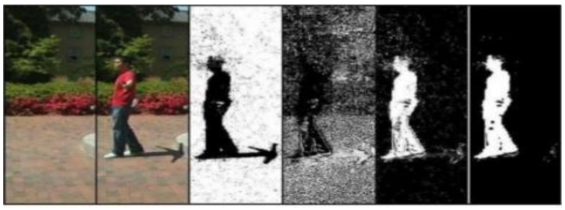

Fig 2: Detection of human from background subtraction

Detection of human from background removal is shown in Figure 6. The image's foreground or subject is recognised and distinguished from the background for additional preprocessing. The separation effect is demonstrated step by step, and then the region of interest is localised.

B. *Object Tracking*

The goal is to track an object's route and speed through video feeds from surveillance cameras and other security systems.

By using object tracking and running classification in a small number of frames taken over a set period of time, the rate of real-time detection can be enhanced.
When searching for objects to lock onto, object detection may proceed at a sluggish frame rate. Once those items are found and locked,object tracking may proceed at a quicker frame rate.

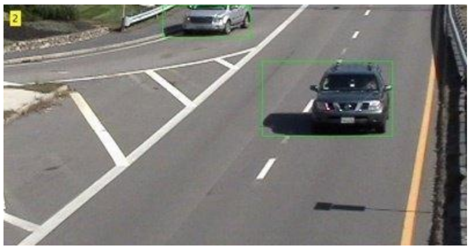

Fig 3: Tracking of car

Fig.3 displays the tracking of the automobile. The object in the aforementioned example can be monitored in two different ways: (1) Tracking in a sequence of detection. With this technique, a CCTV video sequence of moving traffic is captured. In this scenario, if someone wanted to monitor the movement of a car or a person, he would take various pictures or frames at various time intervals. One can target an object, such as a car or a person, with the use of these photographs. Next, by examining how my object has changed during the video's various frames,I can follow it.
The displacement of the object can be verified using various frames captured at various time intervals, and the velocity of the object can then be determined.
This approach actually has a drawback because it relies on detecting the object rather than tracking it throughout time.
"Detection using dynamics" is an improved method.
This method involves estimating the trajectory or movement of the car.
By determining its location at a specific moment, let's say "t," and predicting its location at a later time, let's say "t+10," an accurate image of the car at "t+10" can be suggested with the use of estimation.

iv. SIMULATION RESULTS AND ANALYSIS

A Python programme was created for the technique and implemented in OpenCV [5] based on the SSD algorithm. The model was trained on a total of 21 objects and is run by OpenCV in the Ubuntu IDE. After successfully scanning, detecting, and tracking the video sequence that the camera provided, the results are as follows.

Fig 4: Detection of Bicycle with confidence level of 99.49%

Fig 5: Detection of Train with confidence level of 99.99%

Fig With confidence levels of 99.49%, 98.68%,99.99%, and 8 to 11 respectively, real-time detection of a bicycle, bus, train, and dog is shown, 97.77% of the time.

The model was trained to identify 21 classes of things with an accuracy of 99%, including bicycle, dog, motorbike, human, potted plant, bird, car, cat, sofa, sheep, bottle, and chair.

v. CONCLUSION AND FUTURE SCOPE

Now a days we're seeing that object findings operations are getting popular in utmost of the fields, so based on that perception we 've developed the press grounded operation 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 literacy because we're labeling the data using the LabelImg. So then we're using the YOLO to give faster results, we are using the rearmost interpretation for accurate affair.In real-time circumstances, objects are recognised using the SSD method. Additionally, SSD has demonstrated findings with a high degree of confidence. The primary goal of the SSD algorithm is to identify numerous objects in a real-time video sequence and to follow them. The trained item produced good detection and tracking results, and this model can be used in other contexts to find, follow, and react to the targeted objects in the video surveillance. This ecosystem analysis in real time, which enables security, order, and utility for any organisation, can produce excellent outcomes. Increasing the scope of the investigation to look for weapons and ammunition to raise the alarm in the event of a terrorist attack. The model can be used in CCTVs, drones, and other monitoring equipment to find attacks on numerous locations where weapons are strictly prohibited, including as schools, offices of the government, and hospitals.

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]Hei Law, Yun Teng, Olga Russakovsky, and Jia Deng. CornerNet-Lite: Efficient keypoint based object detection. arXiv preprint arXiv:1904.08900, 2019.

[4] Vinod Nair and Geoffrey E Hinton. Rectified linear units improve restricted boltzmann machines. In Proceedings of International Conference on Machine Learning (ICML), pages 807–814, 2010.

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

[6] YOLO Juan Du1,”Understanding of Object Detection Based on CNN Family”,New Research, and Development Center of Hisense, Qingdao 266071, China.

[7]Dumitru Erhan, Christian Szegedy, Alexander Toshev, “Scalable Object Detection using Deep Neural Networks”, The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 2147-2154.