A Practice for Object Detection Using YOLO Algorithm

Dr. Suwarna Gothane

Associate Professor, Department of Computer Science and Engineering, CMR Technical Campus, Hyderabad, Telangana, India

ABSTRACT

When we look at images or videos, we can easily locate and identify the objects of our interest within moments. Passing on this intelligence to computers is nothing but object detection - locating the object and identifying it. Object Detection has found its application in a wide variety of domains such as video surveillance, image retrieval systems, autonomous driving vehicles and many more. Various algorithms can be used for object detection but we will be focusing on the YoloV3 algorithm. YOLO stands for "You Only Look Once". The YOLO model is very accurate and allows us to detect the objects present in the frame. YOLO follows a completely different approach. Instead of selecting some regions, it applies a neural network to the entire image to predict bounding boxes and their probabilities. YOLO is a single deep convolutional neural network that splits the input image into a set of grid cells, so unlike image classification or face detection, each grid cell in YOLO algorithm will have an associated vector in the output that tells us if an object exists in that grid cell, the class of that object, the predicted bounding box for that object. The model here is progressive so it learns more over time, increasing its prediction accuracy over time. The way the model works is that it makes many predictions in one frame and decides to use the most accurate prediction, thus discarding the other. The predictions are made randomly, so if the model feels like there is an object in the frame which is of a very small pixel it will take that also into consideration. To make it more precise and clearer, the model simply creates bounding boxes around everything in the frame, it would make predictions for each box and pick the one with the most confidence score. All this is done in a small-time frame, thus showing why this specific model is the best to use in a real time situation.

Keywords : Object Detection, YOLO Algorithm, Prediction.

I. INTRODUCTION

"Object Detection" is a dynamic application. The main disadvantage in the existing system R-CNN i.e., Region Based Convolutional Neural Networks where we need to classify huge numbers of regions for the detection. Existing System are outdated, for object detection and resource consuming. It can't be

implemented real time as it takes around 47 seconds for each test proposal. Requirements are far greater than solutions available to store the feature map of the region proposals. So it takes a lot of time to train. To avoid all these limitations and allows to implement it in real world, the system needs to be replaced with a better system. Proposed system work with 2000 regions only which are generated using selective search algorithm. The aim of the proposed system is to eliminate the time consumption. All the previous object detection algorithms have used regions to localize the object within the image. The network does not look at the complete image. Instead, parts of the image which has high probabilities of containing the object. YOLO or You Only Look Once is an object detection algorithm much is different from the region – based algorithms. The system is very simple in design and to implement. The system requires very low system resources and the system will work in almost all configurations. Yolo Object Detection and Open CV allows the user to determine the functionality of the application.

II. LITERATURE SURVEY

Rodrigo Verschae and Javier Ruiz-del-Solar obtained an overview of past research on object detection, outline the current main research directions, and discuss open problems possible future directions. Baohua Qiang et al., proposed an object detection algorithm by jointing semantic segmentation (SSOD) for images. Author constructed a feature extraction network that integrates the hourglass structure network with the attention mechanism layer to extract and fuse multi-scale features to generate high-level features with rich semantic information. Second, the semantic segmentation task is used as an auxiliary task to allow the algorithm to perform multi-task learning. Finally, multi-scale features are used to predict the location and category of the object. algorithm substantially enhances object detection

performance and consistently outperforms other three comparison algorithms, and the detection speed can reach real-time, which can be used for real-time detection. Zhong-Qiu Zhao et al., provided a review on deeplearning based object detection frameworks with Convolutional Neural Network(CNN) focused on typical generic object detection architectures along with some modifications and useful tricks to improve detection performance. As distinct specific detection tasks exhibit different characteristics, survey Experimental analyses several specific tasks, including salient object detection, face detection and pedestrian detection.

Christian Szegedy Alexander and Toshev Dumitru Erhan addressed the problem of object detection using DNNs, Deep Neural Networks classified and also precisely localizing objects of various classes. Author presented simple and powerful formulation of object detection as a regression problem to object bounding box masks using a multi-scale inference procedure which is able to produce high-resolution object detections at a low cost by a few network applications. State-of-the-art performance of the approach is shown on Pascal VOC.

III. PROPOSED APPROACH

A. Proposed Architecture
Architecture describes how the application is going to function. Project Architecture of Object Detection describes how the user's request is taken as input and how the output is delivered. The detailed architecture is shown in Fig 1- Project Architecture of Object Detection
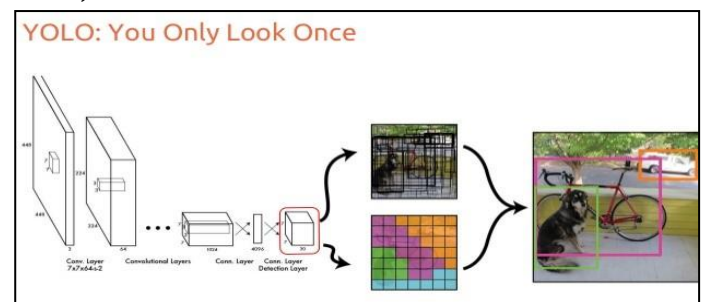
Fig 1. Project Architecture of Object Detection

B. Modules Description

Two main Modules used in the product are User and YOLO USER module perform two functionalities- Run program: This module allows the user to run the program by giving an input video or image file. Attain Feedback: This module allows the user to receive the output in the form of video or image file saved in certain destination. YOLO module perform four functionalities- Object Registration: This module allows the object to be registered with the model. Object Detection: This module allows the product to detect the object based on the training received. Frame Analysis: This module allows the model to analyze the capture frame. Generate Prediction: This module allows the module to generate a text prediction of what the object is and where the object is present Workflow is stated with use case diagram. In the use case diagram, we basically have two actors namely: the User, System with YOLO. The user has the following methods, receive instructions and give requests. The User perform action to enroll the input file and retrieve the output file .System with YOLO will generate the output file by using the COCO dataset which is defined inside. The detailed usecase is shown in Fig 2- Use Case Diagram for User and System for VLX–A Classifieds Platform.
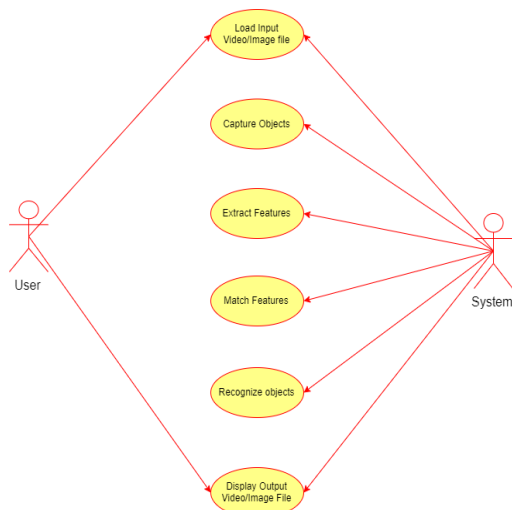
Fig-2. Use Case Diagram for User and System for VLX – A Classifieds Platform

Flow of activities to carry out is shown in Fig 3- Activity Diagram for VLX – A Classifieds Platform.
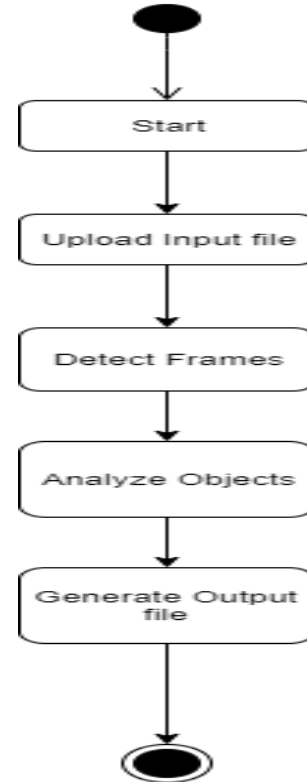
Fig-3. Activity Diagram for VLX – A Classifieds Platform

IV. IMPLEMENTATION

For realistic execution we are using Operating System Multi Platform (Windows 7 & above, Linux GCC),and Backend as Python 3.6 & above. Dataset as COCO (Common Objects In Context)and Machine Learning Model YOLO V3 (You Only Look Once). Using this application, we can detect the objects and specify them. In order to use the application the user has to run the application and can upload a video file or image file to the program by giving the file path. It is designed to detect many objects with the specification. It can easily find out common objects such as chairs, remotes, bus etc. The application gives a glimpse, of where the object is located with the

accuracy. The core features of this project are it provides feedback in the form of video file or image file, it detects most of the common objects in context. The other features include detection of each image is reported with some form of pose information. For example, for face detection in a face detector system compute the locations of the eyes, nose and mouth, in addition to the bounding box of the face.
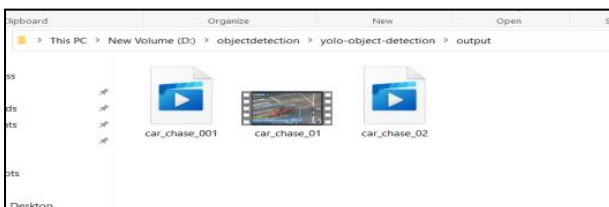
A. Input File Location

Input video files storage location is shown in Fig-4

Fig-4. Input file location

B.Output File Location

The output video files are stored after detecting the objects is shown in Fig-5

Fig-5. Output file location
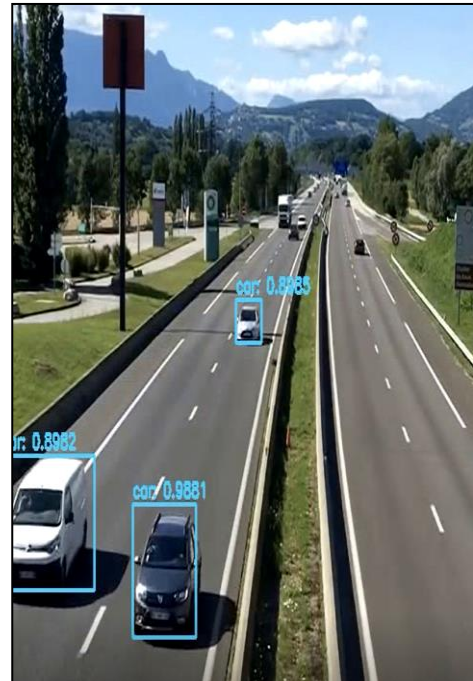
C. Video Frame During Object Detection

The video frame looks when system detect the required object is shown in Fig-6

Fig -6. Video frame during object detection

D. Object Detection

Video Frame looks when it recognizes an object is shown in Fig-7

Fig-7. Object detection

E. Output Generating Command

The command used to run the program and to get the output file is shown in Fig-8

Fig- 8. Output Generating command

V. CONCLUSION & FUTURE ENHANCEMENTS

Approach helps in increasing the accuracy and speed and achieves the desired results. By using method, we are able to detect object more precisely and identify the objects individually with exact location of an object in the picture in x, y axis. implementations of the YOLO algorithm on the web using Darknet is one open-source neural network framework. Darknet was

written in the C Language, which makes it really fast and provides for making computations on a GPU, essential for real-time predictions. The object detection system can be applied in the area of surveillance system, face recognition, fault detection, character recognition etc.

In future we can add a faster model that runs on the GPU and use a camera that provides a 360 field of view and allows analysis completely around the person. We can also include a Global Positioning System and allow the person to detect the objects instantly without any delay in frames and seconds.

VI. REFERENCES

[1]. Rodrigo Verschae, Javier Ruiz-del-Solar, "Object Detection: Current and Future Directions Perspective", Article in Frontiers in Robotics and AI , December 2015.

[2]. Baohua Qiang, Ruidong Chen, Mingliang Zhou, Yuanchao Pang, Yijie Zhai,Minghao Yang,"Convolutional Neural Networks-Based Object Detection Algorithm by Jointing Semantic", Segmentation for Images, Sensors 2020.

[3]. Zhong-Qiu Zhao, Shou-tao Xu, and Xindong Wu, "Object Detection with Deep Learning: A Review", IEEE Transactions on Neural Networks and Learning Systems for Publication, April 2019.

[4]. Christian Szegedy Alexander, Toshev Dumitru Erhan, "Deep Neural Networks for Object Detection", NIPS 2013.

[5]. https://www.analyticsvidhya.com/blog/2018/12/practical-guide-object-detection-yolo-framewor-python