International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

MULTIPLE OBJECT DETECTION IN IMAGES USING YOLO

Garrepalli Gnaneshwari[1] , Goshika Ashritha[2], Guda Srisaipriya[3]
,Dr.B.V.Ramana Murthy[4]

Department of Computer Science and Engineering, Stanley College of Engineering and Technology for Women, Telangana, India

Abstract Object detection is a technology that includes computer vision that deals with detecting instances of objects of a certain class(Such as humans,animals,cars etc.) in images.Object detection has been attracting much interest due to the wide spectrum of applications that use it. Object detection technology has been driven by an increasing processing power available in software and hardware. In this work we present a developed application for multiple object detection based on OpenCV libraries. The complexity-related aspects that were considered in the object detection using YOLO. The proposed application deals with real time systems implementation and the results give an indication of where the cases of object detection applications may be more complex and where it may be simpler.Detection is a process to identify and detects object in images.This project is designed to resemble the way a human brain function.It helps us to detect the objects present in one particular image and also tell us what that object with the help of bounding boxes and name label of that object along with probability score.

Keywords: OpenCV, Convolutional Neural Network, Bounding Boxes, YOLO

1.Introduction

1.1 About Paper

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 applications 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.

YOLOv3 (You Only Look Once, Version 3) is a real-time object detection algorithm that identifies specific objects in videos, live feeds, or images. YOLO uses features learned by a deep convolutional neural network to detect an object.As typical for object detectors, the features learned by the convolutional layers are passed onto a classifier which makes the detection prediction. In YOLO, the prediction is based on a convolutional layer that uses 1×1 convolutions.

1.2 Objective of Paper

The Objective is to detect objects in images using the You Only Look Once (YOLO) approach. This method has several advantages as compared to other object detection algorithms. In other algorithms like Convolutional Neural Network, Fast Convolutional Neural Network the algorithm will not look at the image completely but in YOLO the algorithm looks the image completely by predicting the bounding boxes using convolutional network and the class probabilities for these boxes and detects the image faster as compared to other algorithms.

- When we see or show an image, our brain instantly recognizes the objects contained in it.● On the other hand , it takes a lot of time and training data for a machine to identify these objects, But with the recent advancement in hardware and deep learning, this computer vision field has become a whole lot easier and more intuitive .
- Object detection is a technology that falls under the broader domain of computervision.
- It deals with identifying the objects in images . It has multiple applications such asface detection, vehicle detection, pedestrians, cars etc..
- The main objective of this project is to identify and detect the objects present in the given image.

- The human visual system is fast and accurate and can also perform complextasks like identifying multiple objects and detect obstacles with little conscious thought.
- With available datasets, we can now easily train computers to detect andclassify multiple objects within an image with high accuracy .
- The aim of object detection is to detect all instances of object from a knownas class.
- Object recognition is to describe a collection of related computer vision tasksthat involves activities like identifying objects in digital photographs.
- Image classification involves activities such as predicting the class of one object in an image.
- Object localization is referring to identifying the location of one or more objectsin an image drawing a bounding box around the extent .
- Object detection does the work of combines these two tasks and localizes andclassifies one or more objects in an image.

1.3 Scope of the paper

Object detection is a computer vision technique that allows us to identify and locate objects in an image or video. With this kind of identification and localization, object detection can be used to count objects in a scene and determine and track their precise locations, all while accurately labeling them.The main aim of this project is to recognize multiple objects in images by training the model.

2.Literature Survey

2.1 Existing System

- Earlier we don't have much gpu resources and computational resources .
- The pipeline of the Traditional object detection has been divided in to three stages they are :

✣ **Informative Region Selection**

As different objects may appear in any position of the image and have different aspect ratios or sizes, it is a natural choice to scan the whole image with a multi-scale sliding window . Although this exhaustive strategy can find out all possible positions of the object s , its shortcomings are also obvious . Due to a large number of candidate windows , it is computationally expensive and produces too many redundant windows . However, if only a fixed number of sliding window templates are applied, unsatisfactory regions may be produced .

✣ **Feature Extraction**

To recognize different objects, we need to extract visual features which can provide a semantic and robust representation . SIFT, HOG,and haar-like features are the representative ones . This is due to the fact that these features can produce representations associated with complex cells in the human brain . However due to the diversity of appearances , illumination condition and background , it's difficult to manually design a robust feature descriptor to perfectly describe all kinds of objects .

✣ **Classification**

Besides, a classifier is needed to distinguish a target object from all the other categories and to make the representations more hierarchical, semantic and informative for visual recognition .

➢ **Traditional object detection Methods**

- Scale-invariant Feature Transform (SIFT)
- Haar Features
- Histogram of Oriented Gradients (HOG)
- Thes previous models are not able to detect a real time detection application .

Generic Object Detection

Generic object detection aims at locating and classifying existing objects in any one image , and labeling them with rectangular bounding boxes to show the confidences of existence . The framework of generic object detection methods can mainly be categorized in two types. One follows the traditional object detection pipeline , generating region proposals at first and then classifying each proposal into different object categories . The other regards object detection as a regression or classification problem , adopting a unified framework to achieve final results (categories and locations) directly .

Region Proposals

It is composed of three correlated stages, including regional proposal generation , feature extraction with CNN, classification and bounding boxes regression , which are usually trained separately .

R-CNN

To circumvent the problem of selecting a huge number of regions, Ross Girshick et al. proposed a method where we use the selective search for extract just 2000 regions from the image and he called them region proposals. Therefore, instead of trying to classify the huge number of regions, you can just work with 2000 regions. These 2000 region proposals are generated by using the selective search algorithm which is written below.

Selective Search

1. Generate the initial sub-segmentation, we generate many candidate regions
2. Use the greedy algorithm to recursively combine similar regions into largerones .
3. Use generated regions to produce the final candidate region proposals

These 2000 candidate regions which are proposals are warped into a square and fed into a convolutional neural network that produces a 4096-dimensional feature vector as output. The CNN plays a role of feature extractor and the output dense layer consists of the features extracted from the image and the extracted features are fed into an SVM for the classification of the presence of the object within that candidate region proposal. In addition

to predicting thepresence of an object within the region proposed , the algorithm also predicts four values which are offset values for increasing the precision of the bounding box. For example, given the region proposal, the algorithm might have predicted the presence of a person but the face of that person within that region proposal could have been cut in half. Therefore, the offset values which are given help in adjusting the bounding box of the region proposal.

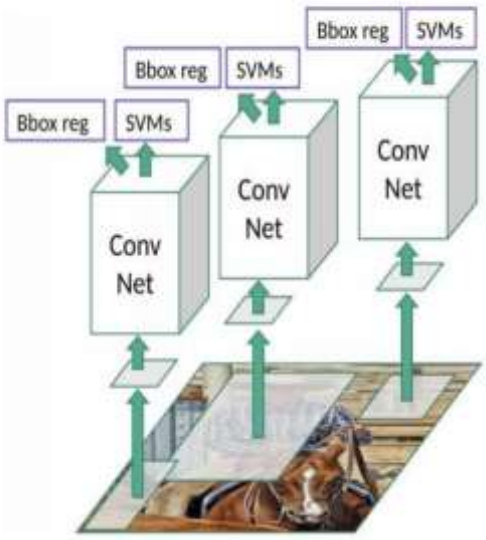


Fig 2.1.1 : R-CNN

Problems with R-CNN :

- ➢ It takes a huge amount of time to train the network .
- ➢ It cannot be implemented real time as it takes around 47 seconds for each test image .
- ➢ The selective search algorithm is a fixed algorithm. Therefore, no learning is happening at that stage. This could lead to the generation of bad candidate region proposals .
- ➢ It still takes a huge amount of time to train the network as you would have to classify 2000 region proposals per image.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

YOU ONLY LOOK ONCE (YOLO)

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 have 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 which are seen above. In YOLO a single convolutional network predicts the bounding boxes and the class probabilities for these boxes. YOLO works by taking an image and splitting it into an SxS grid, within each of the grid we take m bounding boxes. For each of the bounding boxes, the network gives an output a class probability and offset values for the bounding box. The bounding boxes have the class probability above which a threshold value is selected and used to locate the object within the image.YOLO is orders of magnitude faster(45 frames per second) than any other object detection algorithms.and based on the class probability map then it will produce final detection .

The limitation of the YOLO algorithm is that it struggles with the small objects within the image, for example, it might have difficulties in identifying a flock of birds. This is due to the spatial constraints of the algorithm.

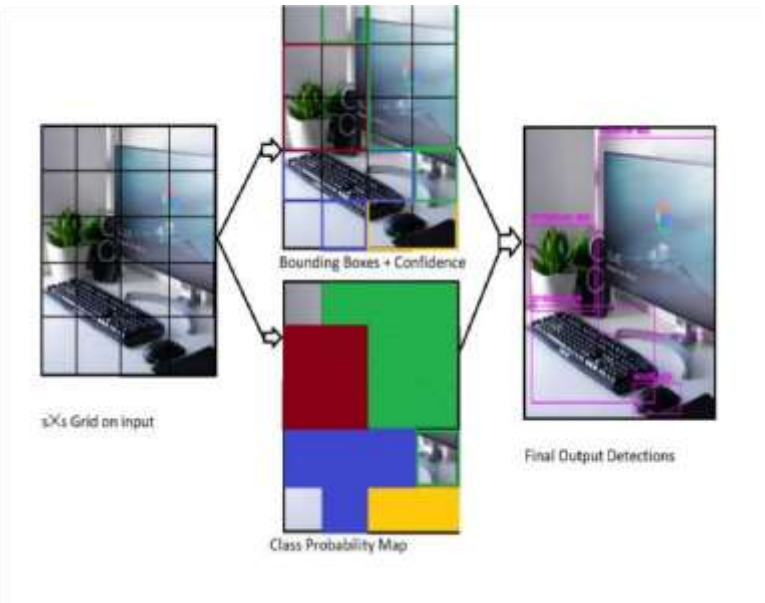


Fig.2.1.2: YOLO Model

It has 24 convolutional layers followed by two fully connected layers. Then we relate these layers on DarkNet classification as half of the resolution as 224✕224 input image and then double the resolution for detection and by using Non-Max Suppression it will remove all thebounding boxes for threshold less than 0.2 so that we can see bounding boxes clearly.

2.1.1Drawbacks of the Existing system

It struggles to detect an object in some rare cases and some compatibility issues that arise , in order to overcome it ,a proposed system has existed to execute a perfect solution without any errors .

2.2 Proposed System

The main aim is to improve the algorithm and work it with a faster FPS (Frames Per Second) Using Online CPUs/GPUs and to use it in real-time scenarios. First, import the necessary libraries in a python file. Then by using the OpenCV ,it detects the object feed via image and loads the class files .

Object detection applications are easier to develop than ever before . Besides significant performance improvements , these techniques have also been leveraging massive image datasets to reduce the need for large datasets .

In addition, with current approaches focussing on full end-to-end pipelines performance has also improved significantly, enabling real -time use cases SSD is widely used for different types of application .

Some applications need accuracy and speed at the same time in order to achieve the main objective . This can be getting more data , inventing or creating more data , re-scaling the data , transforming the data or by feature selection .

Further , these models will be optimized by tuning the algorithm .

coco names,model configuration file, and model weight file. Then define the confidence threshold for detecting an image with the desired accuracy and non-max suppression to reduce the overlapping of the bounding boxes. Next, by defining a function to find the objects, if found the objects then it will be sent to the Bounding boxes function else it will try to detect the object. In the Bounding Box function, It will get x,y,w,h then by using the cv.rectangle function we will draw the boxes for every object it detects. By using the cv.putText function for labeling the rectangle.

3.Proposed Architecture

3.1 About YOLO Architecture

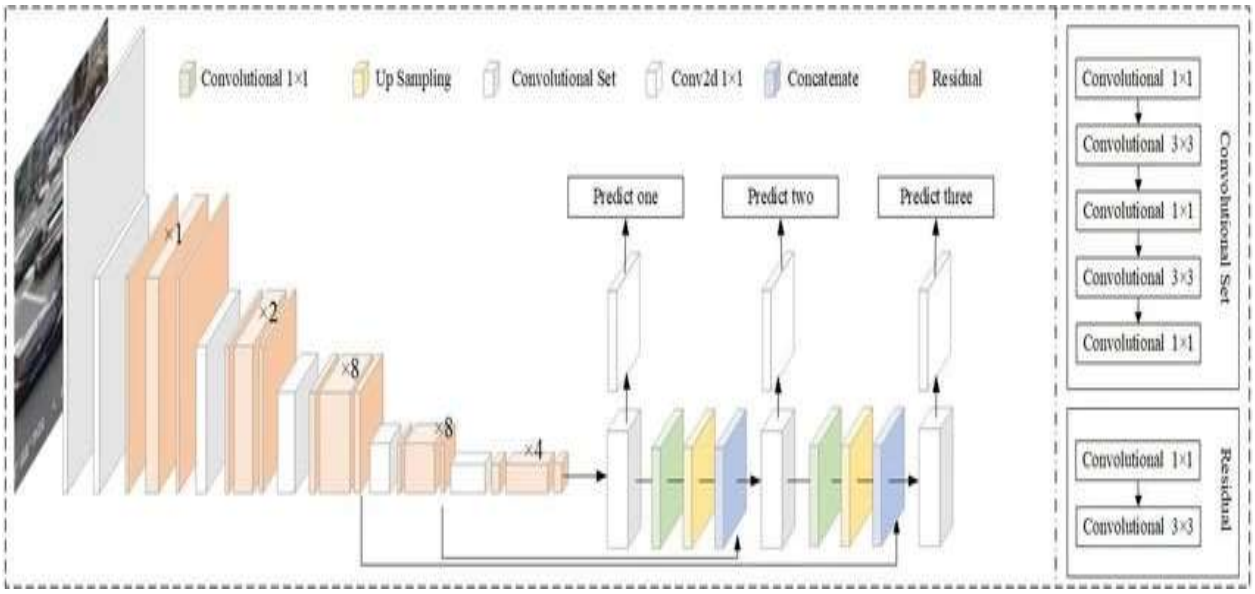
The YOLOv3 algorithm first separates an image into a grid. Each grid cell predicts some number of boundary boxes (sometimes referred to as anchor boxes) around objects that score highly with the aforementioned predefined classes.

Each boundary box has a respective confidence score of how accurate it assumes that prediction should be and detects only one object per bounding box. The boundary boxes are generated by clustering the dimensions of the ground truth boxes from the original dataset to find the most common shapes and sizes.

Other comparable algorithms that can carry out the same objective are R-CNN (Region-based Convolutional Neural Networks made in 2015) and Fast R-CNN (R-CNN improvement developed in 2017), and Mask R-CNN.

YOLOv2 used a feature extractor known as the Darknet-19, which consisted of 19 convolutional layers. The newer version of this algorithm, YOLOv3 uses a new feature

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



extractor known as Darknet-53 which, as the name suggests, uses 53 convolutional layers while the overall algorithm consists of 75 convolutional layers and 31 other layers making it a total of 106 layers [36]. Pooling layers have been removed from the architecture and replaced by another convolutional layer with stride '2', for the purpose of down-sampling .

Fig 3.1.1 Architecture

However, unlike systems like R-CNN and Fast R-CNN, YOLO is trained to do classification and bounding box regression at the same time.

There are major differences between YOLOv3 and older versions in terms of speed, precision, and specificity of classes.

YOLOv3 algorithm trained on COCO dataset which has vast classes and bounding boxes and final detection with a feature map , which is responsible for the detection of objects at particular scale . As a result, YOLOv3 is better at detecting smaller objects when compared to its predecessors YOLOv2 and YOLO.

4.IMPLEMENTATION

4.1 Algorithm

Step 1: Download the weights and configuration files from the following

link : https://pjreddie.com/darknet/yolo/

Step 2: Download the weights and cfg files of YOLO which are named as yolov3.weights and yolov3.cfg.

Step 3: Importing Libraries and Setting Paths.

Step 4: Download coco.names file from the same above link.Once all files are downloaded place them in the project directory.

Step 5: Defining the input for reading class names from coco.names file and appending to the list named classes.

Step 6: Load YOLOv3 Model,Convert the images before feeding to YOLONet or DarkNet and to feed this blob to YOLO network

Step 7:YOLO uses Non-Maximal Suppression (NMS) to only keep the best bounding box.

Step 8: Then this network divides that image into regions which provides the bounding boxes and also predicts probabilities for each region.

Step 9: The output will generate bounding boxes weighted by the predicted probabilities.

4.2 Code implementation

4.2.1. Dataset Collection

Data collection plays the most important role in vehicle classification. Data collection for the mentioned application is achieved with the help of free datasets available on the World Wide Web.

Public datasets contain collections of data for machine learning, some containing millions of data points and an immense amount of annotations that can be re-used for training or fine-tuning AI models. Compared to creating a custom data set through collecting video data or images, it's much faster and cheaper to use a public dataset. Using a fully prepared dataset is favorable if the detection task involves common objects (people, faces) or situations and is not highly specific. The accuracy depends on the dataset sample images used for training different types of vehicles. In the current application.Some datasets are created for specific computer vision tasks such as object detection, facial recognition, or pose estimation. Hence, they may be unsuitable to use for training your own AI models to solve a different problem. In this case, the creation of a custom dataset is required. We have collected data from coco
Datasets .

What is COCO Dataset?

Common Objects in Context (**COCO**) Common Objects in Context (**COCO**) is a database that aims to enable future research for object detection, instance segmentation, image captioning, and person keypoints localization.COCO is a large-scale object detection, segmentation, and captioning dataset.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Fig 4.2.1.1: Sample COCO dataset

4.2.2 Feature Extraction

Feature extraction refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original data set. It yields better results than applying machine learning directly to the raw data. Feature extraction techniques are used to solve common computer-vision problems like object detection and recognition .

Parameters :

A blob is a 4D numpy array object (images, channels, width, height).It has the following parameters:

- the **image** to transform
- the **scale** factor (1/255 to scale the pixel values to [0..1])
- the **size**, here a 416x416 square image
- the **mean** value (default=0)
- the option **swapBR=True** (since OpenCV uses BGR)

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

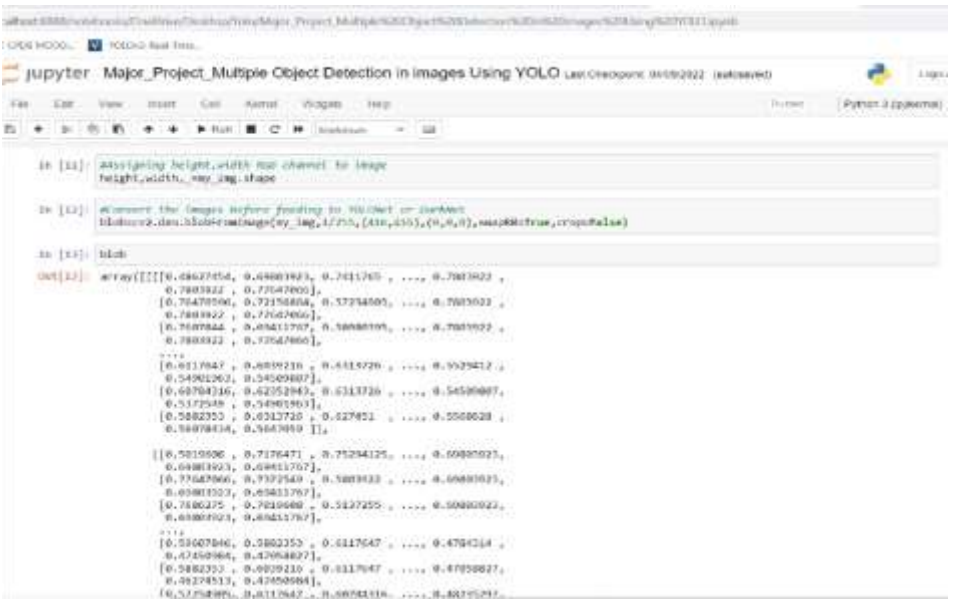


Fig 4.2.2.1: Blob

4.2.3 Training

Image classification is based on the training of the collected data set. For the current application, the training is performed on images that are taken from the internet, and wherever needed, supplemented with few capture sequences of our own.

The training was done using Jupyter notebook with help of Anaconda ,for faster and efficient training of the network. After preprocessing the dataset i.e. creating label file for each image, both images and their respective label files are to be kept together. The yolo.cfg file was used for training configurations which include three yolo layers.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

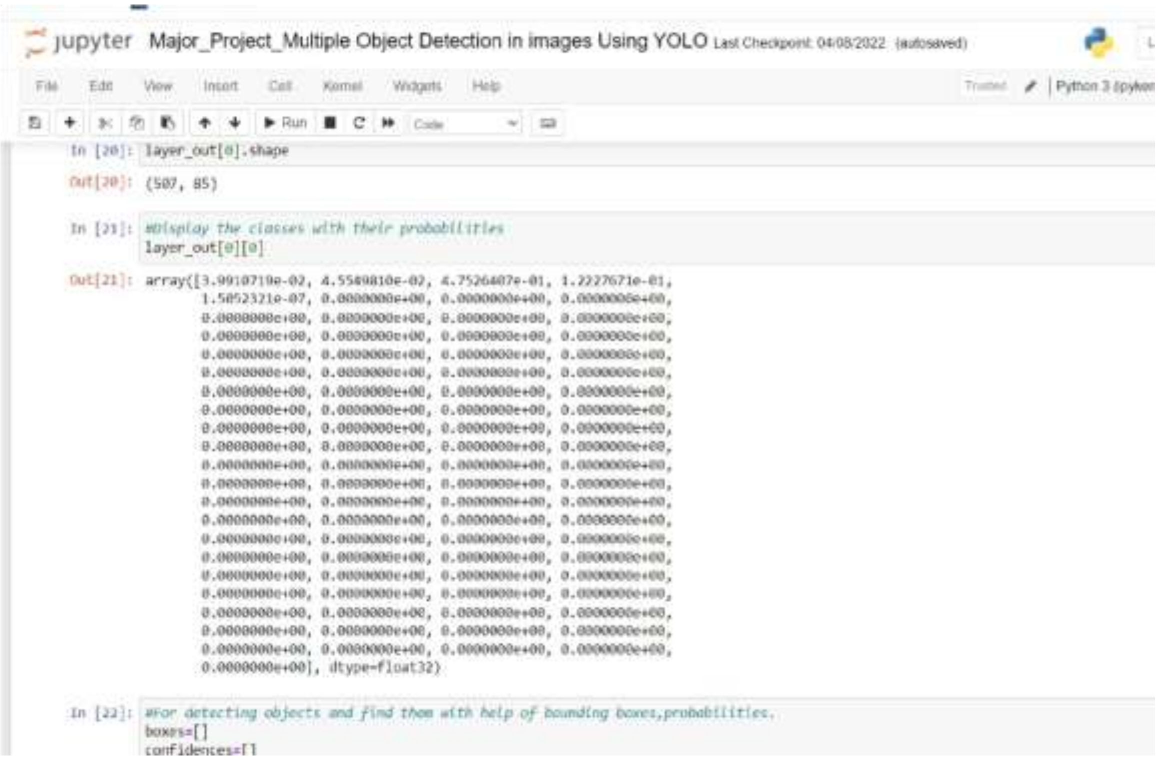


Fig 4.2.3.1: Training Data

OpenCV

OpenCV is a huge open-source library for computer vision, machine learning, and image processing. OpenCV supports a wide variety of programming languages like Python, C++, Java, etc. It can process images and videos to identify objects, faces, or even the handwriting of a human.

Matplotlib

Matplotlib is one of the most popular Python packages used for data visualization. It is a cross-platform library for making plots from data in arrays.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Numpy

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays. ***5.Result***

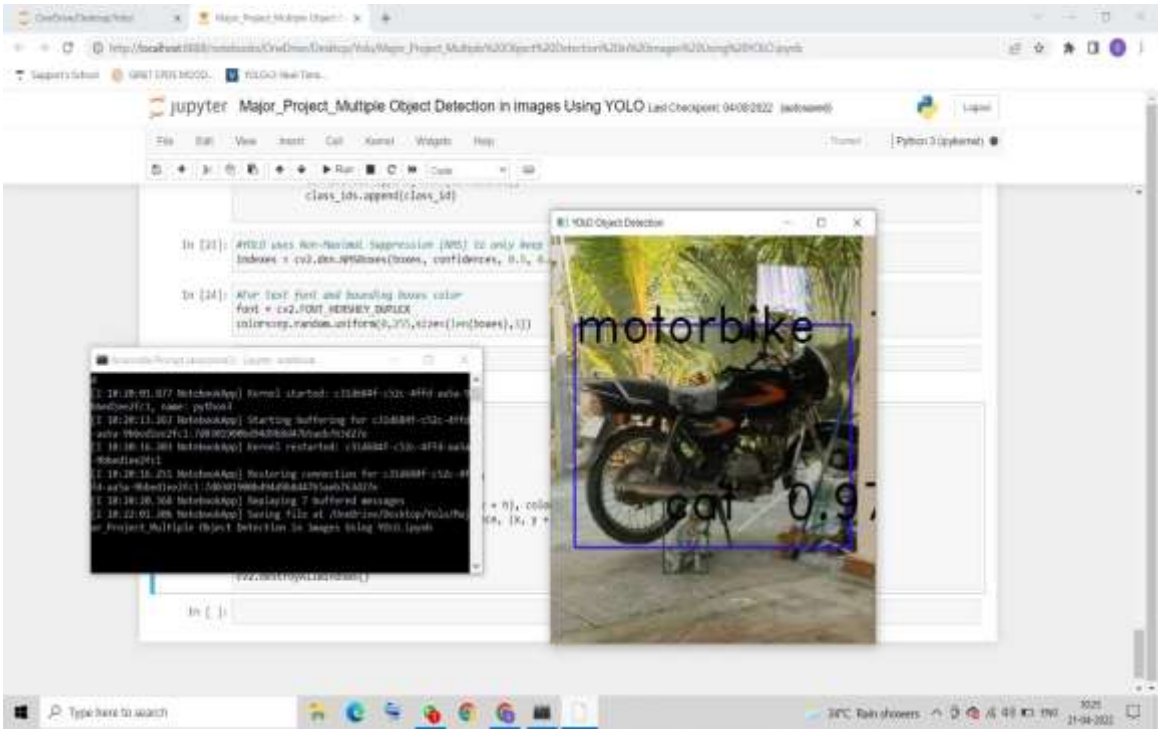


Fig 5.1: Result

Fig 5.2: Detection Of Objects

Some of the real time images that is detecting the objects in images :

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

Fig 5.3: Object Detection output-1

Fig 5.4: Object Detection Output-2

6.Conclusion

In this paper, we proposed the YOLO algorithm for the purpose of detecting objects.This algorithm is generalized, it outperforms different strategies once generalizing from natural pictures to different domains. The algorithm is simple to build and can be trained directly on a complete image.Region proposal strategies limit the classifier to a particular region. YOLO accesses the entire image in predicting boundaries. And also it predicts fewer false positives in background areas.Comparing to other classifier algorithms this algorithm is much more efficient and fastest algorithm to use in real time.

By using this project and based on experimental results we are able to detect objects more precisely and identify the objects individually with the exact location of an object in the picture in the x,y axis.

7.Future Scope

With some of the main useful applications of object detection:

- Vehicle's Plates recognition,self-driving cars, Tracking objects, medical imaging, object counting, object extraction from an image or video, person detection.
- To introduce the system that serves as blind aid(Transform the visual world in to the audio world with the potential to inform blind people objects as well as their spatial locations .
- The future of object detection technology is in the process of proving itself, and much like the original Industrial Revolution, it has the potential, at the very least, to free people from tedious jobs that will be done more efficiently and effectively by machines. It will also open up new avenues of research and operations that will reap additional benefits in the future. Thus, these challenges circumvent the need for a lot of training requiring a massive number of datasets to serve more nuanced tasks, with its continued evolution, along with the devices and techniques that make it possible, it could soon become the next big thing in the future.

8.References

1. [1]YOLOv3, accredited paper on the third version of YOLO: Redmon, Joseph, and Ali Farhadi. "YOLOv3: An Incremental Improvement.

2. R. Huang, J. Pedoeem, and C. Chen, "YOLO-LITE: A Real-Time Object Detection

3. Algorithm Optimized for Non-GPU Computers," IEEE Int. Conf. Big Data, Big Data, pp. 2503–2510, 2019, doi: 10.1109/BigData.2018.8621865.

4. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, realtime object detection," IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit., vol. 2016-Decem, pp. 779–788, 2016, doi: 10.1109/CVPR.2016.91.

5. [4]YOLO Juan Du1,"Understanding of Object Detection Based on CNN Family",New
6. Research, andDevelopment Center of Hisense, Qingdao 266071
7. [5]Andrew Ng YOLO explanation Matthew B. Blaschko ChristophH. Lampert,"Learning to Localize Objects with Structured Output Regression", Published in Computer Vision – ECCV 2008 pp 2-15.

8. A. S. R. H. A. J. S. S. Carlsson, "CNN Features off the-shelf: an Astounding Baseline for

9. Recognition," Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Work.,vol. 7389, pp. 806–813, 2014, doi: 10.1117/12.827526

10. J. Du, "Understanding of Object Detection Based on CNN Family and YOLO," J. Phys. Conf. Ser., vol. 1004, no. 1, 2018, doi: 10.1088/17426596/1004/1/012029.

11. [8]Khushboo Khurana and Reetu Awasthi,"Techniques for Object Recognition in Images and
12. Multi-Object Detection",(IJARCET), ISSN:2278-1323,4
13. [9]Latharani T.R., M.Z. Kurian, Chidananda Murthy M.V,"Various
14. Object Recognition Techniques for Computer Vision", Journal of Analysis and Computation, ISSN: 0973-2861.
15. [10]Allan Zelener - YAD2K: Yet Another Darknet 2 Keras Official_YOLO_website
16. [11]Md Atiqur Rahman and Yang Wang, "Optimizing Intersection-Over-Union in Neural Networks for Image Segmentation," in Object detection, Department of Computer Science,