International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

Multiple Object Detection Using YOLOv3(you look only once)

Prajwal K[1], Mahendra R[2], Rahul mandal[3], Vidyarani H J[4]

[123]BE Information Science Student, Dr. Ambedkar Institute of Technology
[4]Asst.Professor, Dept. Of ISE, Dr. Ambedkar Institute of Technology

---

Abstract – Object detection technology has been steered by software and hardware with high processing power. It is a growing field of computer vision. Identification and classification of objects, either in single scene or in more than one frame, has huge importance in various fields, like detecting anomalies in agriculture field, facial recognition, self-driving autonomous cars, counting the crowd etc. Proposed system falls under the domain of neural network based deep learning techniques, in which we make use of YOLOv3 (You Look Only Once) algorithm. A single convolutional network concurrently predicts multiple bounding boxes and class probabilities for those boxes and also a fastest image processing algorithm compared to all present algorithms. Darknet-53 technique is used for extracting features. Coco dataset is fed for training model. The GUI is built using Flask.

***Key Words*:** Object detection, Coco dataset, Darknet-53, YOLOv3, Python, Flask.

1. INTRODUCTION

With advancements in technology and development in artificial intelligence object detection has become more popular. Object detection is a technology related to image processing and computer vision that allows system to identify and locate objects that are present in an image or video. Explicitly, object detection draws boundary boxes around these separated objects, which allow us to locate where said objects are in a given frame. Most of the people often get confused between object detection and image recognition, it's very much important to know difference between them. Image recognition assigns a label to a whole image. While object detection draws the bounding box around each object detected and labels the box.

We can say that object detection techniques can be broadly divided into two broad groups. R-CNNs (Region-based Convolutional Neural), is a group of techniques which are helpful for addressing object localization and recognition tasks, which are specifically designed for model performance. YOLO can be recognized as the second group of techniques for object recognition which is prominently designed for speed and real-time use. Accuracy, efficiency and flexibility are the main feature of object detection system. The efficiency of the system depends upon the algorithm and extracting features used. The performance of the system has constrained by data present in the given dataset, quality of the image or video provided. YOLO, in a single peep, takes the entire image as input and predicts bounding box coordinates for the objects and their class probabilities. YOLO's greatest advantage is its ace pace, it is extremely fast, and it can process 45 frames per second. Amongst all the three versions of YOLO, YOLOv2 is fastest, but YOLOv3 is fastest and more accurate in terms of detecting small objects. YOLO object detection algorithm is different from other region-based algorithms. In YOLO a single convolutional network predicts the bounding boxes and the class confidence for the objects. It takes an image and split it into a (S x S) grid, within each of the grid we take m bounding boxes. For each of the bounding box, the network gives an output i.e., a class probability and offset values for the bounding box. The bounding boxes with class probability above a threshold value are selected and used to locate the object within the given image. If the center of an object falls into a particular grid cell, that grid cell is solely responsible for detecting that object. Each grid cell predicts bounding boxes and confidence scores for those boxes. These scores show how confident the model is that the box contains an object and also how accurate it thinks the box has that object it predicts.

The application of object detection is extensive. This system makes the object detection easier and faster. It can be used in traffic light detection, autonomous driving cars, crowd detection in highly populated areas. By using IOT technology this system can be further implemented to wireless cameras so that real time object detection will be much easier.

2. LITERATURE SURVEY

2.1 You Only Look Once: Unified, Real-Time Object Detection

YOLO is a new approach for object detection where prior works on object detection make use of classifiers to perform detection. But in this model, we frame object detection as a regression problem to achieve spatially detached bounding boxes and the class probabilities allied with them. YOLO is a single neural network algorithm which predicts bounding boxes, class names and class probabilities directly from full images just in one assessment. The whole object detection pipeline is a single neural network, in other words we can say there is only one layer of input nodes that sends the outputs to succeeding layers of receiving nodes. [1]

2.2 Object Detection and Identification

Object detection is important for computer vision. The problems such as noise, blurring and rotating jitter, etc. with images in real world have an important impact on object detection. The objects can be detected in real time using YOLO, an algorithm based on convolutional neural networks. This paper addresses the various modifications done to YOLO network which improves the efficiency of object detection. [2]

2.3 Literature Survey on Object Detection Using Yolo

Object detection is important for computer vision. The problems such as noise, blurring and rotating jitter, etc. with images in real world have an important impact on object detection. The objects can be detected in real time using

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 07 | July - 2021 ISSN: 2582-3930

YOLO, an algorithm based on convolutional neural networks. This paper addresses the various modifications done to YOLO network which improves the efficiency of object detection.[3]

2.4 YOLOv3: An Incremental Improvement

YOLOv3 is an updated version of YOLOv2 which predicts an objectness score for each bounding box that is detected while using logistic regression. YOLOv3 is sluggish than YOLOv2, but still faster than all the other models. To increase the accuracy of the model, speed has been compromised.

The model uses k-means clustering algorithm to determine the bounding box priors. If the IOU (intersection over union) increases then the performance of the object detection gradually decreases. So, we can say that the model scuffles to get perfectly aligned bounding boxes. [4]

3. PROPOSED SYSTEM

YOLOv3 is one of the fastest algorithms in object detection. With Darknet-53 feature extractor it has become more accurate.

It has proven that the YOLOv3 performs well in accuracy and speed over other algorithms. We used following methodology, design, tools to build the object detection system.

3.1 Methodology

Step 1: Image Acquisition: Getting input i.e., image of .jpeg, .png, .jpg format from system through GUI and grab its spatial dimensions.

Step 2: Determining output layers: After the reading the image we are going to determine the yolo output layers by using the net.getLayerNames() which gives the output layer of yolo.

Step 3: Construction of blob: A blob is constructed from the input images and a forward pass of the yolo object detector is performed; this step gives us our bounding boxes and the probabilities associated for them.

Step 4: Lopping over each of the output layers: We are going to loop through each and every output layers to get layer outputs and also looping over detection then extracting the classIDs (class name) and confidence (i.e., probability) of current object detection. And then the weak predictions are filtered out by ensuring the detected class confidence is greater than the minimum probability specified by the developer. After deleting the weak predictions, model will scale the bounding box coordinates back relative to the size of the input image, but YOLO actually returns the center (x, y) coordinates of the predicted bounding box followed by the width and height of the box. At last, the model uses the center (x, y) coordinates determine the top and the left corner of the bounding box, then it updates the obtained list of bounding box coordinates, class confidences and class names(classIDs).

Step 5: Apply non matrix suppression: In this step the overlapping weak bounding boxes are removed and ensure that at least one detection exists the loop over the indexes what we are keeping, then extract the bounding box coordinates. After obtaining bounding box coordinates then draw a bounding box rectangle and label on the image by function text=" {}: {:4f}". format ().

Step 6: GUI for displaying output: Getting final output of an image with bounding boxes, object name and confidence.

3.2 System Architecture

The feature extractor used by YOLOv3 is Darknet-53, as mentioned in the name, it uses 53 convolutional layers whereas overall algorithm consists of 106 layers i.e.., 75 convolutional layers and 31 other layers. YOLOv3 architecture is constructed of 3 distinct layer forms

- First layer is the residual layer, in this layer image is down sampled. This layer is formed when the activation functions can be easily forwarded to a deeper layer in the neural network. In this setup outputs of layer 1 are concatenated to the outputs of layer 2.

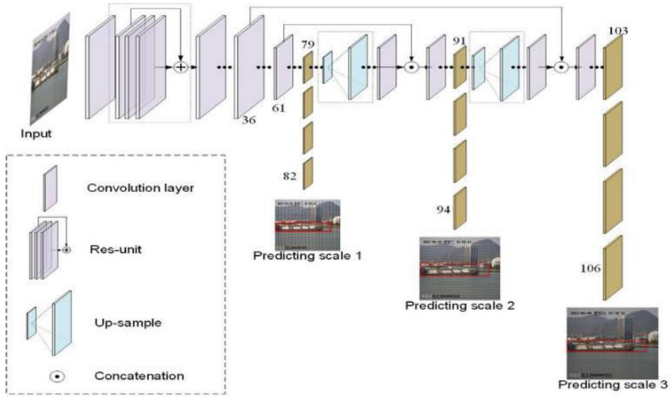


Fig 3.1: System architecture

- Second layer is known as the detection layer. This layer will perform detection in 3 distinct scales. The 1st detection is made by the 82nd layer, 2nd detection is made by 94th layer and final detection is made by 106th layer.
- Third layer is for up-sampling, which's main function is to increases the spatial resolution of an image.

In fig 1 the purpled colored blocks are residual layers; orange colored is the detection layer and the sky-blue colored blocks represents up-sampling layers the image is up sampled before and then it is scaled. concatenation operation is used, to concatenate the outputs one layer to another layer. Addition operation is used to add previous layers.

Fig 3.1 represents the architecture of YOLOv3 algorithm trained with COCO dataset which contains 80 classes and bounding boxes. The kernel size will be 1 x 1 x 255.In YOLOv3, the input image dimensions are down sampled by 32, 16 and 8 to make predictions at scales 3,2 and 1 respectively. If the input image is of size 416*416, the resultant down sampled feature map would be of size 13*13.

3.3 Tools Used

To develop this multiple object detection model, we used Python language in anaconda environment and we made use of many python libraries like OpenCV, NumPy, OS and Flask.

Flask is used to integrate the python code with Html for building an efficient GUI.

4. RESULTS AND DISCUSSION

The results show that the system has performed well against inputs given. For all the elements which are present in datasets we got 100% accuracy. The yolo algorithm works for images with .jpeg .png .jpg formats. Fig-4.1 shows the user interface of object detection system where user chooses the input image which contains object that to be recognized. Fig-4.2 shows the input image fed to as input to the system. Fig-4.3 shows the output image with detected object with bounding boxes and list of detected objects with confidence scores at the left side of the GUI.

4.1 User Interface

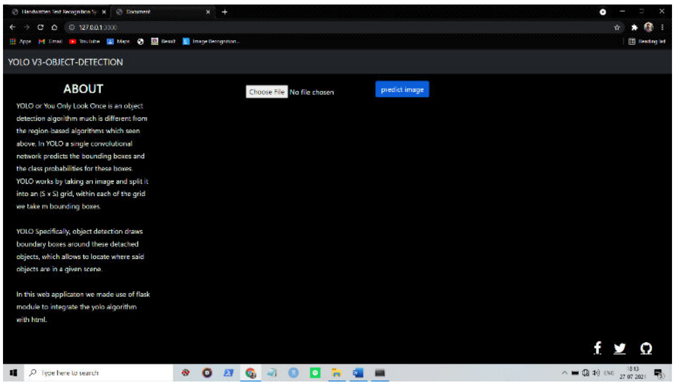


Fig: 4.1 user interface

4.2 Input Image

Fig: 4.2 Input Image

4.3 Output Image and Detected Object List

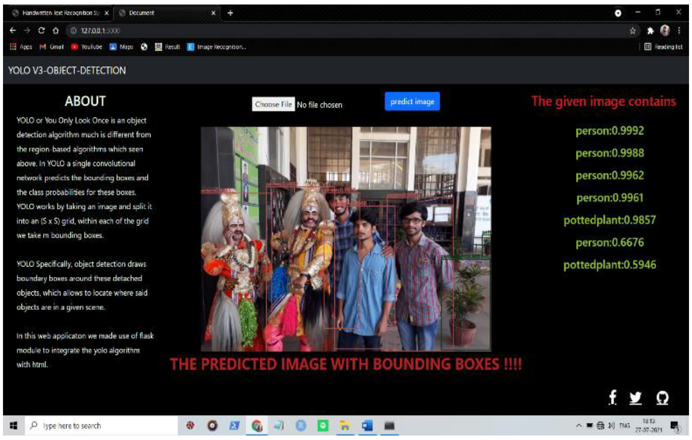
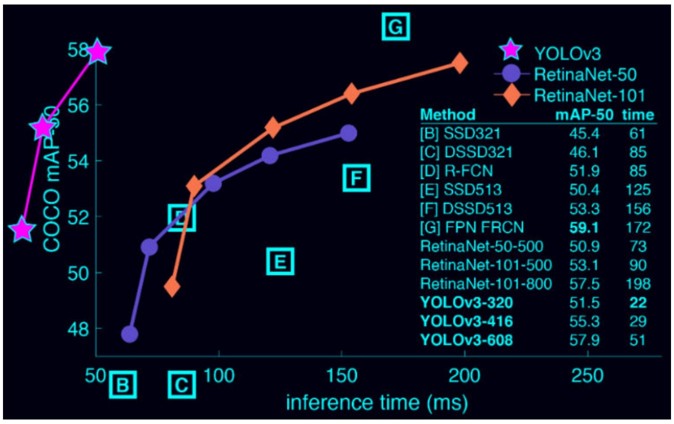


Fig: 4.3 Output Image

4.4 Discussion

The Fig-4.3 shows the efficiency and accuracy of the YOLOv3 system. The image shows the graph plotted for 3 different methods i.e., YOLOv3, RetinaNet-50, RetinaNet-101 with inference time (ms) on X-axis and COCO mAP-50(mean average precision) on Y-axis. YOLOv3 is extremely fast and accurate. In mAP measured at 0.5IOU YOLOv3 in on par with Focal Loss but about 4x faster. The table on the left corner gives us the mAP-50 value and time taken by each method. By looking at all the data present and analysis practically with n number of input and output images we can say that the proposed system is highly efficient and faster in object detection.



5. COCLUSION

We introduced YOLOv3, a unified model for object detection. Our object detection model is easy to develop and implement. It can be trained directly with custom datasets there are plenty of ready datasets available in web. YOLOv3 model is trained on a loss function that directly corresponds to performance of object detection and the whole model is trained jointly.

With all the tests and results given by the model we can conclude that proposed model is one of the accurate, most powerful and fastest object detection model that is present today. YOLOv3 is always the first choice for every real-time identification of objects since its faster and rate of frames per second is higher(67fps).

REFERENCES

[1] Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*.

[2] Viraktamath, S. V., Yavagal, M., & Byahatti, R. Object Detection and Classification using YOLOv3.

[3] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2016)779-788

[4] Vikram, M., A YOLO Based Approach for Traffic Sign Detection. National Institute of Technology Karnataka [April 2, 2018] [Online] Available: https://Vikram-mm. github. io/yolo_report. pdf.

[5] Rekha, B.S. and Athiya Marium, D.G.N.S.S.A.S., 2020. Literature Survey on Object Detection using YOLO. *International Research Journal of Engineering and Technology (IRJET)*.

[6]Abbas, S.M. and Singh, S.N., 2018, February. Region-based object detection and classification using faster R-CNN. In *2018 4th International Conference on Computational Intelligence & Communication Technology (CICT)* (pp. 1-6). IEEE

[7] Vidyarani H J, Srishail math , Robust and Efficient Person Re-Identification Model using K-Nearest Neighbor Graph International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075, Volume-9 Issue-2, December 2019