International Journal of All Research Education and Scientific Methods (IJARESM),
ISSN: 2455-6211, Volume 11, Issue 8, August-2023, Available online at: www.ijaresm.com

# Real Time Human and Object Detection with Yolov3

**Chetlapelly Sai Kiran Goud[1], I. Govardhana Rao[2]**

[1]Master of Technology, Department of Computer Science, University College of Engineering, Osmania University, India
[2]Assistant Professor, Department of Computer Science, University College of Engineering, Osmania University, India

---

## ABSTRACT

**Real-time live human detection is a critical task in computer vision and deep-learning techniques with numerous applications, such as security, surveillance, and robotics. In recent years, deep learning-based object detection methods have demonstrated remarkable in accurately detecting objects in real-world scenarios. Among these methods, the YOLOv3(You Only Look Once version3) algorithm emerged as a popular algorithm for its outstanding speed and competitive accuracy. It is single-short algorithm, which means that it predicts bounding boxes and class probabilities for all objects in a image in single pass. This makes YOLOv3 very fast, making it suitable for real time applications. In this paper, we propose a method for a real time live human detection based on YOLOv3. We leverage the power of convolutional neural networks(CNNs) and YOLOv3's efficient architecture to achieve near real-time performance, ensuring rapid and reliable human detection in dynamic environments. Our method uses pre-trained YOLO3 model to detects human in live video feed. We evaluate real time performance while maintaining high accuracy. Our results show that our method is a promising approach for real time live human detection method is fast, accurate and scalable. The proposed system is capable of detecting humans in complex scenes, handling occlusions, and maintaining high accuracy even in challenging lighting conditions which easy identification of humans and other objects. The contributions of the research include the development of an efficient and reliable real-time live human detection system, based on YOLOv3 architecture. It can be seamlessly integrated into various applications that require real-time human detection, contributing to enhanced safety, security, and automation in numerous domains. Moreover, the insights gained from this study can pave the way for further improvements and advancements in the field of real-time human detection using deep learning techniques.**

**Keywords: YOLOv3, Object detection, real time, live human detection, security, surveillance, robotics.**

## INTRODUCTION

In recent years, there has been a significant surge in the development and implementation of computer vision technologies, particularly in the domain of object detection. Among various state-of-the-art algorithms, YOLOv3 has emerged as a powerful and efficient for real-time live human detection. This advanced deep learning model has revolutionized the way we detect and track humans in real-world scenarios, providing unparalleled speed and accuracy.

Human detection is vital task with wide range of applications like human -computer interaction , medical imaging, autonomous vehicles. The objective is to regardless of their pose ,scale and appearance cluttered backgrounds, and potential occlusions faced unreal-world environments.

YOLOv3 is an extension of its predecessors, YOLO and YOLOv2, that addresses many of the limitations of the traditional object detection methods. YOLOv3 introduces several key innovations that significantly enhance human detection accuracy and real-time performance. One of the main advantages of YOLOv3 lies in its unique approach to object detection instead of dividing the image into a grid of regions and performing multiple region proposals like previous models. YOLOv3 adopts a single-shot detection framework.

It directly predicts bounding boxes and class probabilities for a predefined set of objects including humans, in one go. The design makes YOLOv3 inherently faster than two-stage detectors, making it ideal for real-time applications. YOLOv3 represents a significant milestone in real-time live human detection, offering a potent combination of speed and accuracy.

Its innovative architecture and advanced techniques make it an indispensable tool for a wide range of computer vision application. As research and development in deep learning continue to progress, we can expect further improvements and even more sophisticated approaches to human detection in the future.

## RELATED WORK

There are two well-known best approaches for object detection like one-stage and two-stage detection process models. The regional proposal network or selective search is the first step in two-stage detection process to create regions of possible target objects. In comparison these models promise more accuracy and precision. The one-stage process model directly searches dense sample areas without performing on area of proposal which it faster and efficient than other. The Yolo series algorithms employ single-stage target detection and are very remarkable for object detection. YOLOv3 is more precise and accurate algorithm compared with its predecessors. YOLOv3 is the advanced version of of its older versions YOLOv1 and YOLOv2. YOLOv3 substitutes the mutual exclusion principle by making three predictions at each level which make it perform exceptionally even small size objects. In this paper we will detect humans in live feed for various applications such as self-driving cars, surveillance etc.

### YOLOv3

The YOLOv3 one-stage real time object detection algorithm that identifies objects in video, live feeds or images. The YOLO machine learning algorithm uses features learned by a deep convolutional neural network to detect objects. YOLOv3 algorithm is more accurate and precise version of the original machine learning algorithm .The foundation upon which YOLOv3's object detection algorithm is built is regression. Both object localization and object recognition steps are carried out concurrently, and the output layer transmits the bounding box position and category back.

Darknet 53 is used for feature extraction. In each grid cell of the original input picture's S x S grid, N bounding boxes are forecasted, allowing for the identification of several object categories. Each convolutional layer in Darknet53 is followed by batch normalization and Leaky ReLU activation, making the network entirely convolutional since no pooling layer is employed. For item detection, 53 additional layers are added to the Darknet53 network, resulting in a total of 106 layers in the YOLOv3 convolutional network. The bounding box parameters are predicted by first applying a log space transformation to the network's output and then multiplying it by an anchor. By default, the YOLOv3 model consists of three detection layers. The YOLOv3 default model's three layers are expanded by adding two additional layers based on the concept that a larger detection layer can identify more object details and enhance the effectiveness of small object detection. As a result, YOLOv3 (five layers) has five prediction layers on five scales. Both the three-layered and five-layered YOLOv3 models use Darknet-53 as the backbone. The main improvements in the YOLOv3 network are the choice of anchor boxes, cluster analysis of unique datasets, and the selection of the right quantity and size of candidate boxes. YOLOv3 is the fastest training method. According to the trial findings, YOLOv3 achieves recognition accuracy that is 1% to 3% better than Faster R-CNN and SSD. YOLOv3 outclasses the other two models in terms of recognition speed and can achieve almost real-time performance.

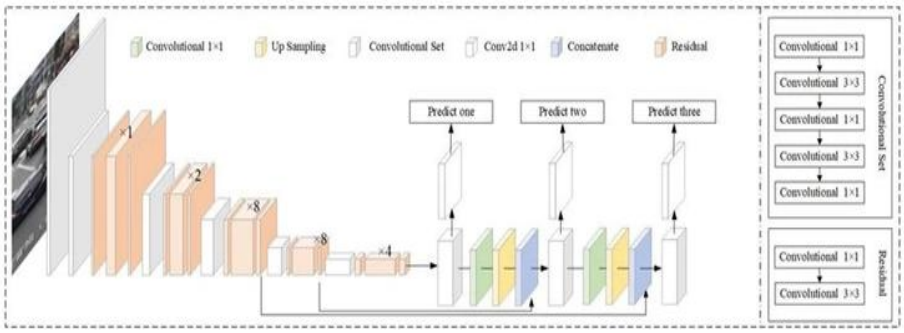


**Fig.3.1.Layered Architecture of YOLOv3**

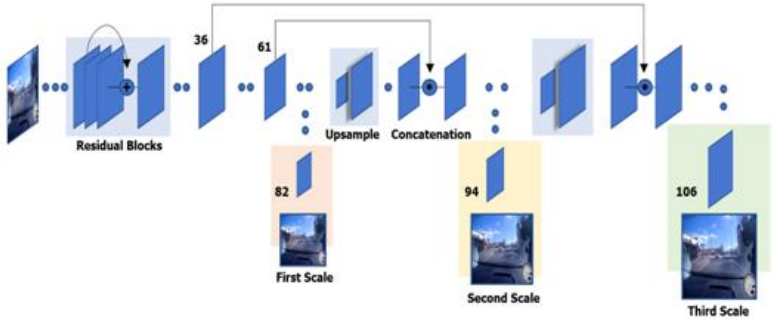


**Fig.3.2.YOLOv3 model diagram representing the residual blocks, different scales of detection, upsampling and concatenation operations**

**Yolo Algorithm:**

```
def yolo(image):
# 1. Divide the image into a grid.
grid = image. reshape((grid_size, grid_size, 3, 5))
# 2. Predict bounding boxes and class probabilities for each grid cell.
boxes = grid[:, :, :, :4]
probs = grid[:, :, :, 4:]
# 3. Apply non-max suppression to remove overlapping bounding boxes.
boxes, probs = non_ max_ suppression(boxes, probs)
# 4. Return the bounding boxes and class probabilities.
return boxes, probs
```

## EXPERIMENT AND RESULTS

Here, we present an overview of the experimental setup, dataset, and performance metrics of the YOLOv3

**Experiment System Details**

1. Apple M1 silicon chip
2. OS: Mac
3. The python programming language used to write the software code.

**Dataset**

The real time images which are captured through the webcam of the mentioned system and also the configuration and weight of pre-trained yolov3 algorithm are being used.

**Performance Metrics**

**Precision:**

The proportion of true positives to all positive predictions is defined as precision. This proportion shows how likely the model is for the expected bounding box to match the actual ground truth box. The below equation gives the formulation to calculate the precision.

$$\text{Precision} = (TP)/(TP + FP)$$

**Recall:**

The fraction of true positives among all real items is what recall is. It displays the probability of genuine things being successfully detected by a model.The following equation is used to calculate recall values.

$$\text{Recall} = (TP)/(TP+FN)$$

**Mean Average Precision (mAP):**

The mean Average Precision (mAP) is yielded by comparing the detected bounding box to the actual ground truth bounding box.

$$mAP = 1/N * \Sigma AP$$

Whereas TP is the True Positive , FP is the False Positive , FN is the False Negative , N is number of classes,AP is the Average precision for a class.

## RESULTS

The below table show the results:

**Table.4.4.1. Results Showing Values**

| | |
|---|---|
| Precision | 0.97 |
| Recall | 0.96 |
| mAP | 0.98 |

| | |
|---|---|
| Precision | 0.97 |
| Accuracy | 0.97 |

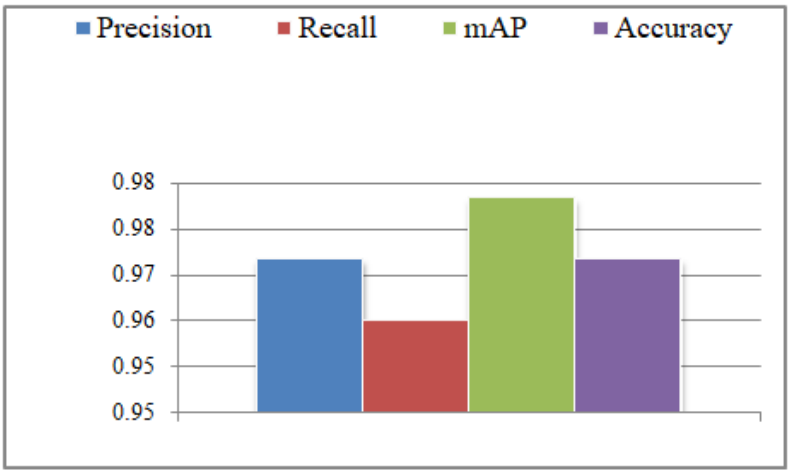


**Fig.4.4.2.Bar Graphs**

The below figures showing Real Time detection of objects and persons with accuracy rate using system webcam :

**Fig.4.4.1. The real time live image captured using webcam detected 2-persons and objects in the frame**

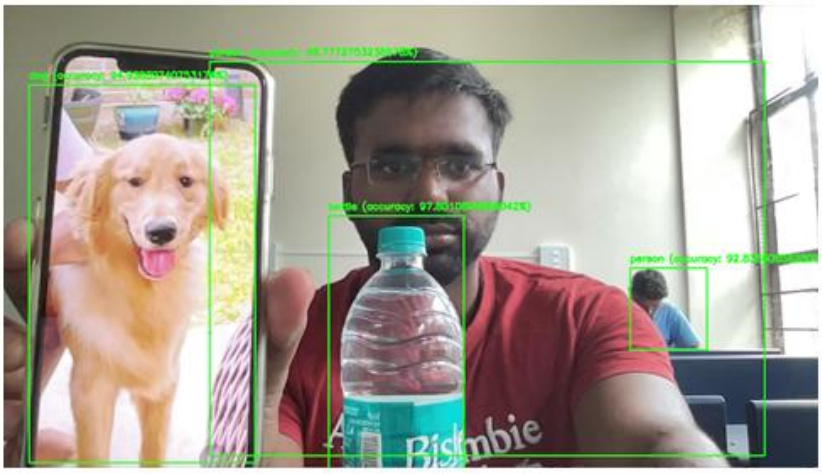

**Fig.4.4.2. The Yolov3 detected the persons and animal and a bottle with great accuracy**

## DISCUSSION

As we know YOLOv3 works by dividing the image to a grid of cells. Each cell predicts the probability of the presence of an object, as well as the bounding box for the object. The model has pre-trained data set which contained labeled objects. YOLOv3 uses a technique called inference which is the process of using a trained model to make predictions on the new data in this live feed(real - time video ). The Inference process for YOLOv3 is relatively fast. It can detect humans and object in real, even on low-powered devices

From the experiment it is observed that the accuracy the YOLOv3 model is above 98.89% .The real time live images captured using system webcam detected persons, a water bottle, a cell phone in Fig.4.4.1 and dog in frame in Fig .4.4.2 simultaneously with great accuracy above 98% which helps robots to identify the obstacles and also self-driving cars to drive safely on roads with our any obstacles. The security surveillance camera can also detected the objects and humans with great accuracy. Therefore, the YOLOv3 model has strong applicability in detecting objects and humans in real time which is great for real time use in security , autonomous driving cars, robotics, and there also one important use of his model in detecting humans and animals in incase on fire accident in a building application. In general any model is not perfect same as YOLOv3 sometimes makes mistakes in detecting objects like it may not identify an object or misidentify an object.

Overall YOLOv3 is powerful tool for real-time human and object detection. As it very fast , accurate and versatile.

## CONCLUSION

In this paper, we wanted to detect humans and objects in real time using live webcam of the mentioned system by applying YOLOv3 one-stage model algorithm. The YOLOv3 model has pre-trained weights for the feature extractor, with our live feed. We observed that YOLOv3 model works efficiently in real time live feed. Humans and Objects in the frame are detected very fast and achieved great results with an accuracy of almost all 97%. With additional training and pre-trained weights we can achieve more accuracy and detection of all objects in the frame. As research and development in deep learning continue to progress, we can expect further improvements and even more sophisticated approaches to human and object detection in the future.

## REFERENCES

[1]. Nguyen, D., Li, W. & Ogunbona, P. O., ”Human detection from images and videos: a survey”. Pattern Recognition, 51 148-175, 2016.
[2]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, ‘‘You only look once: Unified, real-time object detection,’’ in Proc. IEEE Conf. Computer. Vis. Pattern Recognition. (CVPR), Jun. 2016, pp. 779–788.
[3]. N. Dalal and B. Triggs, ”Histograms of oriented gradients for human detection,” in 2005 IEEE Computer Society Conf. Computer Vision and Pattern Recognition (CVPR’05), 2005,vol. 1, pp. 886-893.
[4]. A. Krizhevsky,I. Sutskever and G.E. Hinton, ”Image Net Classification with Deep Convolutional Neural Networks,” in Neural Information Processing Systems, 2012, doi: 25. 10.1145/3065386.
[5]. M. Vandersteegen, K. V. Beeck and T. Goedeme ́,”Real-Time Multi- spectral Pedestrian Detection with a Single-Pass Deep Neural Net- work,” in Int. Conf. Image Anal. and Recogni., 2018, pp. 419–426.
[6]. W. Liu et al.,” SSD: Single Shot Multi Box Detector,” in European Conf. computer vision, Spinger, 2016, vol 9905, pp. 21–37.
[7]. J. Redmon and A. Farhadi,”YOLOv3: An Incremental Improvement,”2018.
[8]. N.-D. Nguyen, T. Do, T. D. Ngo and D.-D. Le,” An Evaluation of Deep Learning Methods for Small Object Detection,” in Journal Elect. and Computer Eng., Apr. 2020.
[9]. A. K. S. Kushwaha, C. M. Sharma, M. Khare, R. K. Srivastava and A. Khare, ”Automatic multiple human detection and tracking for visual surveillance system,” 2012 International Conference on Informatics, Electronics & Vision (ICIEV), Dhaka, 2012, pp. 326- 331.
[10]. A. Kathuria.”What’s new in yolo v3?.” towardsdatascience.com. https://towardsdatascience.com/yolo-v3-object-detection- 53fb7d3bfe6b (accessed Jul. 10, 2020).
[11]. J. Redmon. Darknet: Open source neural networks in c. http://pjreddie.com/darknet/ (accessed Jul. 10, 2020).
[12]. N. Gaba, N. Barak and S. Aggarwal, ”Motion Detection, Tracking and Classification for Automated Video Surveillance”, IEEE 1st International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEICES), pp. 1-5, 2016.
[13]. A. Nurhopipah, A. Harjoko: ”Motion Detection and Face Recognition For CCTV Surveillance System”, IJCCS (Indonesian Journal of Computing and Cybernetics Systems) Vol.12, No.2, July 2018, pp. 107 118, 2018.
[14]. Lokesh Heda, Parul Sahare: ”Performance Evaluation of YOLOv3, YOLOv4 and YOLOv5 for Real-time Human Detection.”, IEEE 2nd International Conference on Paradigm Shift in Communications Embedded