International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Special Issue: 1 Volume No.7 April – 2023
DOI:10.46647/ijetms.2023.v07si01.081 ISSN: 2581-4621

Object Detection with Voice Feedback

U. Prem sagar[1], C Indraja[2], N Divya[3], M Haripriya[4], A Harikrishna[5]
Assistant professor, Student, Department of CSE, AITS, Tirupati

Abstract

Vision is one of the very essential human senses and it plays the most important role in human perception about surrounding environment. The main aim of our project is to develop an application which recognizes the object and gives the feedback in the form of audio. With the rapid development of deep learning, many algorithms were improving the relationship between video analysis and image understanding. All these work differently with their network architecture but with the same aim of detecting multiple objects within complex image. To design this system, we are using YOLO (You Only Look Once) approach. The main aim of this study is object detection with latest version YOLO_V3 premsagarrsp@gmail.com algorithm with audio feedback that can help blind persons recognize

1. INTRODUCTION

Humans almost by birth are trained by their parents to categorize between various objects as children self is one object. Human Visual System is very accurate and precise that can handle multi-tasks even with less conscious mind. When there is large data then we need more accurate system to correctly recognize and localize multiple objects simultaneously. Here machines come into existence, we can train our computers with the help of better algorithms to detect multiple objects within the image with high accuracy and preciseness. Object Detection is the most challenging application of computer vision as it requires complete understanding of images. In other words, object tracker tries to find the presence of object within multiple frames and assigns labels to each object [1]. There might be many problems faced by the tracker in terms of complex image, Loss of information and transformation of 3D world into 2 D image. To achieve good accuracy in object detection we should not only focus on classifying objects but also on locating the positions of different objects that may vary image to image [2]. It is very important to develop the most effective real time object tracking algorithm which is a challenging task. Deep learning since 2012 is working in these kinds of problems and has revolutionized the domain of computer vision. This paper aims to test the performance of both the algorithms in different situations in real time using webcam and is made primarily for the visually impaired peoples. Blind peoples have to rely on someone who can guide them or on their physical touch which is sometimes very risky also.

Daily navigation of blind peoples in unfamiliar environments could be the frighten task without the help of some intelligent systems. They key concern behind this contribution is to investigate the possibility of expanding the counts of objects at one go to expand the support given to the visually impaired peoples. Some common limitations of the previous techniques are less accuracy, complexity in scene, lightening etc. To overcome all those challenges two algorithms are analysed on all possible grounds and from every perspective to achieve good accuracy.

2. RELATED WORK

In recent years many algorithms are developed by many researchers. Both machine learning and deep learning approaches work in this application of computer vision. This section outlines the journey of the different techniques used by the researchers in their study since 2012.

The study [1] authors focus on Real time object detection and tracking is an important task in various computer vision applications. For robust object tracking the factors like object shape variation, partial and full occlusion, scene illumination variation will create significant problems. Here object detection and tracking approach that combines Prewitt edge detection and Kalman filter is introduced. The target object’s representation and the location prediction are the two major aspects for object tracking this can be achieved by using these algorithms. Here real time object tracking is developed through webcam. Experiments show that our tracking algorithm can track moving object efficiently under object deformation, occlusion and can track multiple objects.

In this study [2], the authors proposed a brief introduction on the history of deep learning and its representative tool, namely Convolutional Neural Network (CNN). Then we focus on typical generic object detection architectures along with some modifications and useful tricks to improve detection performance further. As distinct specific detection tasks exhibit different characteristics, we also briefly survey several specific tasks, including salient object detection, face detection and pedestrian detection.

In another study [3], the authors employed histograms of oriented gradients for human detection. This study explains the question of feature sets for robust visual object recognition; adopting linear SVM based human detection as a test case. After reviewing existing edge and gradient based descriptors, we show experimentally that grids of histograms of oriented gradient (HOG) descriptors significantly outperform existing feature sets for human detection. This explains the influence of each stage of the computation on performance, concluding that fine-scale gradients, fine orientation binning, relatively coarse spatial binning, and highquality local contrast normalization in overlapping descriptor blocks are all important for good results. The new approach gives near-perfect separation on the original MIT pedestrian database, so we introduce a more challenging dataset containing over 1800 annotated human images with a large range of pose variations and backgrounds.

3. ANALYSIS

This paper explains the application developed for object detection with voice feedback by using below algorithms which are explained in brief.

3.1 Object Detection

The main purpose of object detection is to identify and locate one or more effective targets from still image or video data. It comprehensively includes a variety of important techniques, such as image processing, pattern recognition, artificial intelligence and machine learning. Before Object recognition has developed rapidly, starting with the deep learning–based convolutional neural network (CNN) technique [5] that drew attention at the ImageNet 2012 competition. The CNN, however, was accurate with object classification, but it was difficult to determine where inside the image the object was located. Subsequently, the model for solving this problem was the region-based consolidated neural network (R-CNN), which uses a linear regression method. However, due to the slow speed of the R -CNN, Fast R-CNN was developed. It utilizes a deep learning technique to not only classify the object but also to find the area the object is located in. Nonetheless, there was a limit in that the above model's object recognition processing speed was insufficient for real time object recognition. Since then, You Only Look Once (YOLO), which comprises all the processes of object recognition as a deep learning network, has emerged, and technologies with fast detection speeds, such as Single Shot Multi Box Detector (SSD), have been developed. YOLO estimates the type and location of objects using regression inference on the problem of area selection and classification. It is faster to train than the Faster RCNN and is more accurate than YOLO because it uses different sizes of feature map.

3.2 Image Processing

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. It is a type of signal processing in which input is an image and output may be image or characteristics/features associated with that image. Image processing basically includes the following three steps:

- Importing the image to the system.
- analysing and manipulating the image.
- Output in which result is displayed

3.3 OpenCV

Techniques for Object Recognition in Images and Multi Object Detection and segmentation is the most significant and testing central undertaking of Computer vision. It is a basic part in numerous applications, for example, image search, scene understanding, and so far. However, it is as yet an open issue because of the assortment and multifaceted nature of item classes and foundations. The most effortless approach to identify and fragment an item from a picture is the shadingbased techniques. The term and the foundation ought to have a critical shading distinction so as to effectively portion objects utilizing shading-based strategies.

3.4 Yolo v3

In this project we have used YOLO v3 which is faster than the prior version. It works three times faster, at 320 ×320 YOLOv3 runs 22ms at 28.2 map. It has a similar performance but 3.8× faster. The most notable characteristic of v3 is that it makes 3 distinct scales of detections. YOLO v3 is a fully convolutional neural network and it generates its resultant output by applying a 1 x 1 kernel to a feature map. In YOLOv3, the recognition is obtained by implementing 1x1 detection kernels to three-size feature maps at three different regions in the network. Within each boundary the network predicts 4 coordinates tx, ty, tw, th. Whereas if a cell is offset in the upper left corner of the image by (cx, cy) and prior bounding boxes has pw, ph width and height respectively then the prediction is done.

4. WORKING

The entire system is present as an Android based smartphone application. We are using Python3 for this project, the camera is initialized by using OpenCV library and the camera starts capturing frames with the rate of 30 frames per second to the algorithm. Then the system uses YOLO v3 which is trained on the COCO dataset and Deep Neural Network (DNN) to identify the object kept before the user. The object identified is later converted to an audio segment using web speech Api. The audio segment is the output of our system that gives the spatial location and name of the object to the person. Now by using this information the person can have a visualization of the objects around him. The proposed system will even protect the person from colliding to the objects around will secure him from injuries.

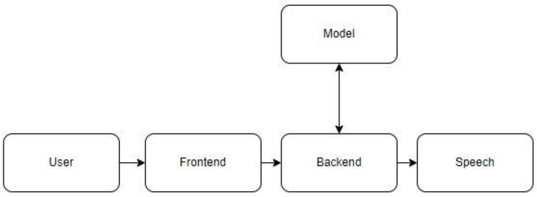


Fig 1: Block Diagram

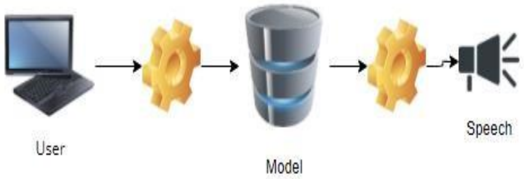


Fig 2: Proposed Platform

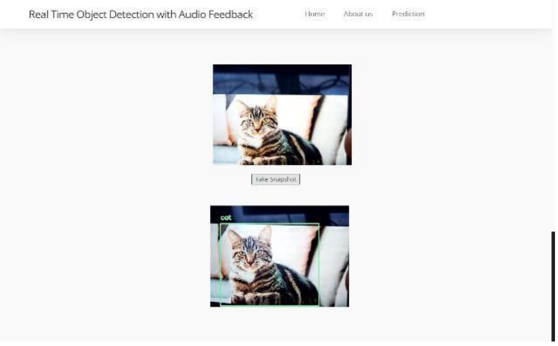
5. RESULTS

5.1 Android Application

When the user decides to start the object detection process, he opens the Camera View of the Android App. This triggers the object detection process. The video frames are pre-processed to ensure that the frames do not have noise disturbance, the frames are not blurry, etc. The individual video frames are sent asynchronously to the YOLO object detection system.

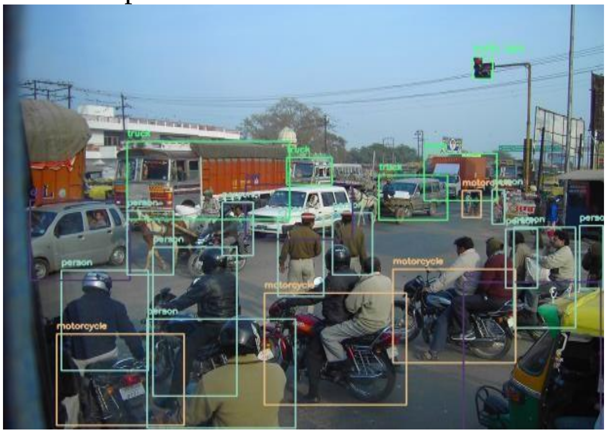
5.2. Object Detection

To use YOLO algorithm, it is necessary to establish what is actually being predicted. Ultimately, we aim to predict a class of an object and the bounding box specifying object location. Each bounding box can be described using four descriptors: 1. Centre of a bounding box (bx by) 2. width (bw) 3. height (bh) 4. value c is corresponding to a class of an object (such as: teddy bear, chair, dog etc.).



5.3. Audio Output

The audio system converts the object's label and location into audio format. The audio is then played on the smartphone speaker as an output for the user.



6. CONCLUSION

This project is for the blind people who are incapable to see this beautiful world, our initiative will support them to have a better life. By this project one will be able to understand what object is present in front of him and by continuous research and development our team will be able improve this product by feeding more data to the Deep Learning algorithm by which the accuracy of the model will increase as well as the power of the algorithm to recognize more objects will increase. This application aims to enable people with visual impairment to live more independently. People with visual impairment will be able to overcome some threats that they may come across in their day-to-day life that may be either while reading a book or traveling through the city by making efficient use of the application and its associative voice feedback. Therefore, it will help to prevent possible accidents. The mobile devices can be carried easily and the camera of the device can be used to detect object from the surroundings and give output in audio format.

6. REFERENCES

S. Cherian, & C. Singh, "Real Time Implementation of Object Tracking Through webcam," International Journal of Research in Engineering and Technology, 128-132, (2014)

Z. Zhao, Q. Zheng, P.Xu, S. T, & X. Wu, "Object detection with deep learning: A review," IEEE transactions on neural networks and learning systems, 30(11), 3212-3232, (2019).

N. Dalal, & B. Triggs, "Histograms of oriented gradients for human detection," In 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886893). IEEE, (2005, June).

S. Geethapriya, N. Duraimurugan, & S.P. Chokkalingam, "RealTime Object Detection with Yolo," International Journal of Engineering and Advanced Technology (IJEAT), 8(3S), (2019).

Rahul Kumar and Sukadev Meher, "Assistive System for Visually Impaired using Object Recognition, M.Sc. Thesis at Department of Electronics and Communication Engineering, National Institute of Technology Rourkela, Rourkela, Odisha769 008, India, May 2015.

R. Bharti, K. Bhadane, P. Bhadane, & A. Gadhe,
"Object Detection and Recognition for Blind Assistance," International Research Journal of Engineering and Technology (IRJET) e-ISSN:
2395-0056 Volume: 06, (2019).

[7J Redmon & A. Farhadi, "Yolov3: An incremental improvement," ArXiv preprint arXiv: 1804.02767, (2018).

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

G. Peng, "Performance and Accuracy Analysis in Object Detection," (2019).

Kedar Potdar, Chinmay Pai and Sukrut Akolkar, "A Convolutional Neural Network based
Live Object Recognition System as Blind Aid", arXiv:1811.10399v1 [cs.CV] 26 Nov 2018
https://arxiv.org/pdf/1811.10399.pdf

X. Wang, A. Shrivastava, & A. Gupta, "A-fastrcnn: Hard positive generation via adversary for object detection," In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2606- 2615), (2017).