International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 10 | Oct - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

Review on "Real -Time Object Detection with Speech Feedback for Visually Impaired"

Mr. M. S. Khan[1], Ritul Ghumare[2], Prerna Malode[3], Rutuja Murkute[4], Sakshi Pagare[5]

**1 Assistant Professor, Department of Information Technology, Matoshri College of Engineering & Research Centre, Eklahare, Maharashtra, India*
**2,3,4,5 Students, Department of Information Technology, Matoshri College of Engineering & Research Centre, Eklahare, Maharashtra, India*

ABSTRACT –

It's a known fact that estimated number of blind persons in the world is about 285 million, approximately equal to the 20% of the Indian Population. They are mostly dependent on someone for even accessing their basic daily needs. In our project, we used TensorFlow, it\'s a new library from Google. TensorFlow model our neural networks. The TensorFlow Object Detection API is used to detect many objects. We have Introduce an algorithm (YOLO). YOLO uses a similar phase while training, to match the appropriate anchor box with the bounding boxes of each ground truth object within an image. Essentially, the anchor box with the highest degree of flap with an object is responsible for predicting that object\'s class and its location. It has microcontroller which has wi-fi inbuilt module. This guide is convenient and offers data to the client to move around in new condition, regardless of whether indoor or open air, through an ease to use interface. Then again, and so as to lessen route challenges of the visually impaired, a deterrent location framework utilizing ultrasounds is added to this gadget. The proposed framework identifies the closest hindrance through ultrasonic sensors and it gives an alert to illuminate the visually impaired about its confinement.

***Key Words*:** Python, Machine Learning, YOLO lib. Datasets, Opencv.

1.INTRODUCTION

The Technology for navigation ofthe blind is not sufficiently accessible devices rely heavily on infrastructure requirements. Without vision it can be challenging for visually im-paired persons to navigate through rooms or different road paths .The main aim to develop the project is to help the visually impaired people and to detect the obstacles to detect the road traffic signs. The blind persons life become easier and they can go anywhere where they wants without anyone helps .They can walk alone through street they does not need anyone to assist them they can handle their self correctly. The preventing users from dangerous location our aim is to collected from environment (cameras, sensors, scanners, etc) and transmitted to the users to the audio format. When sensors identify objects it send the data to the audio module and get converted into audio clip. An obstacle as close as 4 cm can be detected by these module. The objective of these project is to provide cost effective way to allow path planning for blind people. With the help of object detection system and navigation system the blind person can easily detect the object through cameras or through some sensors. For example, it is very difficult for to find a particular room in an unfamiliar environment. The blind and visually impaired people find it difficult to know a person is talking to them or someone else during a conversation.

The Computer vision technologies, especially neural network, have been rapidly developed. The aim of to guide the blind people through the output of processor or controller to navigate them. The purpose ofthe project is to a system to help blind people in their life. With the help ofthese system the blind person do theirwork more effectively. They does not depend on any others to perform any activity. With the help of these system they can walk alone on the road or traffic areas or the park anywhere they want. They are very curious about the beauty of the world, what happening in front so with the help of these system they can enjoy their life. The portable device, designed for visually impaired people to assist them with getting around. Most commercially available devices, these device should provide direction to location and alert the blind people of obstacle in their path one of the most important things is to alert the blind peoples to any obstacle in their path and navigate the road or different areas to the blind persons through the voice assistance. It converts the message and these message send to the users through the audio signal, it also detect the road traffic signs,

with all of these module the blind persons life become easier. In these way we are try to developing the project to help the blind people and they can enjoy their life.

2. LITERATURE SURVEY

Abdul Muhsin M, Farah F. Alkhalid, Bashra Kadhim Oleiwi have proposed their work on "Online Blind Assistive System Using Object Detection" in 2020. In this work, the function of computer vision is to detect indoor objects accurately. The visually impaired people can be assisted by navigating the purposes of the CNN framework.4,5,14 To identify the specific objects first, we need to detect the pixels available in the images. If the lighting conditions are wrong, then it is challenging to capture and identify the objects with high accuracy. To detect the indoor objects, the algorithm needs to extract the image features with a particular class, and it can be done by RetinaNet.25 To enable the network for small object detection by a Region Proposal Int J Cur Res Rev

The object detection system in [3] is used to detect objects in the traffic scenes. Here they have used the combination of optimized you only look once (OYOLO), which is 1.18 times faster than YOLO and R-FCN (Regression based Full Convolution Network). It is used to detect and classify the images such as cars, cyclist and pedestrian. Use of YOLO makes location errors, to avoid that we use OYOLO. Paper [4] presents a prototype that extracts the text from image and is converted to speech. Extraction of text is done by using the Tesseract Optical Character Recognition (OCR). This method is carried out by using Raspberry Pi. Text recognition is done by using Open Computer Vision (Open CV), considering the large library of functions when compared to MATLAB. Capturing of image is done by using a portable camera and the image is converted to gray scale and filtered by Gaussian filtering. Then it is binarized and cropped. The cropped image is given to Tesseract OCR. The e-Speak creates an analog signal of the text and is given to the headset.

Object detection using machine learning for visually impaired people Networks (RPN), which involves subsampling to obtain the image information. The Resort with 152 samples achieved an average precision with 83.1%, and Dense Net with 121 samples achieved an average precision with 79.8%.

Dr. K. Sreenivasulu, P. kiran Rao have proposed their work on "A Comparative Review on Object Detection System for Visually Impaired" in 2016. This model is used for detecting the patterns in urban areas such as public streets, raining, restaurants, etc.13 This method characterizes the audio clips, which yields the patterns. The main limitation of this model is to require a trained data set. 6

3. METHODOLOGY

The aim of our project is to design a system at a low cost to enable a person who's vision is not clear to lead a normal life without depending on others in a selected environment like our house, workplace etc. All the things that are used for the day to day activities can be detected using a web cam. The process is as shown in the flowchart in Figure 1.

The object detection, recognition and communicating speed of the object to the user should be very fast. Since the proposed system is intended for visually impaired persons, the response time of the system plays a key role. Delay in communicating about the obstacles to the user will not meet the purpose. To overcome processing speed and delay issues, Yolo V3 algorithm is used in the proposed system. Yolo V3 algorithm is faster than many other real time object detecting algorithms.

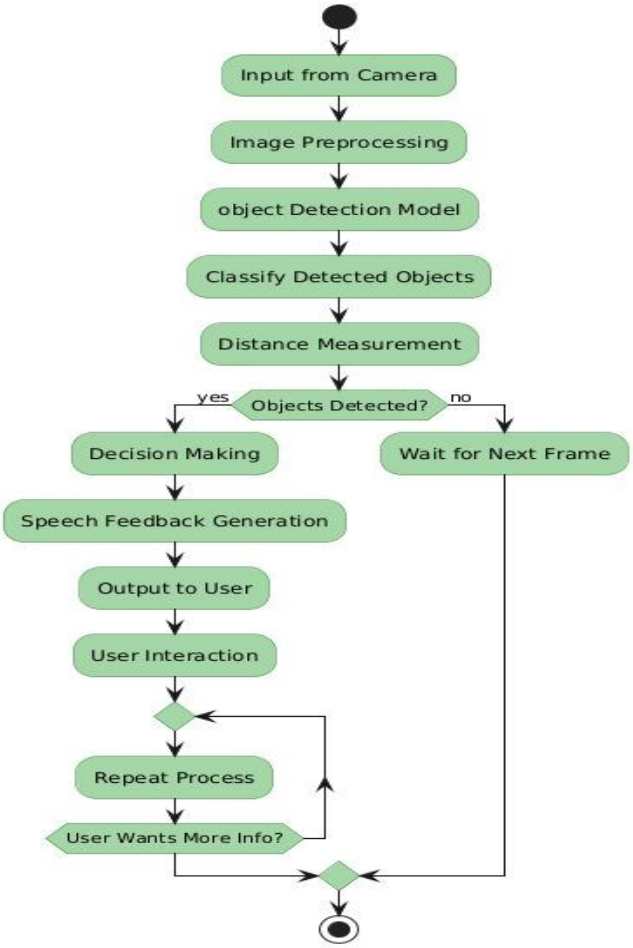

Fig no. 01: Flow chart

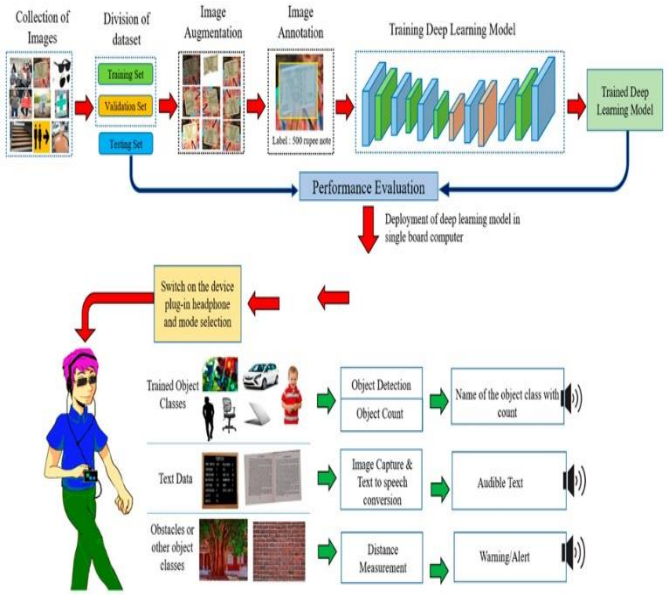

Fig no. 02: System Architecture

4. YOLO ALGORITHM :-

YOLO algorithm [6],[8] is primarily used for the prediction of bounding boxes accurately from an image. Images are divided into N x N grids and for each grid the prediction of the bounding boxes are done as well as the class probabilities. After performing object localization and image classification for each grid of the input image, each grid is given a different label.

The following project is designed as a YOLO algorithm applied separately on each grid and the objects in it are marked with their particular label and corresponding bounding boxes are also highlighted. The grids having no object are labeled as 0. Initially, YOLO algorithm is applied to the received image. In our project, the real time image is divided into grids of matrices. As the image complexity varies the image can be split into any number of grids. After division of the images, both classification and the localization process are done on each grid containing the object. The confidence score is computed for all the grids. The confidence score and also the bounding box for each of the grid will differ based on whether the object is detected or not. For no object it displays the value as 0 and displays the value 1 if object exist. Bounding box value will show how confident the network is, that is, how much the detected object matches to the object under observation.

5. Proposed System

Visually impaired persons face many challenges in identifying objects and performing the day to day activities. Most of the time they are dependent on others in moving from one place to other. They face a lot of difficulty especially when moving in outdoor environment, where objects are continuously moving. This object detection system would assist the visually impaired persons by providing the perception of the surrounding environment or position of the objects. This device can assist visually impaired persons in avoiding the obstacles in both indoor and outdoor environment. The proposed system also helps the user in identifying the objects around them. It would minimize the visually impaired person's difficulties and help them lead a quality life.

The proposed system aims to be simple, user-friendly, handy, economical and efficient solution. Most of the existing assistive systems are highly sophisticated and expensive. Smartphones are widely used these days and the usage of smartphones by all persons has become very usual in the recent past. Therefore this system uses all the advanced built-in features of a smartphone. Smartphone's integrated camera is used to capture the real time video to detect the objects present and headphones or smartphone's speaker is used to communicate to the user through audio instructions. The proposed

application can easily be accessed by the visually impaired persons.

The Yolo V3 (You only look once) algorithm is used to detect the real time objects captured from the continuous streaming by the smartphone camera. It is considered as one of the most powerful real time object detector algorithms. Unlike other algorithms like R-CNN and Faster R-CNN which examine several regions of the image to identify objects, Yolo passes image or video only one time through its network and uses a unique neural network using the characteristics of the complete image to predict multiple boxes containing an object. This is a significant feature in Yolo, which reduces the processing speed when compared with other algorithms. Processing speed plays a key role particularly while detecting objects in real time video stream. OpenCV provides a function that facilitates image pre-processing for deep-learning classification. Group of connected pixels in an image that share common properties (Blob ex: Grayscale value) of each frame captured is identified. Bounding boxes are created for each object identified and based on the Yolo V3 pre-trained weights, confidence score and coco dataset each object is processed and labelled.

The coco dataset contains all the objects or class names on which the model is trained. Non-maximum suppression (NMS) uses function called "Intersection over Union (IoU)" which is used to determine the best boxes. In order to select the best box, NMS follows three steps. 1. It selects the box with highest score. 2. Computes the overlap with other boxes, removes the overlap which has more than a certain

6. EXPERIMENTAL PROCEDURE

The experimental procedure started with collecting the database, training the YOLO network and then testing the output by giving test images. The network was trained to detect 5 classes namely C1 (class 1) for bottles, C2 (class 2) for watches, C3 (class 3) for keys, C4 (class 4) for pens and C5 (class 5) for glasses.

6.1. Collection of Data Base
6.2. Network Output Analysis
6.3. Bounding Box Dimensions
6.4. Output Filtering
6.5. Training
6.6. Text to speech

CONCLUSIONS

In recent years, there were many assistive solutions developed for the visually impaired persons to provide assistance in detecting objects present in their surrounding environment and in moving from one place to another. But most of the existing solutions are expensive, highly sophisticated, difficult to handle, designed as a dedicated aid, require training etc. Our primary goal is to provide an assistive solution to the visually impaired persons which is simple, user-friendly, affordable and handy and assist the visually impaired persons in understanding the surrounding environment and help them in moving from one place to another place independently. The real time video stream is captured by the smartphone camera and then the label of the object detected using object detection algorithm is communicated to the user in speech through the speakers or headphones. The audio output communicated to the user would help them in performing their day to day activities and lead a quality life

REFERENCES

[1] Chen X, Yuille AL. A time-efficient cascade for real-time object detection: With applications for the visually impaired. In2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)-Workshops 2005 Sep 21:28-28.

[2] Chi-Sheng, Hsieh. "Electronic talking stick for the blind." U.S. Patent No. 5,097,856, 24 Mar. 1992.
[3] WafaMElmannai, KhaledM.Elleithy. "A Highly Accurate and Reliable Data Fusion Framework for Guiding the Visually Impaired". IEEE Access 6 (2018) :33029-33054. [1]

[4] Ifukube, T., Sasaki, T., Peng, C., 1991. A blind mobility aid modelled after echolocation of bats, IEEE Transactions on Biomedical Engineering 38, pp. 461 - 465.

[5] Cantoni, V., Lombardi, L., Porta, M., Sicard, N., 2001. Vanishing Point Detection: Representation Analysis and New Approaches, 11th International Conference on Image Analysis and Processing.

[6] Balakrishnan, G. N. R. Y. S., Sainarayanan, G., 2006. A Stereo Image Processing System for

Visually Impaired, International Journal of Information and Communication Engineering 2, pp. 136 145.

[7] C.S. Kher, Y.A. Dabhade, S. sK Kadam., S.D.Dhamdhere and A.V. Deshpande "An Intelligent Walking Stick for the Blind." International Journal of Engineering Research and General Science, vol. 3, number 1, pp. 1057-1062

[8] G. Prasanthi and P. Tejaswitha "Sensor Assisted Stick for the Blind People." Transactions on Engineering and Sciences, vol. 3, number 1, pp. 12-16, 2015