International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-8 Issue-6, March 2020

Recognize Objects for Visually Impaired using Computer Vision

Deven Pawar, Mihir Raul, Pranav Raut, Sharmila Gaikwad

Abstract*: *Visually impaired people are often unaware of dangers in front of them, even in familiar environments. Due to lack of vision either partial or complete, such people are highly dependent on the sense of hearing to perform day to day activities. One more form of vision impairment is colour blindness. Individuals with colour blindness find it hard to distinguish between colours. This research proposes making software that can help such individuals for solving the unawareness of the surrounding of the visually impaired people which allows them to have a greater awareness of their surroundings. The software needs an input device typically a camera and an audio feedback system. The camera will continuously capture images and the algorithm recognize the objects in the image and output the result using the audio feedback system. The system also proposes to include colour extraction to also correctly identify the colour of the object and a further addition is to identify individuals if enough datasets are provided. If any suspicious/dangerous objects detected in the surrounding the software will inform the user about the imminent danger. This study has analysed Faster R-CNN, SSD (Single Shot MultiBox Detector) and YOLO (You only look once) for their accuracy and rate of object detection. This research also studied different operating scenarios of the device which includes operation at night and operation in various orientations. The results of the object recognition system while using YOLO have an accuracy of 59.7% and 10fps during real-time operation, which is sufficient for assisting visually impaired people in realizing the types and localities of the objects around them.

Keywords* : *Computer Vision F-CNN, Raspberry-Pi, Visually Impaired, YOLO.

I. INTRODUCTION

visual impairment is the decreased ability to see, which causes problems that are not correctable by usual means, such as glasses, contact lenses or medical treatments. Visual impairment causes people difficulties in their normal daily activities such as driving, reading, socializing, and walking. According to facts[3] by WHO (World Health Organization), globally, it was estimated that approximately 2.2 billion people have a vision impairment or blindness out of which 205.3 million have a moderate or severe distance vision Whereas severe vision-impairment or blindness is untreatable and people suffering from such impairment have no choice but to use a certain aiding device such as a walking cane to help them navigate. The need for this research was that there have been fewer aids that help the visually impaired. Visually impaired people have been struggling to detect and locate objects, so there was a strong motivation to build a software that would do the detection and identification of the objects that lie in front of the user. Rather than an object even if a person is in front of the camera the software will be able to detect who the person is. This task is done by first detecting the object towards which the camera is pointed, the algorithms then identify the object and then give audio feedback about what the object or person lies in front of them.

This research focuses on helping these people to ease their navigation and increase awareness of the objects in the surrounding. This study proposes a making of software for solving the unawareness of the surrounding of visually impaired people which allows visually impaired people to navigate smoothly with greater awareness of the surrounding. The software helps visually impaired in detecting objects along with their colors and also helps them in detecting the color of the object helping them distinguish between colors also the software helps in face-detection in case the visually impaired person should be aware of the person that is in their vicinity. The software could be installed on a compact device.

II. OBJECTIVES

The main aim of this paper is to capture images, analyze them through an image detection algorithm, get a suitable and relatable output and get audio feedback regarding the output. This is done through state-of-the-art YOLO[1] (You Only Look Once) via camera module feeding data. YOLO, which was originally written in C, this research uses its Python implementation called DarkFlow. The algorithm is a Single-Shot MultiBox Detector. The objectives are met by running YOLO on real-time and detecting the object in the frame, the output labels are then sent to our audio synthesizing module e.g. Espeak etc. The further functionalities that this research provides is the detection of suspicious objects e.g. Guns, knives etc.

Manuscript received on March 15, 2020.
Revised Manuscript received on March 24, 2020.
Manuscript published on March 30, 2020.
* Correspondence Author
Deven Pawar*, U.G. Student Computer Department, Rajiv Gandhi Institute of Technology, Mumbai, India. Email: devenpawar8@gmail.com
Mihir Raul, U.G. Student Computer Department, Rajiv Gandhi Institute of Technology, Mumbai, India. Email: mihirraul18@gmail.com
Pranav Raut, U.G. Student Computer Department, Rajiv Gandhi Institute of Technology, Mumbai, India. Email: pranavraut033@gmail.com
Sharmila Gaikwad, Asst. Professor, Rajiv Gandhi Institute of Technology, Mumbai, India. Email: sharmila.gaikwad@mctrgit.com

III. RELATED THEORY

"Object detection is the craft of detecting instances of a certain class, like animals, humans and many more in an image or video." - Gilbert Tanner [2].

The software processes for any danger or objects in the image and output the result using the audio feedback system. After recognition the audio synthesizer helps use give an audio output.

A) Different Object Detection Algorithm:

1) Faster R-CNN

Quicker RCNN is a system that objects recognition. As clarified by its name, its quicker than its relatives RCNN and Fast RCNN. How quick? Practically constant quick. This system has use cases in self-driving vehicles, assembling, security, and is even utilized at Pinterest.

How Faster RCNN works:

A. Run the picture through a CNN to get a Feature Map
B. Run the Activation Map through a different system, called the Region Proposal Network (RPN), that yields fascinating boxes/regions
C. For the fascinating boxes/locales from RPN utilize a few completely associated layers to yield class +Bounding Box organizes.

The methodology resembles the R-CNN calculation. Yet, rather than bolstering the district recommendations to the CNN, it feed the information picture to the CNN to produce a convolutional highlight map. From the convolutional highlight map, we recognize the locale of proposition and twist them into squares and by utilizing a RoI pooling layer we reshape them into a fixed size with the goal that it very well may be encouraged into a completely associated layer. From the RoI highlight vector, we utilize a softmax layer to anticipate the class of the proposed locale and the counterbalance esteems for the jumping box. The explanation "Quicker R-CNN" is quicker than R-CNN is on the grounds that it doesn't need to encourage 2000 area recommendations to the convolutional neural system inevitably. Rather, the convolution activity is done just once per picture and a component map is created from it. Figure 3.1 show the working of Faster RCNN.

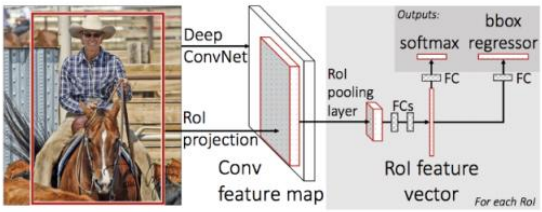


Figure 1: F-RCNN [6]

2) Single Shot MultiBox Detector:

This idea was discharged toward the finish of November 2016 and arrived at new records as far as execution and exactness for object discovery assignments, scoring over 74% mAP (mean Average Precision) at 59 edges for each second on standard datasets, for example, PascalVOC and COCO. To more readily comprehend SSD, how about we start by clarifying where the name of this design originates from:

A. Single Shot: this implies the undertaking of item limitation and arrangement are done in a solitary forward of the system.
B. MultiBox: This is the name of a procedure for jumping box relapse created by Szegedy et al.
C. Detector: The system is an article identifier that likewise arranges those distinguished items

As should be obvious from the figure 3.2, SSD's design expands on the respected VGG-16 engineering however disposes of the completely associated layers. The explanation VGG-16 was utilized as the base system is a direct result of its solid execution in great picture order errands and its prominence for issues where move learning helps in improving outcomes. Rather than the first VGG completely associated layers, a lot of assistant convolutional layers (from conv6 onwards) were included, hence empowering to extricate highlights at different scales and logically decline the size of the contribution to each consequent layer.

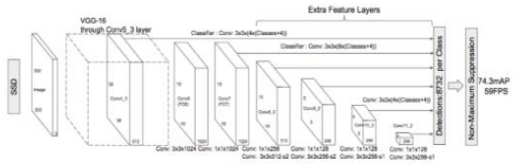


Figure 2:Architecture of Single Shot Multibox Detector; input is 300x300x3

A. Confidence Loss: This measure how certain the system is of the objectness of the figured bouncing box. All out cross-entropy is utilized to process this misfortune.

B. Location Loss: This measure the distance away the system's anticipated bouncing boxes are starting from the earliest stage ones from the preparation set. L2-Norm is utilized here

The articulation for the misfortune, which marks how far away our forecast "landed", is along these lines:

$$Multibox\_loss = confidenceloss + alpha * locationloss$$

The alpha term causes us in adjusting the commitment of the area misfortune. As normal in profound learning, the objective is to discover the parameter esteems that most ideally diminish the misfortune work, along these lines carrying our forecasts nearer to the ground truth.

3) YOLO (You Only Look Once)

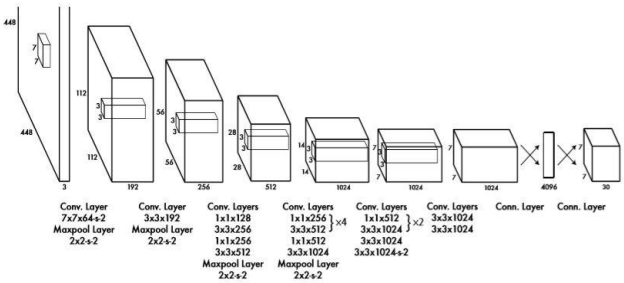
Table 1. Comparison with respect to accuracy (mAP) and frame rate (fps)

Algorithm	Input Image	mAP	Frame Rate
SSD	500x500	46.5	19
R-FCN	1000x1000	51.9	12
Tiny YOLO	608x608	23.7	24.4
YOLOv3	608x608	57.3	20

Article identification repurposes classifiers to perform discovery. Rather, YOLO outline object recognition as a relapse issue to spatially isolated jumping boxes and related class probabilities. A solitary neural system predicts jumping boxes and class probabilities straightforwardly from full pictures in a single assessment. Since the entire discovery pipeline is a solitary system, it very well may be advanced start to finish legitimately on recognition execution YOLO reframe object identification as a solitary relapse issue, directly from picture pixels to jumping box organizes and class probabilities.

YOLO model has a few points of interest over classifier-based frameworks.

It takes a gander at the entire picture at test time, so its expectations are educated by worldwide setting in the picture. It additionally makes expectations with a solitary system assessment dissimilar to frameworks like R-CNN which require thousands for a solitary picture. This makes it amazingly quick, more than 1000x quicker than RCNN and 100x quicker than Fast R-CNN. Table 1. shows correlation between various calculations regarding precision (mAP) and casing rate (fps).



Yolo detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1×1 convolutional layer reduces the features space from preceding layers. We pretrain the convolutional layers on the ImageNet classification task at half

IV. IMPLEMENTATION

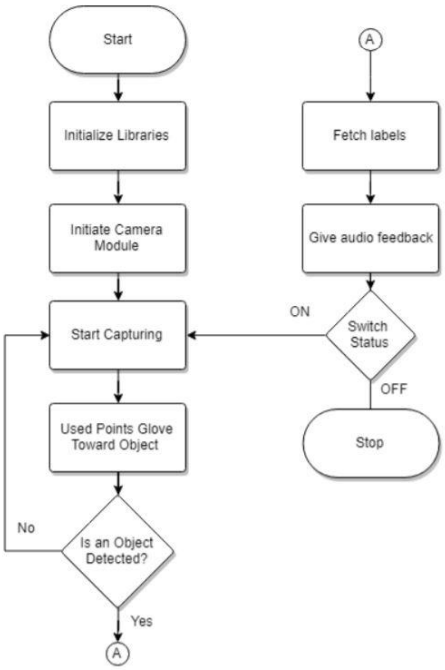


Figure 3: Flowchart of the proposed design

Taking many input images of a wide range of objects increases the possibility of recognition. The code implementation begins by importing all the packages and parsing the arguments, specifying each object class and dedicate the required boundaries such that if an object is detected in that boundary and is identified as part of the inbuilt class then the recognition is done, and the output is displayed. The program run in a continuous loop detecting object in each frame and if the probability value of the detected object is above threshold i.e. 60% then and only then the output will be displayed.

1) Dataset used for Training the Data:

A very important aspect of deep learning is finding a good dataset to train the model. A good dataset will result in better precision and recall. Some commonly used datasets are ImageNet, Pascal, SUN, and COCO. Yolo is a pre-trained object detection algorithm. The dataset used as training data for the YOLO object detection algorithm is COCO (Common objects in context). COCO is a large-scale detection, segmentation, and captioning dataset. The coco dataset contains around 80 classes, 80,000 training images, and 40,000 validation images. Currently, there are about 2 releases of the COCO dataset.

Figure 4: An example of classes presents in COCO dataset

As hinted by the name, images in the COCO dataset are taken from everyday scenes thus attaching "context" to the objects captured in the scenes. Let's assume we need to detect a person or an object in an image. A non-contextual, isolated image will be a close-up photograph of a person. Looking at the photograph, we can only tell that it is an image of a person. However, it will be challenging to describe the environment where the photograph was taken without having other supplementary images that capture not only the person but also the studio or surrounding scene.

COCO was an initiative to collect natural images, the images that reflect the everyday scene and provides contextual information. In the everyday scene, multiple objects can be found in the same image and each should be labeled as a different object and segmented properly. COCO dataset provides the labeling and segmentation of the objects in the images. A machine learning practitioner can take advantage of the labeled and segmented images to create a better performing object detection model.

V. ANALYSIS

1	Name:	Object Recognizer for visually impaired
	Actor:	Wearer/Device
	Description:	Process where the object is detected by the device.
	Successful completion:	1.Wearer points towards an object 2.Object is detected 3.Audio feedback given about the object
	Pre-condition:	Well-lit environment

2	Name:	Object Recognizer for visually impaired
	Actor:	Wearer/Device
	Description:	Process where a suspicious object is detected, and warning is provided.
	Successful completion:	1.Wearer points towards an object 2.Object is detected 3.Audio feedback given about the object 4.Warning provided to the wearer.
	Pre-condition:	Well-lit environment
3	Name:	Object Recognizer for visually impaired
	Actor:	Wearer/Device
	Description:	Process where color of the object is detected.
	Successful completion:	1.Wearer points towards an object 2.Object is detected 3.Audio feedback given about the object
	Pre-condition:	Well-lit environment

A. Analysis of Color Extraction using K-means:

Figure 5: Image under consideration

We have used Fig.5 as a sample image after object detection for color extraction. The clipped image i.e. the portion clipped by the rectangle is considered as the parameter for the color extraction algorithm. Clusters of the pixel values represented as RGB is formed using K-means clustering algorithm. The label of each cluster is the average value of all the pixels inside the cluster.

Fig. 6 & Fig. 7 is a visual representation of the cluster formed. X-axis represents the cluster label i.e. The average color value of the cluster. The color name of the RGB value is obtained by converting finding the color name from CSS3 color variables which the closet value.

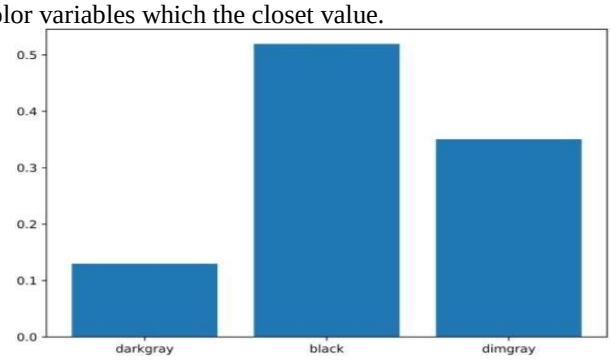


Figure 6: Bar of cluster; X-axis The label of the cluster; Y-axis Color Probability

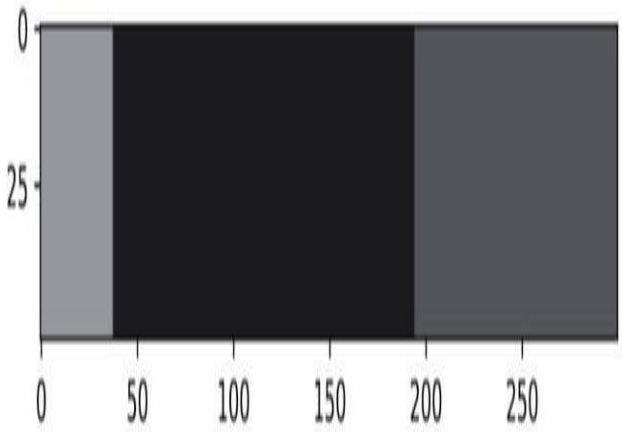


Fig. 7: Histogram of computed clusters

VI. CONCLUSION

This research aims at building a software that can detect and classify and extract color of the object towards which the camera is pointed. This research accomplishes this by using state of the art object detection algorithm called YOLO (You only look once). The color extraction is performed by detecting the different hues in the image, clustering them accordingly using K-means and then considering the color that has the largest cluster size with the background of the image masked. The software's camera module takes in the object and after comparison with the dataset; it recognizes the object in the image frame if any. After recognition, the audio synthesizer helps use give an audio output as audio cues are the only way the detection can be conveyed to the wearer.

The software will contain the ability to detect suspicious objects like knives, guns, etc. A warning will be given to the wearer about the suspicious object via audio feedback as well. With this paper being able to detect suspicious objects it'll be of help to the wearer of the device who'll be a person with impaired vision to be more aware of his/her surroundings making him feel secure by our research. This paper plays a very important role in making the visually impaired people feel in a safe environment. The software for color blind people provides with color information as well proving it to be of immense help to color blind users of the software. The delay between the detection of the object and the audio feedback to the user is made the minimum and the audio feedback is made clear enough for the user to be understood.

ACKNOWLEDGMENT

This project and technical paper were greatly supported by Dr. Sharmila Gaikwad. His appropriate mentoring and total guidance helped us at every stage of this project and this paper would not have been possible without his valuable contributions. We thank our colleagues who provided insight and expertise that greatly contributed to the research and helped us with the entire process.

REFERENCES

1. J. Redmon, S. Divvala, R. Girshick, A. Farhadi, "You only look once: Unied real-time object detection", 2015.
2. J. Redmon, "Darknet: Open source neural networks in c", 2016, http://pjreddie.com/darknet
3. World Health Organization (WHO), "Blindness and vision impairment", 8 October 2018, https://www.who.int/news-room/fact-sheets/detail/blindness-and-visual-impairment.
4. Gilbert Tanner, Towards Data Science, "Creating your own object detector", 6 February 2019, https://towardsdatascience.com/creating-your-own-object-detector-ad69dda69c85
5. Alessandro Dionisi, Emilio Sardini,Mauro Serpelloni, "Wearable object detection system for the blind", Published in 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings.
6. Sik-Ho Tsang, Towards Data Science, "Review: Faster R-CNN (Object Detection)", 15 Septembet 2018, https://towardsdatascience.com/review-faster-r-cnn-object-detection-f5685cb30202
7. Heetika Gada, Abhinav Kashyap, Vedant Gokani, Amit Deshmukh, "Object Recognition for the Visually Impaired" IEEE XPlore,Proceedings of ICNTE, Vashi, 4th-5th January 2019, India
8. Rebecca Copeland, Noel Crespi, Analyzing Consumerization – "Should Enterprise Business Context Determine Session Policy?", 2012 IEEE Proceedings of 16th International Conference on Intelligence in Next Generation Networks (ICIN), pp.187-193 ,2012. 19

AUTHORS PROFILE

Mr. Deven Pawar U.G. Student Computer Department, Rajiv Gandhi Institute of Technology, Mumbai, India.

Mr. Mihir Raul U.G. Student Computer Department, Rajiv Gandhi Institute of Technology, Mumbai, India.

Mr. Pranav Raut U.G. Student Computer Department, Rajiv Gandhi Institute of Technology, Mumbai, India.

Dr. Sharmila Gaikwad Asst. Professor in Computer Engineering Department, Rajiv Gandhi Institute of Technology, Mumbai, India.