

# Object Detection and Translation for Blind People Using Deep Learning

Mayuresh Banne<sup>1</sup>, Rahul Vhatkar<sup>2</sup>, Ruchita Tatkare<sup>3</sup>

<sup>1,2,3</sup>Department of Information Technology, Vidyalkar Institute of Technology, Wadala

\*\*\*

**Abstract** - There are million of people in India alone that are visually impaired. so, it's essential to understand for visually impaired people to recognize a product of their daily use so we made a system to identify products in their everyday routine by this system. There are many papers on this topic that will help a blind person. This paper helps a blind person in their daily use. This system consists of a camera, a speaker and an image processing system. This project tries to detect the object and transform that object into the audio form and inform blind person about those objects. Our system consists of a box which has a portable camera and a system which will process that image. image are captured with a portable camera device with real-time image recognition on existing object detection models. after detecting an object that information is translate into audio.

**Key Words:** System, Camera, Audio, Image processing, Object.

## 1. INTRODUCTION

Millions of people live in this world who can't see environment due to visual impairment. Although they can develop alternative approaches to deal with daily routines, but sometimes there are from certain objects they just can't tell without feel of touch A variety of object is processing and machine learning techniques have been applied to the problem, including matrix factorization, dictionary learning and most recently mask region Convolution neural networks (MASK RCNN). In particular, mask region Convolution neural network (MASK RCNN) are, in principle, very well suited to the problem of object detection and recognition First, it generates proposals about the regions where there might be an object based on the input image. Second, it predicts the class-id of the object and define a bounding box and generates a mask at pixel level of the object based on the first stage proposal. Mask-RCNN is additional feature of Faster RCNN. Fast RCNN has an output for each object that are class label and bounded-box offset. But in Mask-RCNN it has one feature that is object masking. Mask RCNN has additional mask output which are distinct from class and box that why it requires finer spatial layout of an object [1]. It also includes pixel-to-pixel alignment which was not present in Faster RCNN.

possible explanation for the limited exploration of CNNs and the difficulty to improve on simpler models is the relative scarcity of labelled data for object detection.

## 2. LITERATURE SURVEY

### 2.1 Prof. Seema Udgirkar, Shivaji Sarokar, Sujit Gore, Dinesh Kakuste, Suraj Chaskar, "Object Detection System for Blind People".

In these paper author tries to convey that they proposed a smart vision whose objective is to move anywhere in the environment through a user-friendly interface system. his project mainly focusses on computer vision module. In these authors made a system which will able to find the obstacle which are near to his head specially while entering from door.in short it is made to protected his head from getting injury. This product is design to navigate blind person in any environment and it guides the user about that object and provide information about that obstacle using buzzer and vibrater as a two-output mode of the user. User control mode include switch that allows the user to choose project mode of operation. There are of two mode of operation first is buzzer mode and second is vibration mode these mode are provided as an output for a blind person ,mode are used because user might not be comfortable with one of these mode .sometimes vibration motor are uncomfortable because it can irritate him with their vibration .similarly if there is a lot of noise in the surrounding buzzer cannot be used as he can't here the buzzer noise. sensor control tells whether to take the measure and receive output from the sensor and normalize it to control value for the sensor which is mounted on stepper motor .this stepper motor is continuously move in 90 degree in which it divide the image into 3 portion which are left ,right and central .for obstacle detection. So basically, when blind person is walking when optical is appeared it is sensed by sensor and tells it through one the different outputs[2]By using these papers, we can use their image processing technique .in our project since these project uses camera to detected an obstacle, we can use same technique in our object detection.

### 2.2. Amira S. Mahmoud, Sayed A. Mohamed1 Reda A. El-Khoribi2 Hisham M. AbdelSalam, "Object Detection Using Adaptive Mask RCNN in Optical Remote Sensing Images".

In this paper author has talked us about mask region-based convolutional network (Mask-RCNN) is used for utilization for multi-class object detection for sensing the images. Transfer learning, fine tuning and data augmentation are used to overcome

object scale variability, the density of object. Also, adaptive Mask-RCNN was compared to deep object detection methods. Mask-RCNN is an extended version to faster RCNN that allow an accurate pixel-based segmentation it consist of two stages feature pyramid network (FPN) and region proposal network (RPN). In feature pyramid network based on input images the different number of proposals were generated about that regions. Before that we utilized a standard convolutional neural network as a feature extractor by using art architecture AlexNet and VGGNet layers respectively. The network suffers from vanishing gradient problem which result in performance saturation and degrading so ResNet50 architecture were introduced. Which skip connection or shortcuts that allow to take activation from one layer and feed to another layer. ResNet50 uses seminal architecture to different computer application in these they have used pre trained architecture on ImageNet (1000 class). Which are small due to global average pooling rather than fully connected layers. the FPN extracts region of interest from different levels and gives as an input to RPN. RPN scanners individual and predicted whether an object is present or not .it actually scans the feature map and makes it much faster. Then each region of interest proposed by RPN as input and output a classification and bounded box and Mask-RCNN is added a new branch to output which indicate whether the given pixel is a Part or not a part of an object.[3]. From these papers we can Mask-RCNN for an object detection since Mask RCNN is faster than any deep learning object detection and provide more accurate info because of it has more added feature then faster RCNN.

### **2.3. N. Saranya, M. Nandinipriya, U. Priya, “Real Time Object Detection for Blind People”.**

In these paper author explained us about object detection from the image and represented it by their name and speech. And it also helps the blind people in location and encoded the audio into 2 channel audios with the help of 3D binaural sound. In these a video is capture with portable camera device from client side and it is streamed to a server for real time image recognition with object detection. which mean it identify and follow the same object in sequence of video frames sometimes video may have some noise. to remove that noise from frames noise reduction technique is used that improve the image quality and extraction of object frame is used to detected the object based on color of the moving frame .using different feature extraction of object from frame is used they are called object detection every object has specific feature based on shapes by using this a rectangular bounding box and centroid is plotted and position of that centroid is stored with bounded box. Now by using data streaming a pipeline is developed that enable quick communication. A raw image is taken from camera encodes it into string and send it to a client to the server. It sends information directly to the unity sound generator and play the binaural sound and output is received to unity sound generator and it plays through a wireless cord. This paper can be used in conversion of sound which will be helpful for blind person in object detection.

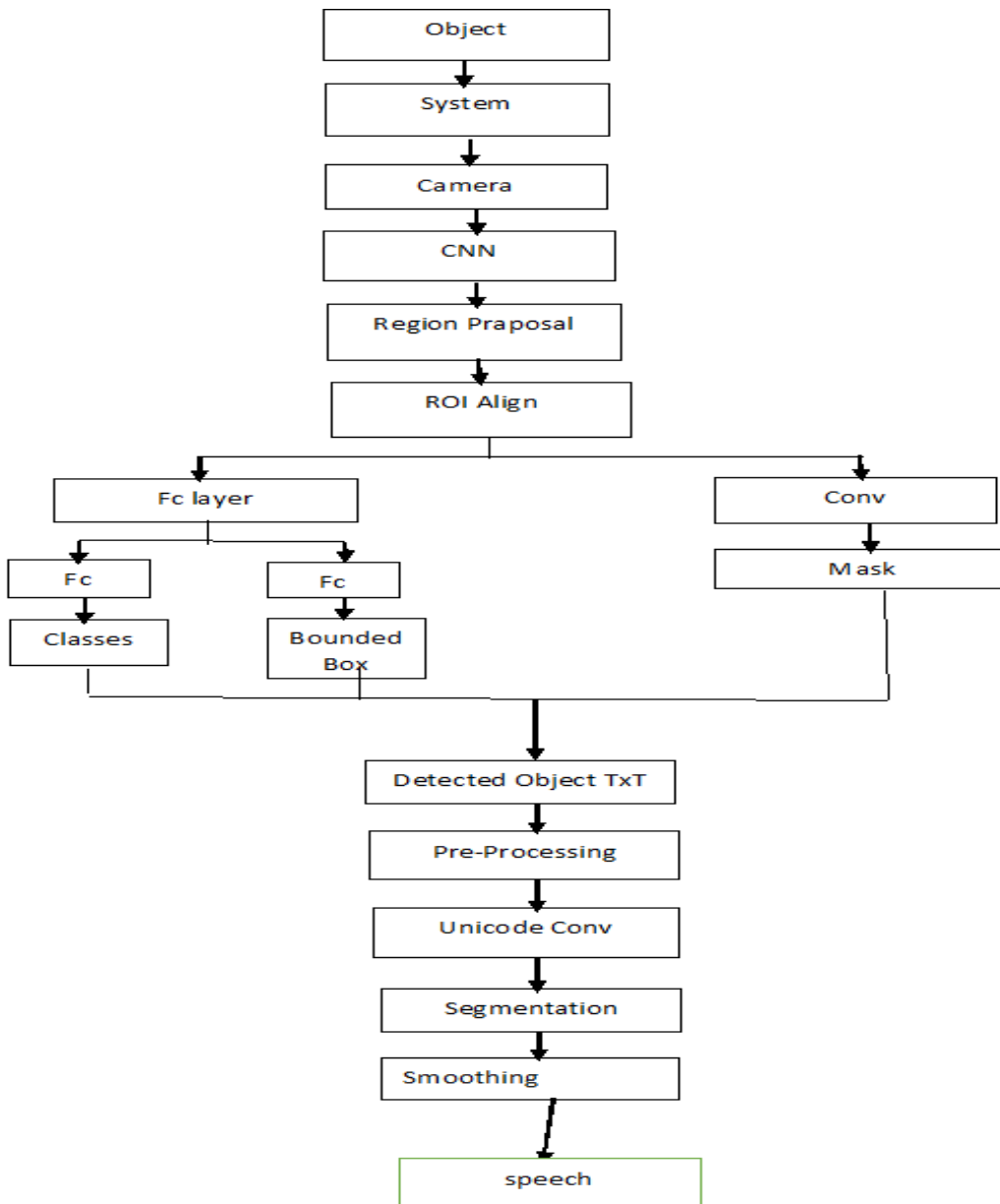
From the past event and the existing approach, the below Drawback are been noted:

1. Cost Effective System.
2. Lack of user acceptance.
3. High chances of system breaking due to hardware failure.

Considering all the drawbacks into account we have formulated a proposed system which covers all the above-mentioned drawbacks.

1. Low cost system is used.
2. No special external hardware is needed.
3. Software based system leading to low chances of complete system failure.

### 3. FLOWCHART



As shown in given figure in this image is taken from camera of an object:

Step 1. place the object inside the box.

Step 2. object will get scan by the camera which is placed inside the box.

Step 3. if object is not detected then Change the position of an object until object is detected by the camera.

Step 4. object is detected by the camera Extract region proposals using an algorithm such as Selective Search.

Step 5. Detection is identify the class id attribute of a detected object from which class that are belonging.

Step 6. After detecting frame of an object, a score is generated and bounded box is created.

Step 7. pyttsx3 lib is used to convert txt into a speech operation in which class name is identified

mask RCNN and gets converted to voice speech.

Step 8. Pyttsx3 is used for text to speech conversion in python. it is a libraries which is used for conversion. unlike other libraries it works on an office and it is compatible with python 2 and python 3.

Step 9. pyttsx3 initialize the engine after that we can set the speaking rate volume level and we can get the details of current voice.

Step 10. engine. Say() is used to output the information of currently detected objected region proposals using an algorithm such as Selective Search.

#### 4. IMPLEMENTATION

In this section the implementation of Blind-Aid application has been described. In section A of this topic, the actual methodology used is been described with respect to all the modules. Section B, are the snapshots of the application with their description showing the implemented application in detail.

[A] Methodology Used:

In these we are using a system from which blind person can place an object into that system after processing voice is generated through which blind person can identify that object. These systems are using deep learning from which a clicked image of an object is converted and transforms into speech which become handy for blind person to understand that object. Image processing is used in the following way it is done by taking an image as an input goes through Mask-RCNN procession which: it generates proposals regions of that object from a given input image which is taken from camera. It predicts the class of that object, define the bounding box and generates a mask at a pixel level of the object based on that stage proposal. with the reference images we can describe an object detection and translation.

[B] Screenshots:

System:



The above figure is the system in which it consists of a box in which camera is mounted on a top when object is placed in the box camera detected an object.

Process:

```
In [35]: capture = cv2.VideoCapture(0)

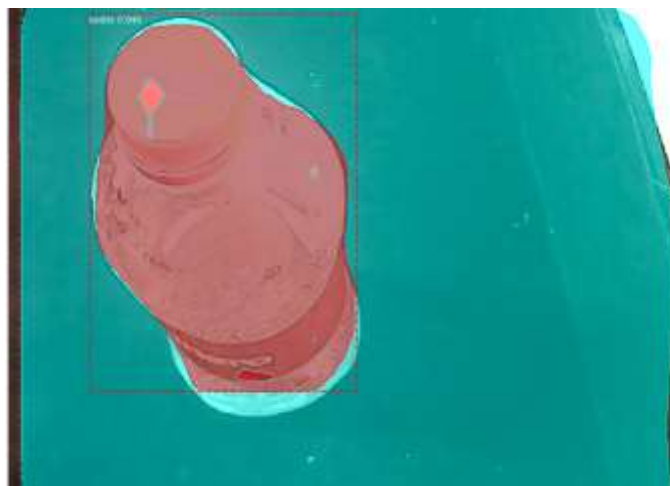
# these 2 lines can be removed if you dont have a 1080p camera.
capture.set(cv2.CAP_PROP_FRAME_WIDTH, 1920)
capture.set(cv2.CAP_PROP_FRAME_HEIGHT, 1080)

while True:
    ret, frame = capture.read()
    results = model.detect([frame], verbose=0)
    r = results[0]
    frame = visualize.display_instances(
        frame, r['rois'], r['masks'], r['class_ids'], class_names, r['scores']
    )
    class_ids=r['class_ids']
    N=r['rois'].shape[0]
    for i in range(N):
        class_id = class_ids[i]
        label = class_names[class_id]
        engine = pyttsx3.init()
        if(label=='bottle'):
            engine.say("A plastic bottle is a bottle")
            engine.say(label)
            engine.runAndWait()
    cv2.imshow('frame', frame)
    if cv2.waitKey(1) & 0xFF == ord('q'):
        break

capture.release()
cv2.destroyAllWindows()
```

Above figure shows how object detection and translation works in these OpenCV uses a videocapture() in these it captures an image of that object and by using frame which has rois, masks, class\_ids, scores object is detected and masked. And By using pyttx3 lib we can translate class\_ids object into speech by using engine.say().

Detected Object:



This is image of an object detected that is bottle in which it shows score. Score can range from 0 to 1. Bigger the scores more the accurate the result.

### 5. FUTURE SCOPE

In future scope of this application, the System will detect more object that are used daily. It will increase its accuracy so to get better result and it will also consider complex shape which was difficult by the blind person to detected. Since this project is using a system which is stationary by using smartphone, we can overcome this problem. smartphone is portable so it come easy to carry.

## 6. CONCLUSION

The Project entitled “Object Detection and Translation for Blind People Using Deep Learning” has been developed and this satisfies all proposed requirements. The system is highly usable and user friendly. All the system objectives have been met. All these phases of development are done according to methodologies. The application will execute successfully by fulfilling the objectives of the project. Further extensions to this application can be made as required with minor modifications.

## ACKNOWLEDGEMENT

We are pleased to present “Object Detection and Translation for Blind People Using Deep Learning” as our project and take this opportunity to express our profound gratitude to all those people who helped us in completion of this project.

We thank our college for providing us with excellent facilities that helped us to complete and present this project. We would also like to thank the staff members and lab assistants for permitting us to use computers in the lab as and when required.

We express our deepest gratitude towards our project guide Prof. Ichhanshu Jaiswal for his valuable and timely advice during the various phases in our project. We would also like to thank him for providing us with all proper facilities and support as the project coordinator. We would like to thank him for support, patience and faith in our capabilities and for giving us flexibility in terms of working and reporting schedules.

Finally, we would like to thank everyone who has helped us directly or indirectly in our project.

## REFERENCES

- [1] Facebook AI Research (FAIR), Kaiming He Georgia Gkioxari Piotr Dollar Ross.
- [2] Prof. Seema Udgirkar, Shivaji Sarokar, Sujit Gore, Dinesh Kakuste, Suraj Chaskar, Object Detection System for Blind People.
- [3] Object Detection Using Adaptive Mask RCNN in Optical Remote Sensing Images, Amira S. Mahmoud, Sayed A. Mohamed1 Reda A. El-Khoribi2 Hisham M. AbdelSalam.
- [4] Real Time Object Detection for Blind People, N.Saranya , M.Nandinipriya , U.Priya
- [5] Convolutional Neural Network for Object Detection System for Blind People, Y.C. Wong, J.A. Lai, S.S.S. Ranjit, A.R. Syafeeza, N. A. Hamid
- [6] Object Detection and Recognition for Visually Impaired People, Shuihua Wang
- [7]medium.com\_@alittlepain833/simple-understanding-of-mask-rcnn , Xiang Zhang
- [8] TEXT TO SPEECH CONVERSION MODULE, Hussain Rangoonwala , Vishal Kaushik , P Mohith and DhanalakshmiSamiappan