e-ISSN: 2582-5208

International Research Journal of Modernization in Engineering Technology and Science

(Peer-Reviewed, Open Access, Fully Refereed International Journal)

Volume:07/Issu:04/April-2025 Impact Factor- 8.187 www.irjmets.com

EMPOWERING THE VISUALLY IMPAIRED WITH REAL-TIME AUDIO FEEDBACK ON OBJECTS AND ENVIRONMENT

Mr. Rahul Dhokane[*1], **Rohit Sanap**[*2], **Vaishnavi Ghuge**[*3], **Pooja Jadhav**[*4], **Ankita Kangane**[*5]

[*1]Assistant Professor, Department Of Information Technology, Sir Visvesvaraya Institute Of Technology, Nashik, Maharashtra, India.

[*2,3,4,5]Department Of Information Technology, Sir Visvesvaraya Institute Of Technology, Nashik, Maharashtra, India.

ABSTRACT

Vision is one of the very essential human senses and it plays the most important role in human perception of our environment, unfortunately there are many people who are visually impaired. Blind people today rely on sighted guides, seeing-eye dogs and canes even a century after these came into existence. The present work aims to aid the blind through a wrist wearable. The wearable is designed to capture the user's environment through a camera and recognize the objects present in image. These identified objects are informed to user through an audio output.

The object recognition is achieved through OpenCV.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.

In addition to object detection, the device can adapt to different environments and user preferences, offering personalized feedback. The system's accuracy and reliability are bolstered by extensive use of diverse datasets and real-time processing capabilities, ensuring the wearable remains effective in a wide range of real-world scenarios. This innovation has the potential to enhance the independence of visually impaired individuals, providing them with more autonomy and a greater sense of security while navigating unfamiliar spaces.

Keywords: Python, Machine Learning, YOLO lib. Datasets, Opencv.

I. INTRODUCTION

Blind people lead a normal life with their own style of doing things. But, they definitely face troubles due to inaccessible infrastructure and in new environment. It is difficult for others to help the visually impaired all the time. In 2015, there were an estimated 253 million people with visual impairment worldwide. Of these, 36 million were blind and a further 217 million had moderate to severe visual impairment (MSVI). By 2020, it is projected to 76 million people. The need for identifying objects in the surroundings among blind people and a broader look at the advanced technology available in today's world is the reason to develop this project. Object recognition is one of the fundamental tasks in computer vision.

It is the process of finding or identifying instances of objects (for example faces, dogs or buildings) in digital images or videos. Globally, the causes of vision impairment are cataract, uncorrected refractive errors, trachoma, diabetic retinopathy, corneal opacity, age related muscular degeneration and eye injuries. It limits visually impaired people to navigate and perform everyday tasks. Advancement in technologies such as hand gesture, text recognition, Eye-ring project are came up with solutions. However, these solutions have some disadvantages like expensive, heavyweight, less accuracy, less speed, etc. So, the objective of the project is to design and implement a system that captures the user's environment through a camera module. The images that are captured are transformed to Raspberry pi and it is analyzed using Haar Cascade and YOLO Model. Then the recognized objects are given to the user via audio through a speaker or Bluetooth headset. The image data is transferred to data server for further use.

Despite advancements in technology, many existing tools still remain expensive, bulky, or impractical for everyday use by visually impaired individuals. For example, while some wearable devices and smartphone applications attempt to assist blind users with navigation, they often require specialized hardware or fail to deliver accurate, real-time object detection. The reliance on guides, canes, and seeing-eye dogs remains common, but these solutions are limited in terms of providing detailed feedback about the surroundings, especially in dynamic environments.

The objective of this project is to develop a software-based system that leverages the mobile phone camera to assist blind people in navigating their environment. The system utilizes the smartphone's camera to capture the user's surroundings, with images processed using object detection algorithms such as YOLO (You Only Look Once) and Haar Cascade. These algorithms are employed for efficient real-time object identification. The identified objects are then communicated to the user via audio feedback, either through the phone's speaker or a Bluetooth headset, providing useful, immediate information to help navigate both indoor and outdoor environments. This system is completely software-based, relying on the power of mobile devices to offer a lightweight, affordable, and easily accessible solution for visually impaired individuals, ultimately enhancing their independence and overall safety in various environments.

II. 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.

III. PROPOSED SYSTEM

The main motive of the proposed system is to assist visually impaired persons by providing the perception of the environment, which helps them in avoiding obstacles or barriers and in moving from one place to another. The goal to provide a simple, user-friendly and handy solution is achieved. The proposed system is capable of detecting the objects present in the surrounding environment with good speed and accuracy. It detects objects effectively in both outdoor and indoor environment. The system is able to successfully detect the multiple objects present in the surrounding environment and communicate the same to the user in audio through headphones or speaker. The proposed system is tested in detecting objects in indoor environment, outdoor environment and objects which are more than 10 m from the camera. The system is capable of detecting the objects in the surrounding environment and provide audio output to the user. The performance of the application in the above mentioned three categories is satisfactory.

Object Detection: In this phase are very easy to detect the real world objects. Like humans in still images or Videos. Below diagram are the examples of the how to object are detected easily.

There can also be obstacles with no disappearing points, for which default hypothesis of bottom to top depth gradient is used. Depending on the position of these disappearing points, a combination of bottom to top depth hypothesis and the depth hypothesis corresponding to the disappearing points is used for assigning depth values for the obstacles of the image. Hence depth is estimated for the separated obstacles regaining the variation of depth values within the same obstacle. The depth map estimated for the obstacles detected.

Furthermore, the system has been rigorously tested across various scenarios, including detecting objects in both indoor and outdoor environments, in low and high-light conditions, and at distances greater than 10 meters. The results demonstrate the system's robustness and versatility, with satisfactory performance in real-world applications. The accuracy and efficiency of the object detection algorithms, combined with real-time audio feedback, ensure that visually impaired individuals can confidently navigate a variety of environments. The system also features continuous updates and improvements, allowing it to adapt to different contexts and maintain optimal performance over time.

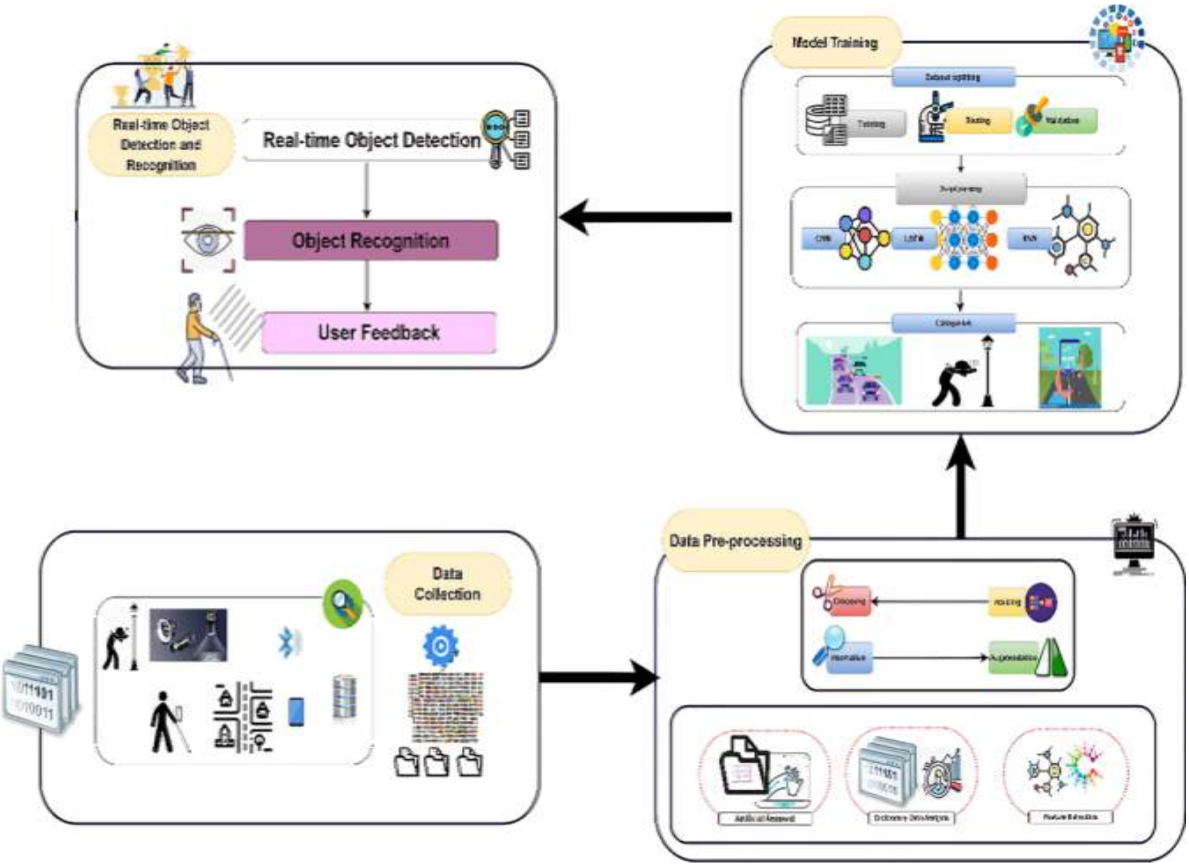

Fig 1: System Architecture

IV. ALGORITHM

4.1 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. The prediction of bounding box is illustrated below.

4.2 COLLECTION OF DATA BASE:

For any machine learning model, the collection and preparation of data is a crucial step. The performance and effectiveness of the object detection system heavily depend on the quality and diversity of the dataset used for training. In this system, we collect a large set of images that represent various objects in different environments, lighting conditions, and angles. This diverse dataset ensures that the model can generalize well when it encounters new, unseen environments.

The dataset can include images taken from various sources such as open datasets (e.g., COCO, ImageNet, Open Images) or custom data gathered specifically for the purpose of object detection for the blind. For example, images might include common objects like chairs, doors, trees, cars, and people, and may feature different indoor and outdoor settings.

Each image in the dataset is annotated with information about the object it contains. This process is called labeling, and it involves drawing bounding boxes around objects in the image and assigning them a label (such as "car," "dog," "person," etc.). Tools like LabelImg or RectLabel are typically used for this task. Proper labeling is vital because the algorithm will use these labels to learn how to distinguish between different object classes.

In addition to annotating images, a dataset should ideally contain images that vary in:

- **Lighting conditions**: To simulate real-world scenarios where lighting may not always be ideal.
- **Backgrounds**: The objects in the dataset should appear on various backgrounds to avoid the model overfitting to a single type of environment.
- **Angles and views**: Objects should be captured from different angles and perspectives to improve the model's robustness.
- **Distance**: Objects should be present at different distances from the camera, as it will affect how the model detects them.

4.3 NETWORK OUTPUT ANALYSIS:

After training the YOLO model on the dataset, the network generates output during inference, which consists of predicted bounding boxes, class labels, and associated confidence scores for each detected object in an image.

The **output analysis** phase is where the system interprets and processes these results. The YOLO algorithm divides the image into a grid, and each grid cell predicts a bounding box and class probability for each object within it. For every bounding box predicted by a grid cell, the system computes a confidence score that reflects the model's certainty that the bounding box corresponds to an actual object.

Key elements of the network output include:

1. **Bounding boxes**: These are the rectangular boxes predicted around objects in the image. The coordinates of the bounding box (center, width, height) are calculated relative to the image's dimensions.
2. **Class labels**: The system assigns a label to each bounding box based on the type of object detected (e.g., "cat," "car," "tree").
3. **Confidence scores**: Each bounding box has an associated confidence score, which is a measure of how confident the model is that the box contains the object and that the prediction is accurate. The higher the score, the more confident the model is in its prediction.

During this phase, the system analyzes the predictions for each object, filtering out irrelevant or low-confidence predictions. This analysis helps ensure that only the most accurate object detections are considered, eliminating false positives or low-quality detections.

4.4. BOUNDING BOX DIMENSIONS:

Bounding boxes are crucial in object detection as they provide a spatial representation of where the object is located in the image. The dimensions of the bounding box are calculated based on the detected object and represent the area of the image where the object is found.

In the YOLO algorithm, each grid cell in the image predicts multiple bounding boxes, and for each bounding box, it predicts the following:

- **Center coordinates** (x, y): The center of the bounding box, relative to the grid cell.
- **Width (w) and Height (h)**: The size of the bounding box. The system uses the width and height of the bounding box to estimate the object's size.
- **Confidence score**: Reflects how likely the bounding box contains an object.
- **Class probabilities**: Indicating the likelihood of the object being of a particular class.

The bounding box is defined by its top-left and bottom-right corners (x1, y1, x2, y2), which are calculated from the predicted center coordinates, width, and height. These values are scaled to the dimensions of the image, so they can be drawn on the image during visualization.

4.5 OUTPUT FILTERING:

Output filtering is a critical step in improving the accuracy and reliability of the object detection system. After the YOLO model generates its predictions, it may produce multiple bounding boxes for the same object, especially if that object spans multiple grid cells or if it's detected by different grids. The goal of output filtering is to eliminate unnecessary predictions, focusing only on the most accurate ones.

Common output filtering techniques include:

- **Non-Maximum Suppression (NMS)**: This technique is used to remove redundant bounding boxes that overlap significantly. NMS works by selecting the bounding box with the highest confidence score and discarding the other boxes that overlap with it by more than a certain threshold (usually 0.5). This helps ensure that only one bounding box is selected per object.
- **Thresholding**: Bounding boxes with confidence scores below a certain threshold are discarded. For example, if a bounding box has a confidence score of less than 0.5, it may be discarded because the model is not confident enough in the detection.

Output filtering helps reduce false positives and ensures that only the most relevant and accurate objects are detected and passed on to the user.

V. COMPONENTS USED

Fig 2: Raspberry pi 4b Module

In an object detection project for visually impaired people, a Raspberry Pi is used as the central processing unit to handle tasks such as capturing images, processing data, and providing feedback. The Raspberry Pi can be connected to a camera module to capture live video or images from the surroundings. It then uses machine learning algorithms (often with pre-trained models like TensorFlow or OpenCV) to detect objects in real-time. Once an object is detected, the Raspberry Pi can provide audio feedback (via speakers or a headphone) to inform the user of the object's presence, location, and distance, helping them navigate their environment safely. Its compact size, affordability, and ease of integration make Raspberry Pi an ideal choice for such assistive technology projects.

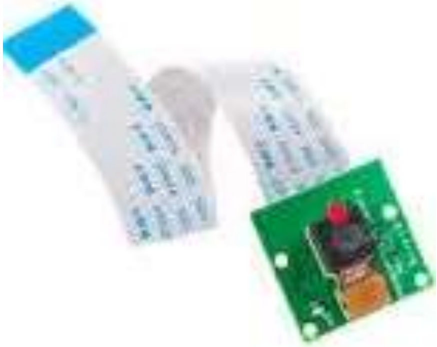

Fig 3: Raspberry pi Camera Module

The **Raspberry Pi Camera Module** is a small camera that attaches to a Raspberry Pi. It captures pictures and videos, which are useful in projects like helping visually impaired people. The camera takes images of the surroundings, and the Raspberry Pi processes them to detect objects. It then gives audio feedback to the user, letting them know what's around them. It's small, easy to use, and perfect for projects where space and power are limited.

VI. CONCLUSION

In this project we present a visual system for blind people based on object like images and video scene. This system uses Deep Learning for object identification. In order to detect some objects with different conditions. Object detection deals with detecting objects of inside a certain image or video. The TensorFlow Object Detection API easily create or use an object detection model Blind peoples they have a very little information on self-velocity objects, direction which is essential for travel. The navigation systems is costly which is not affordable by the common blind people. So this project main aim is to the help of blind people.

VII. 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.