Artificial intelligence enabled smart glove for visually impaired

Jagdeep Rahul
Department of Electronics Communication Engineering,
Rajiv Gandhi University,
Doimukh, India
jagdeeprahul11@gmail.com

Ashutosh Kumar
Department of Electronics Communication Engineering,
Vellore Institute of Technology,
Amrawati, India
ashutosh.kumar007@gmail.com

Lakhan Dev Sharma
Department of Electronics Communication Engineering,
Vellore Institute of Technology,
Amrawati, India
devsharmalakhan@gmail.com

***Abstract*— Object detection is a technique to tag objects present in the frame of an image, video sequence, and real-time video. In recent years, the world has been reshaped around deep learning algorithms. This paper makes the user aware of the obstacles present in his environment. There are two fundamental parts in this paper: the software part and the hardware part. The state-of-the-art You Look Only Once (YOLO) algorithm was applied in the present work for object identification. The overall analysis shows that this algorithm produces accurate results for real-time object detection and can be considered faster object identification.**

***Keywords*— Tensor flow, Object detection, Real-time video YOLO, Raspberry pi 3.**

I. Introduction

The world we know is mainly designed keeping in mind the comfort of people without any visual problems. As a result, this creates a gap in communication for visually impaired people. Around 37 million individuals worldwide are blind or visually impaired, accounting for a significant part of the global population. Because of their handicaps, these visually challenged people encounter several challenges in their daily lives. They have the same right to education and information as everyone else. However, their capacity to learn new things is severely limited [1],[13-18]. They must rely on conventional and orthodox means of obtaining information, such as tactile sensations like finger Braille, manual alphabets, and the print on palm technique, but these methods are arduous, slow, and inefficient, and are not suitable for usage in a computer context. As a result, they are unable to access the internet, which is the information hub [2]. The proposed model helps to solve the object detection and identification problem for visually impaired people both indoors and outdoors. This project also adds a security layer for these specially-abled people.

II. Literature Review

Many researchers have proposed various models for helping visually impaired people, and few of them have a very good extent to abolish these problems.

P. Angelov et al. have proposed a model for real time independent object detection and tracking [3] was based on object detection and object tracking with color specific implementation. It was a visual project where video/image data was applied as an input into the model using a webcam with the Raspberry Pi connected. The model proposed by the Aydin Akan et al. [4] for obstacle detection and tracking for visually impaired people on sidewalks . A proposed concept for an obstacle detection and warning system on sidewalks was designed to assist visually impaired persons. On the hardware side, this model was created using a Raspberry Pi and three vibration sensors. The speaker was utilised as an output to communicate the object's name to the user. Python was selected as the programming language because to its speed, compatibility with the Raspberry Pi, and availability of packages. Tensorflow, a Python tool, was utilised to conduct object identification and classification using a pre-trained model in this study.

The method proposed by *Amir et al.* [5] was the *Implementation of the Line Tracking Algorithm using the Raspberry Pi*. The hardware implementation of the model was done on the Raspberry Pi due to its optimum size. The accuracy of the model was sufficient, although the FPS rate was not enough. The work done b*y Srivastava et al.* [6] was a prototype of a hand glove for visually impaired people which converts the OCR text to speech. Problems can be reduced to some extent. MATLAB and Arduino were the two software tools used to implement this work. The image processing operations were done using MATLAB, while the ultrasonic sensor programming was done with Arduino. The effort then moved on to obstacle detection, picture to text conversion, and finally text to voice translation.

978-1-6654-4577-1/21/$31.00 ©2021 IEEE

In another work a smart glove was developed to assist the visually impaired people in [7]. A prototype of the smart glove was designed utilising a deep neural network and an object tracking algorithm. It is capable of directing the blind to the desired item in an interior setting. A USB camera is included in the palm of the glove, which transmits live video to the Raspberry Pi for processing. Its entire implementation cost was in the neighbourhood of sixteen thousand dollars, which is fairly high. It also made use of Intel movidius NCS, which distinguished it from the previous work. Our suggested model is adaptable to practically any maritime environment and may be used to accomplish the objective for which it was developed.

The suggested method entails creating a portable system capable of labelling objects using the OpenCV and Tensorflow frameworks. The second component of this study is the translation of tagged text to voice and then the output in the form of auditory signals to make the blind person aware of the item or barrier in front of him. This model consists of a hand glove attached to a pi cam. The camera captures real-time video and the algorithm performs object detection. It gives the output to the user as an audio signal through a speaker/earphone. This work is also implemented using the Raspberry Pi (version 3) platform, which makes the model portable. The model detects the object from a predefined and trained class that contains a particular label. This model is also capable of detecting objects from an image or from a video. Hence, this model can contribute greatly to the mobility and safety of visually impaired people.

.

III. PROPOSED METHODOLOGY

In the proposed work, the YOLO algorithm has been used, considering its advantages above all other algorithms.
-YOLO stands for You Look Only Once.

-This algorithm outperformed all of them.

-It divides the image into certain grid cells.

-The output is based on a confidence score.

-The limitation of YOLO is that it struggles with small images.

The object detection and categorization method You Only Look Once (YOLO) is a helpful one-stage technique. YOLO is an object identification method that employs just one neural network. Unlike previous object detection algorithms, which sweep the picture bit by bit, this approach analyses the entire image and reframes object detection as a single regression issue, going straight from image pixels to bounding box coordinates and class probabilities. The flow chart diagram in fig.1 depicts the steps that we have used during implementation of this work.

The proposed system consists of the following steps given below.

1. Real-time video is given as input

2. Image is captured for each frame

3. The captured image is compared to the pre-existing trained data set with the help of the YOLO model

4. The converted image is then further processed for text to speech conversion using the GTTS library.

5. In the last step, the converted speech is given as output with the help of an earphone/speaker

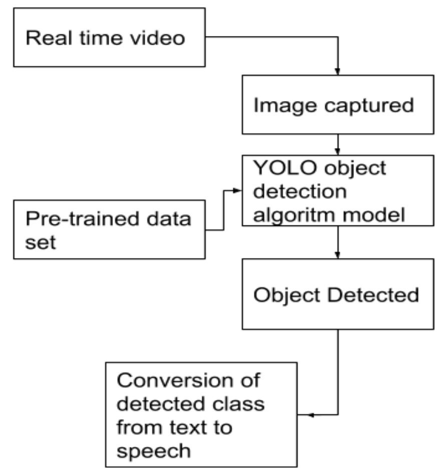

Fig. 1 Flow chart of the proposed model

The entire system can be divided into two sections:

1. Hardware section
2. Software section

1. Hardware section

Raspberry Pi 3:

The Raspberry Pi [8] is a very small, low-cost, and low-powered system on a chip (SoC).It has significantly less processing power than a regular PC or even most modern smartphones, but due to its versatility and cost, it has become popular even outside the initial target audience. The Raspberry Pi 3, which has been used here, boasts some new and improved features, but uses the same SoC as the Model B, at slightly higher clock rates, but with the same amount of RAM. The systematic diagram of the Raspberry Pi 3 is given in fig.2.

Table 1: Specifications of Arduino

Parameter	Specifications
Raw Voltage input	5V, 2A power source
Maximum total current drawn from all I/O pins	54mA
Clock Frequency	1.2GHz

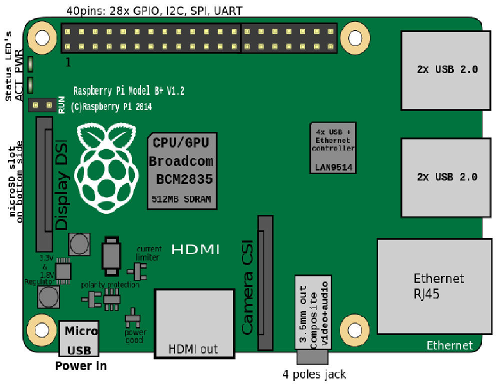

Fig. 2. Raspberry Pi 3 model.

Pi Camera :

Pi camer [9] is used in this work to get a real-time video of the environment. Other cameras could also be used, but the pi cam is more portable with a raspberry pi, so we have used it for this work. The specifications of the used Arduino are given in table 1, and the used Pi camera is shown in fig 3.

.

Fig. 3 Pi Camera used in model.

Audio Output -

The detected object is converted from text to speech, to get audio output from the used speaker. We have used a bluetooth speaker particularly for demonstration purposes to make this model more handy and portable earphones could also be used.

2. Software section:

The algorithm code for object detection was written in the Python language. There were a series of steps involved, from writing the code to implementing it. In the first step, all the required Python packages were installed, such as YOLO packages, num pi, openCV [10], and tensor flow for running a shell script command in a software environment. The installation of all the libraries was performed in a sequential manner, one by one. If any image is captured by the Raspberry Pi webcam, the width and height of the image are calculated. Based on this height and width, several classes are created, such as boxes, classes, and scores. That boxes were used to particularly bind the objects detected with random colours, and classes were used to name the objects detected, while scores were used to give the confidence value. This whole process gets calculated in a loop for all the detected objects. One key point of this model is that any object is detected only when the confidence value is greater than the threshold value (in this case, the threshold value assigned is 60%).

To convert the detected object into speech, a separate class was created in the code. The Python library used for converting text to speech is gTTS [11]. A new object is created which contains the object to be detected (having an accuracy equal to or greater than 60) along with the language in which the audio will be played and the pace at which the audio is going to play as an argument. The audio after getting converted from text gets stored as an mp3 file which gets played after each image detection.

IV. RESULT AND DISCUSSION

The results were obtained after implementation of the above algorithm for object detection. The steps of the proposed model are presented in fig. 4. This model can work for three kinds of input:-

a.) Image (.png)
b.) Video
c.) Real-time video

Fig. 4. Proposed model step

a) The portable network graphics (.png) file was given as input. The screenshots of the obtained results are shown in fig.5.

.

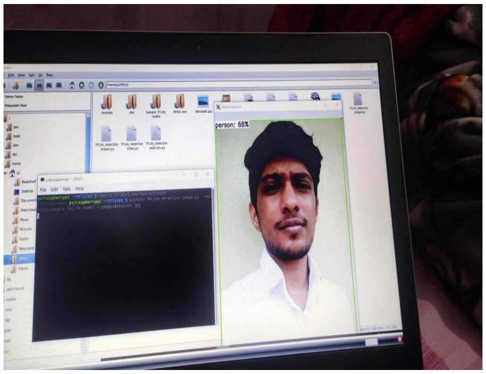

Fig. 5. Detected object in uploaded image.

1.) *Video as an input:* - The uploaded video is shown in fig.6

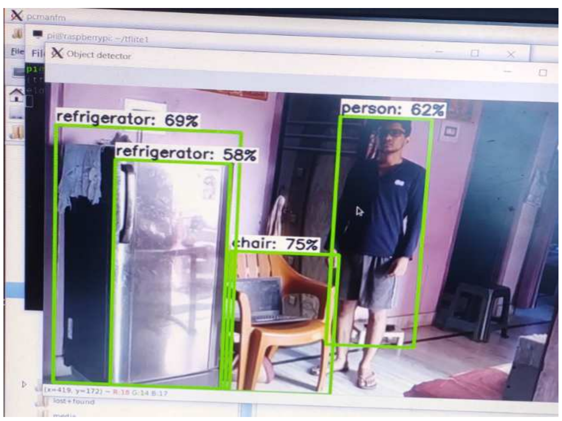

Fig. 6. Detected object in a uploaded video.

2.) *Real-time video as an input*

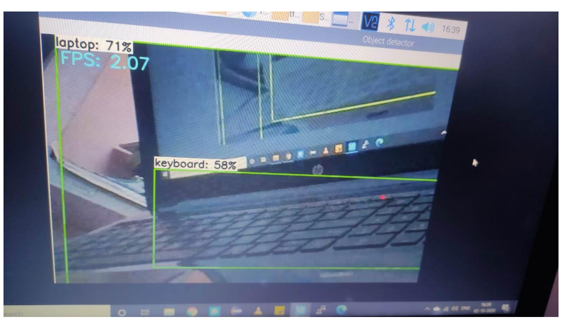

Fig. 7. Detected object during real-time video.

The proposed model worked completely well, satisfying all the conditions. However, real-time video processing took a bit of extra time to process when compared to the feed video. As mentioned above all the results obtained were as expected.
The proposed model was tested on all three given input formats, where the achieved results were found satisfactory. The speaker was connected to the Raspberry Pi with Bluetooth. The speaker produces the output as soon as the confidence value score crosses 60%. The limit was given as such to increase efficiency. However, there is sometimes negligible audio lag when compared to each frame processed.
The proposed model was further compared with the existing models to validate the performance of the model. The method given in [3] only uses the YOLO algorithm and neither portable nor audio as an output. The methods proposed in [4] and [6] were capable of producing audio as an output, but neither the portable nor the YOLO algorithm used there. The models in [5] and [12] were portable, but the YOLO algorithm was not used there. Our proposed model produces the audio as an output, is portable, and uses the YOLO algorithm to achieve object identification. The overall comparison of the proposed model was performed in table 2.
Table 2. Comparison of the models

S. No.	Existing Model	Authors name	Audio Output	Portable	Yolo algorithm Used
1	Real time approach to autonomous novelty detection and object tracking	P. Angelov [3]	No	No	Present
2	Designing an obstacle detection and alerting system for visually impaired people on sidewalk	Aydin Akan [4]	Yes	No	No
3	NETRA(2019)	N.K. Srivastava[6]	Yes	No	No
4	Line tracking algorithm using Raspberry Pi	Samreen Amir [5]	No	Yes	No
5	Object detection and human identification using Raspberry Pi	R.S.M. [12]	Yes	Yes	No
6	Proposed Model	J. Rahul	Yes	Yes	Yes

V. CONCLUSION

A prototype of an AI-enabled Smart Glove has been designed which is effective in helping the navigation of visually impaired people. It detects any object indoors as well as outdoors with the help of a trained data set. The device is compact and portable. The accuracy obtained using the YOLO algorithm is comparatively higher when compared to other algorithms like CNN and RCNN. When compared to other existing works, it also takes less time to process the video file. We have observed better accuracy when the image is given as an input as compared to that of video and real-time video. The idea of detecting the object only after it reaches a minimum threshold value makes the model more precise and accurate, hence avoiding any false detections.

REFERENCES

[1] Bandodkar, Mukul, and Virat Chourasia. "Low cost real-time communication braille hand-glove for visually impaired using slot sensors and vibration motors." *International Journal of Information and Communication Engineering* 8.6 (2014): 973-980.

[2] Linvill, John G., and James Charles Bliss. "A direct translation reading aid for the blind." *Proceedings of the IEEE* 54.1 (1966): 40-51.

[3] Angelov, Plamen, Pouria Sadeghi-Tehran, and Ramin Ramezani. "A real-time approach to autonomous novelty detection and object tracking in video stream." *International Journal of Intelligent Systems* 26.3 (2011): 189-205.

[4] Pehlivan, Sude, Mazlum Unay, and Aydin Akan. "Designing an obstacle detection and alerting system for visually impaired people on sidewalks." *2019 Medical Technologies Congress (TIPTEKNO)*. IEEE, 2019.

[5] Amir, Samreen, et al. "Implementation of line tracking algorithm using Raspberry pi in marine environment." *2014 IEEE International Conference on Industrial Engineering and Engineering Management*. IEEE, 2014.

[6] Srivastava, N. K., and Satyam Singh. "Netra: Smart Hand Gloves Comprises Obstacle Detection, Object Identification & OCR Text to Speech Converter for Blinds." *2018 5th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*. IEEE, 2018.

[7] Jain, Sambhav, et al. "Design and implementation of the smart glove to aid the visually impaired." 2019 International Conference on Communication and Signal Processing (ICCSP). IEEE, 2019.

[8]Raspberry:https://www.raspberrypi.org/products/raspberry-pi-3-model-b.

[9] Picam:https://www.raspberrypi.org/products/camera-module-v2/.

[10] Guennouni, S., Ahaitouf, A., & Mansouri, A. (2014). Multiple object detection using OpenCV on an embedded platform. 2014 Third IEEE International Colloquium in Information Science and Technology (CIST).

[11] Gtts: https://cloud.google.com/text-to-speech/docs/apis.

[12] Khan, Fathima, Gowtham Ghatge, and S. Hemaya. "Object detection and human identification using raspberry pi." *2019 1st International Conference on Advances in Information Technology (ICAIT)*. IEEE, 2019.

[13] Rahul, Jagdeep, Marpe Sora, and Lakhan Dev Sharma. "Exploratory data analysis based efficient QRS-complex detection technique with minimal computational load." *Physical and Engineering Sciences in Medicine* 43.3 (2020): 1049-1067.

[14] Rahul, Jagdeep, et al. "An improved cardiac arrhythmia classification using an RR interval-based approach." *Biocybernetics and Biomedical Engineering* 41.2 (2021): 656-666.

[15] Rahul, Jagdeep, Marpe Sora, and Lakhan Dev Sharma. "A novel and lightweight P, QRS, and T peaks detector using adaptive thresholding and template waveform." *Computers in Biology and Medicine* 132 (2021): 104307.

[16] Rahul, Jagdeep, Marpe Sora, and Lakhan Dev Sharma. "Dynamic thresholding based efficient QRS complex detection with low computational overhead." *Biomedical Signal Processing and Control* 67 (2021): 102519.

[17] Rahul, Jagdeep, Lakhan Dev Sharma, and Vijay Kumar Bohat. "Short duration Vectorcardiogram based inferior myocardial infarction detection: class and subject-oriented approach." *Biomedical Engineering/Biomedizinische Technik* (2021).

[18] Rahul, Jagdeep, and Lakhan Dev Sharma. "Artificial intelligence-based approach for atrial fibrillation detection using normalised and short-duration time-frequency ECG." *Biomedical Signal Processing and Control* 71 (2022): 103270.