ISSN: 2633-4828 Vol. 5 No.4, December, 2023

International Journal of Applied Engineering & Technology

DESIGN AND IMPLEMENTATION OF REAL TIME ASSISTIVE DEVICE FOR BLIND PEOPLE USING OBJECT RECOGNITION

Hawraa Kamil Abbas and Ammar Ibrahim Majeed
College of Information Engineering, University of Al-Nahrain, Baghdad, Iraq

ABSTRACT

This paper presents the design and implementation of a real-time assistive device using object recognition for visually impaired people face difficulties in safe and independent movement which deprive them from regular professional and social activities in both indoors and outdoors. Similarly they have distress in identification of surrounding environment fundamentals. Computer vision technologies, particularly the Deep Convolutional Neural Network, were developed rapidly in recent years.

The proposed device utilizes a Raspberry Pi board coupled with the You Only Look Once (YOLO) object recognition algorithm to recognize objects in the surrounding environment.

The object recognition yolo algorithm trained on the Common Objects in Context (COCO) dataset which is a large dataset of images, and it is capable of detecting objects such as chairs, tables, doors, and other obstacles to identify the object present before the person. Algorithm is trained the device then provides voice output through an audio interface to warn the user about the presence of objects in their path. The device is relatively low-cost, portable, and can be easily carried by the user. The proposed solution is expected to improve the mobility and independence of the blind people by providing a real-time assistance system.

Keywords: Object Detection, COCO, YOLO, CNN, Computer Vision, Raspberry PI, visually impaired people, voice output.

1. INTRODUCTION

Millions of people worldwide suffer from vision impairments that make it difficult for them to comprehend their surroundings. According to estimates from the World Health Organisation, 285 million people worldwide suffer from visual impairments; of them, 246 million are losing their ability to see well, and 39 million are blind (World health organisation, 2019). They can do everyday work with a variety of tactics, but they also have trouble navigating and acting awkwardly in social situations. People with visual impairments have difficulties moving around safely and independently, which makes it difficult for them to engage in normal social and professional activities both indoors and outdoors. In a similar vein, they struggle to recognise basic elements of their environment.

For those who are blind or visually impaired It might be quite difficult for them to move around comfortably and to identify the people and things in their environment, both inside and outside. They could run into a lot of adjacent barriers and have trouble differentiating items from individuals. In locations with a lot of car and pedestrian traffic, they will also encounter a lot of obstacles including potholes, bumps, power pylons, signboards, etc (Duman et al., 2019).

As a result of these problems, several assistive gadgets have been developed to help those who are visually impaired. Among these are methods for computer vision. Making a computer or other machine able to perceive like humans do is the aim of computer vision. A visual person uses both their eyes and their brain to understand their environment. Humans are able to locate items in their surroundings and to detect, recognise, and understand those objects' whereabouts. It is possible to equip a computer with comparable capabilities using a variety of hardware and software tools [3]. This makes it possible to create assistive technology for the blind.

In recent years, there has been a lot of interest in deep learning-based object identification techniques. Many object detection techniques based on convolutional neural networks (CNNs) have been suggested recently as a

result of deep learning technical developments. Strong CNN-based algorithms have significantly increased the accuracy of object detection.

Techniques that employ complex structures to enhance network representation capabilities include Regions with Convolutional Neural Networks (R-CNN), Fast Regions with Convolutional Neural Networks (Faster-RCNN), Single Shot Multi-Box Detector (SSD), and Faster Regions with Convolutional Neural Networks (Faster-RCNN) (Sharma et al., 2021). These exceptional CNN-based detectors provide state-of-the-art item detection accuracy, but they are challenging to integrate into mobile devices for practical uses. Neural networks rely heavily on powerful graphics processing units (GPUs) to complete the inference process, which is one of the primary causes as they include a lot of parameters and require a lot of processing power (Mounir ET AL., 2021).

It is so challenging to construct these networks on embedded devices with constrained memory and processing power. Due to this limitation, standalone terminal detection devices that need near-real-time detection cannot generally use these high-performance networks (Roy et al., 2020).

Consequently, CNN-based object identification neural networks that are small and light were needed for edge devices with constrained processing power (Peeples, 2020).In order to lower computing costs, several lightweight single-shot deep-learning-based detectors have been created recently. Among them is the Yolo series, an inventive and endearing family of object identification networks that significantly lowers processing expenses and model size. Additionally, there are several variants (Yi et al., 2019). This study aims to develop and build an assistive system for visually impaired individuals. The system consists of two components: software and hardware. A Raspberry Pi board and a single camera are used to create the hardware. The programme makes use of bounding-box based distance estimate techniques, such as random forest regression, and convolutional neural network (CNN) based object identification approaches, such as YOLO. In addition, the system may provide blind users with guidance by producing auditory alerts on item names and distances.

The structure of the paper is as follows. The literature's most closely connected works are surveyed in section 2. The design of the suggested system and its modules are displayed in Section 3. The suggested system is implemented and the outcomes of the experiments are displayed in Section 4. The conclusion and potential future projects are discussed in the final part.

2.RELATED WORK

Convolutional neural networks and computer vision technologies are being used in more and more investigations these days. At the pinnacle of technology, convolutional neural networks for computer vision have been built to help the visually impaired. There are several tangible tools available to support researchers across multiple disciplines in their everyday work.

Modern computer science has made great strides in image processing and data science in particular, which have opened up a lot of possibilities for enhancing assistive technology for the blind. These studies are the most pertinent ones:

Sonay DUMAN, Abdullah ELEWİ, and Zeki YETGİN's paper, "Design and Implementation of an Embedded Real-Time System for Guiding Visually Impaired Individuals" (Duman et al., 2019).intends to develop and put into use a portable system that will let blind individuals perceive the environment, identify nearby objects and people, and determine how far away such items are. Their suggested method uses a single camera installed on a Raspberry Pi board and YOLO (You Only Look Once), a CNN-based real-time object identification algorithm. In addition, the system calculates the separation between identified items and speaks this information out to those with visual impairments. The findings demonstrate that the system has a 98.8% accuracy rate in detecting and estimating a person's distance.

R. Gatti, J. L. Avinash, N. Nataraja, G. R. Poornima, S. Santosh Kumar, and K. Sunil Kumar 2020] in their study "Design and Implementation of Vision Module for Visually Impaired People" have proposed and implemented a vision model that will help visually impaired people access daily needs independently of others. The vision

module that was built is inexpensive to implement and portable. When deep learning methods were combined with artificial neural networks (ANNs), object detection and identification performance increased to 85% accuracy.

“Object Detection System with Voice Output using Python.”. The goal of a research by Sivaraman et al. (Sivaraman et al.,2024) is to help blind or visually impaired individuals live more freely. By effectively utilising the programme and its associated audio input, people with visual impairment will be able to overcome various hazards in their daily life, such as while reading a book or travelling around the city. It will therefore assist in averting possible mishaps. In addition to being portable, mobile devices have cameras that may be used to record audio and identify things in their surroundings. People who are blind or visually challenged can therefore "See Through the Ears."

“Real Time Object Detection for Visually Impaired Person”. Mahmood et al. (2012) conducted a research with Drs. Raghad Raied Mahmood and Majid Dherar Younus and Emad Atiya Khalaf. In order to assist visually impaired individuals in their daily lives, this article suggests a real-time object detection system. A Raspberry Pi and the deep learning algorithm YOLO (You Only Look Once) make up the system.

They employ the YOLOv3 real-time item Detection algorithm, which was trained on the COCO dataset, to recognise the item in front of the user. The anticipated output is then obtained by identifying the object's label and converting it into audio using Google Text to Speech (GTTS).

“Assistive Object Recognition System for Visually Impaired “. Shaikh S, Karale V, and Tawde G. (2020) conducted a study (Shaikh et al., 2020). The suggested study aims to give blind individuals a pleasant and dependable way to identify their environment. Using a USB camera, the sophisticated system takes pictures in front of the users in real time.

The machine learning and feature extraction method applied in this instance is called YOLO. The YOLO framework divides a picture into grids, selects the full image in one instance, and then predicts the bounding box coordinates and class probabilities for each box. This process is how object identification is accomplished. Singing YOLO has two main benefits: it's really fast and it understands generalised object representation. They could travel with confidence since the device employed text-to-speech technology to deliver audio explanations of their surroundings. The system in question is reliable, portable, and effective. Additionally, a virtual environment is produced, and assurance is given by the system stating the name of the recognised object.

3.SYSTEM ARCHITECTURE

The System architecture is displayed in the fig. [1] that follows. The three main parts of the system's design are a camera, a Raspberry Pi, and headphones, as shown in the diagram. The user turns on the system and turns on the associated camera. Live video streaming has started now that the camera is turned on. As soon as the live streaming is taken, the item will start to be collected. The object pattern from the recorded scene is then found utilising the YOLO API's pattern matching feature. The procedure of feature extraction is utilised by the system to ascertain the kind of item. Feature extraction uses the object's characteristics to identify it. Using the camera's focus length, the system will print the name in text. The Python Audio Library may be used to provide audio signals for the objects that have been detected. The audio output of the generated signal is transferred to the earphone that is attached to the ser's ear.

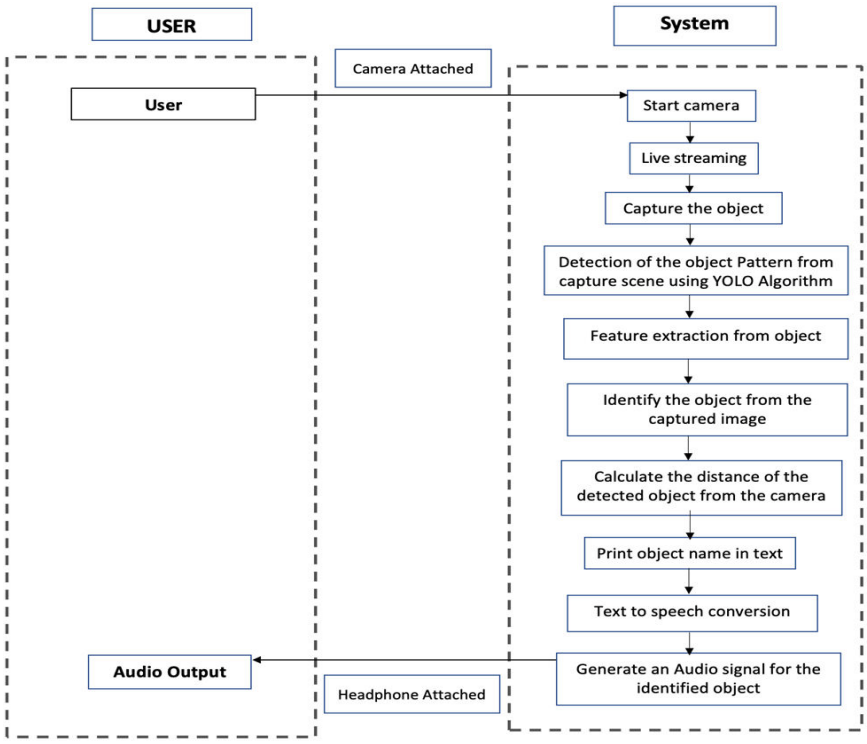

Figure 1 System Architecture of the Proposed System

This research provides a way to let blind persons safely explore new outside environments until they come upon an interesting object. The Processing Device is a high-frame-rate camera and a Raspberry Pi. Using a light and a modest processing speed, the little gadget recognises items of interest, identifies impediments in the user's route in real time, and computes a trajectory for safely achieving the target. The techniques used by the system are a combination of computer vision and machine learning.

4.METHODOLOGY

4.1. You Only Look Once (YOLO) Algorithm

YOLO is a novel technique for instantly identifying and drawing bounding boxes around many items in a picture. To get the result, it merely runs the picture through the CNN algorithm once. Because of its simplified design, YOLO performs better than Faster R-CNN even though it is almost exactly the same as R-CNN. YOLO can concurrently conduct bounding box regression and classification, unlike Faster R-CNN. Using the class label of an item, YOLO can forecast its position. By spatially separating bounding boxes and the class possibilities they represent—which are predicted using a single neural network—YOLO approaches object detection as a regression issue. Compared to the traditional CNN pipeline, this is a significant change (Shaikh et al., 2020).

The speed and precision of YOLO's real-time object identification are explained by the YOLO algorithm, which expands on GoogleNet equations to be employed as its fundamental forwarding transport computation. As opposed to R-CNN architectures, which reevaluate probability scores after running a classifier on a hypothetical bounding box (Kumar et al., 2021). Bounding box predictions and class probability for those bounding boxes are made concurrently by YOLO. The GoogLeNet model for image classification, which is akin to a conventional convolutional neural network, served as the model for YOLO's design. First, the network collects characteristics from the picture, and then its fully connected layers predict the output coordinates and probabilities. Twenty-four convolutional layers, two fully connected layers, one-by-one reduction layers, and three-by-three convolutional layers were used in the construction of the YOLO network model (Huang et al., 2022).

In this Paper, use YOLO v3 more quickly than it did with version 1. On a 28.2 map at 320 × 320 YOLOv3, it operates three times quicker at 22 ms. It is 3.8 times quicker, yet it performs the same. The v3's ability to generate three distinct acquisition scales is its most significant feature. The YOLO v3 neural network is highly adaptive and uses a 1 x 1 core regarding distinctive blueprint to accomplish its desired effect. One way to highlight YOLOv3 in figure 2 is to put 1x1 discs on three-dimensional feature maps in three different parts of the network.

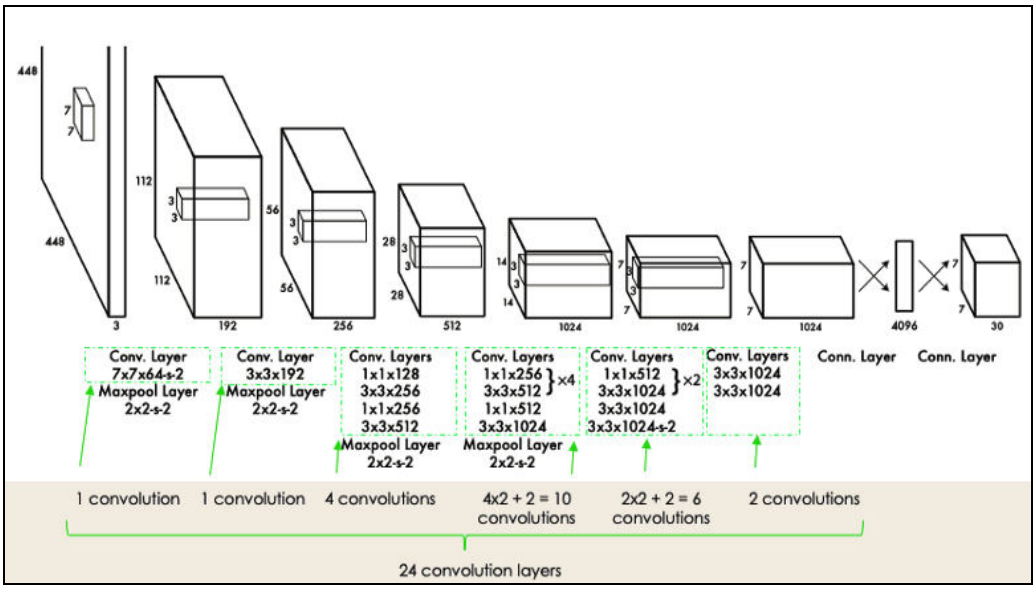

Figure 2 YOLO algorithm Architecture

O Steps

1. An input picture is divided into a grid of SxS cells, where m bounding boxes are indicated, to create a YOLO network. Using these bounding boundaries, the network predicts class probability and offset values, as shown in figure (3).

2. In these bounding boxes, networks forecast class probabilities and offset values.

3. These bounding boxes are utilised to find the item within the picture when their class probability exceeds the threshold value.

4. For this, the intersection over union (IOU) strategy is employed.

5. Compared to the other methods, Yolo finds things in the image 45 frames per second quicker.

O Reasons to Choose YOLO Algorithm:

1. It is capable of making predictions using a single network evaluation rather than hundreds of assessments for a single image. The algorithm is 1,000 times and 100 times quicker than existing object detection algorithms like R-CNN and Fast R-CNN.

2. Yolo is extremely quick, processing at 45 frames per second.

3. A forecast (object locations and classes) is created based on a single network. To increase the accuracy of the model, end-to-end training is possible.

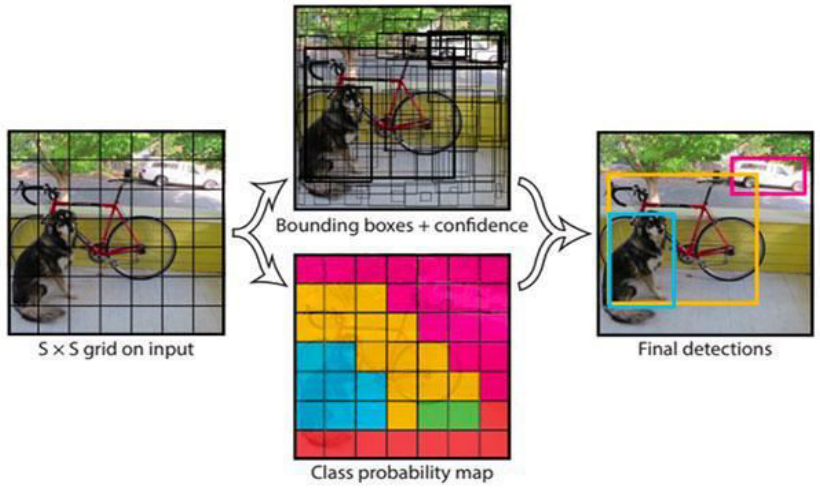

Figure 3 the YOLO Model

4.2. YOLO Working Mechanism

Yolo is an algorithm based on regression. It forecasts bounding boxes that describe an item's position throughout the picture as well as the probability of object classes. An object's bounding boxes are expressed as b_x,b_y, where the box's centre is defined by the x and y coordinates in relation to the boundaries of the grid cell. The object's class is indicated by the number c, while the image's width and height are denoted by the numbers bw and bh, respectively. YOLO is given the image, which is divided into (S×S) grids (3×3) in order to properly identify the item. Two tactics that can be employed are Non-Max Suppression and Intersection Over Union (IOU). IOU uses both predicted and real bounding box variables, and the IOU for both is calculated in Eq (1), as shown below (Qing et al.,2021). :

$$IOU = \frac{Intersection\ Area}{Union\ Area} \qquad (1)$$

The next method, called Non-Max Suppression, uses high possibility boxes and suppresses boxes with high IOU values. Until a box is identified as the object's bounding box, this process is repeated. The item in each grid cell is forecasted to have "c" conditional class probabilities. These odds are determined by the grid cell in which the object is placed. A grid cell projects just one set of class probabilities, regardless of the number of bounding boxes it contains (Zhu et al., 2018).

5. SYSTEM IMPLEMENTATION

5.1. Video Capturing

We connected our webcam to this model using cv2.VideoCapture() so that OpenCV could receive input in real time. Tfnet is necessary in order to recognise the item. All training is mostly done on tfnet by Yolo. The TfNet class uses all of the libraries before utilising a variety of techniques. The whole optimizer may be found in the _TRAINER directory. The goal of each of their optimizers is to increase network performance. The Frame Rate, or frequency of the video, is what is returned by the _get_fps function. The smoother the action of the video, the higher the FPS. We used the capture.read () method, which returns Boolean values, to show our results. The camera will frame and offer actual value when it is switched on. TFNet received this frame as an input by means of the tfnet.return_predict() method. In this way,, we will obtain the results of object identification. Then, for each object, we create different colored boxes.

5.2. Object Detection

The first step in using the YOLO algorithm is to ascertain what is being forecasted. Lastly, we wish to forecast the class of an item and the bounding box that indicates its placement.

The bounding box of an object, its class, and the likelihood that the class of object would be present within it are all predicted by the YOLO detector. The following parameters apply to each bounding box.

Each bounding box can be described using one of four descriptors, as figure (4) illustrates:

1. The bounding box's centre (bx, by)

2. Breadth (bw)

3.Height (bh)

4. An object's class (c), such as an automobile, traffic signals, etc.

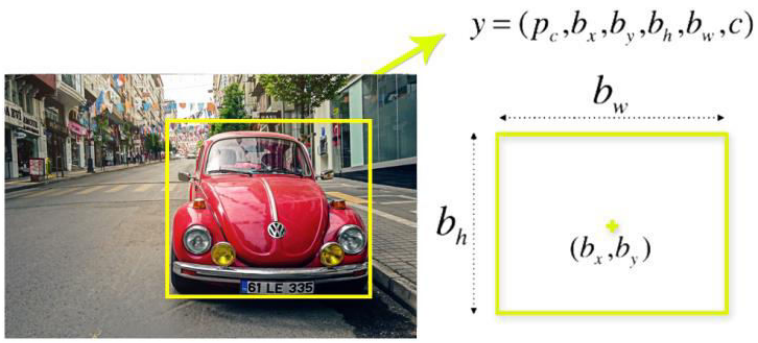

Figure 4 Object detection

A probability value (pc) for each bounding box indicates the likelihood that a certain kind of item will be found inside that bounding box.

Cells inside the picture are separated, usually into (19x19) grids. If a cell includes more than one object, each cell must predict five bounding boxes. Most of these bounding boxes and cells will be empty. Thus, we forecast the value (pc), which is employed in the non-max suppression procedure to exclude boxes with low object probability and bounding boxes with the maximum common area, as seen in figure (5).

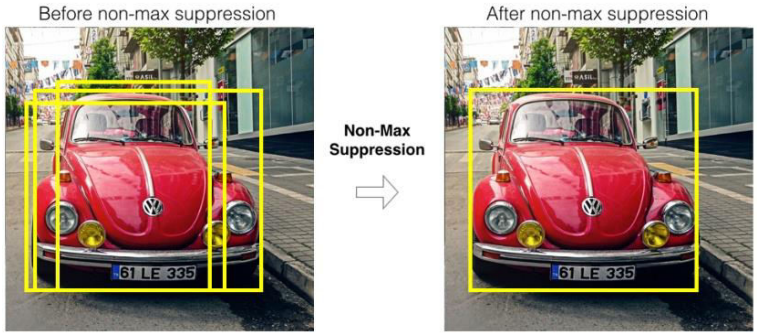

Figure 5 Non-max suppression

5.3. Audio Output

The object's location and label are converted into audio format by the audio system. The user then hears the sounds through the speaker.

Speech synthesis is the artificial creation of human speech. A computer system that may be used in hardware or software products is called a speech computer or voice synthesiser. Some systems translate symbolic language

representations, including phonetic transcriptions, into voice; text-to-speech (TTS) systems translate text into speech.

6. RESULTS AND SYSTEM IMPLEMENTATION

Provide the findings of using the suggested deep network for object detection in this part. A prototype of the suggested system has been put into practice. Both hardware and software are included in the prototype.

The hardware components are:

1. Raspberry Pi 3 Model B board
2. Raspberry Pi Camera Module
3. Power Supply (5 V)
4. Headphones

The software components are:

1. Programming language python 3.6.5 under the Idle environment is utilized to build the application program, and PyCharm CE.
2. YOLOv3 (Tiny) Real-Time Object Detection System with NNPack Darknet(Duman et al.,2019). the opencv4.
3. Google Text-to-Speech Module

In this paper Open CV module was used with the pre-trained YOLO model to do objects detection. The module was trained to track objects in the surrounding environments. This model is trained on COCO dataset from Microsoft. It is capable of detecting 80 common objects such as person ,book, mobile,bottle,remote as shown below figure 7 . After that, the testing process is carried out by inserting images and recognized by the CNN, and the results were excellent in identifying the objects, as ready images and images taken by the regular camera and Raspberry camera were inserted, and this work together was done with an algorithm called YOLO which means that you only look once at the image which is very quickly.the program has been implemented in Python using pycharm. The experiments have been performed with different lights conditions.

The result of this paper is as shown in figure 6.the system detect with an accuracy of 99% , and different object detection with an average accuracy of 85% as the result shown in figure:

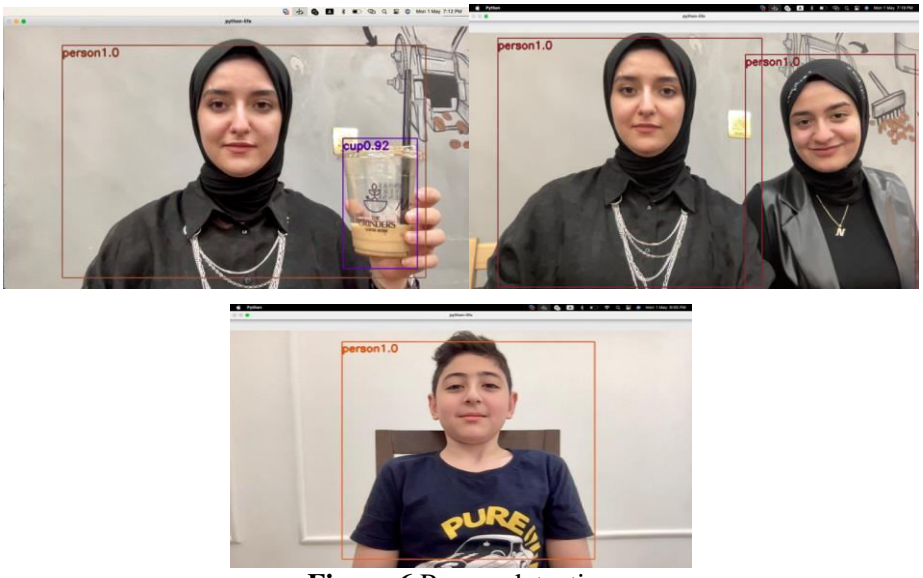

Figure 6 Person detection

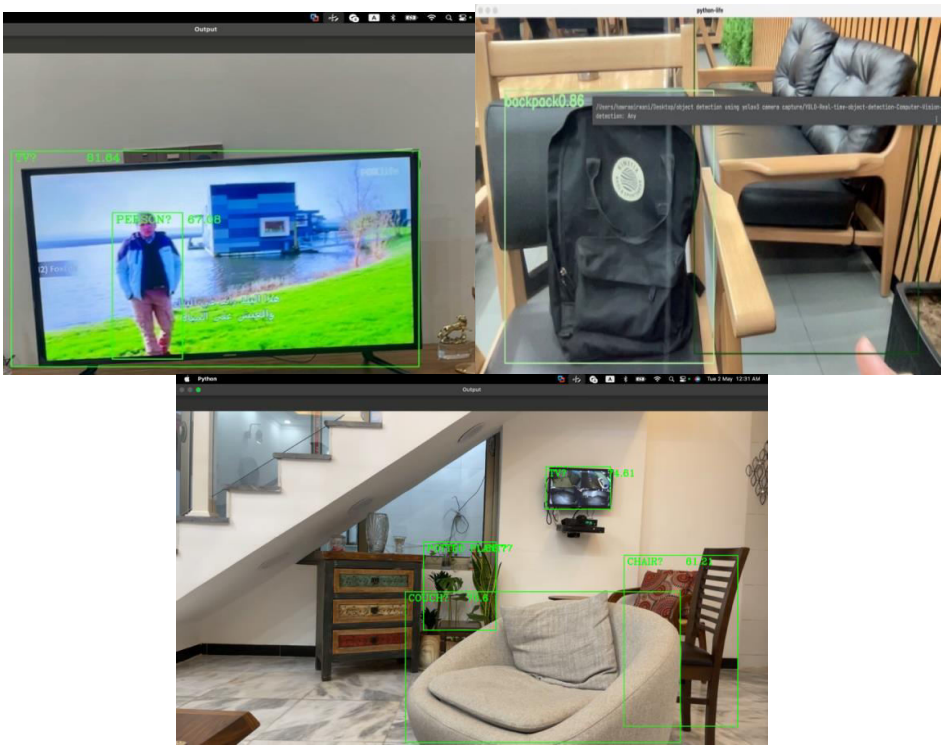

Figure 7 Object detection

7. CONCLUSIONS

This paper presents an object detection model for visually impaired people based on the YOLO algorithm.. The suggested approach for text-to-speech (voice) and real-time object recognition For visually challenged persons, machine learning feedback on a Raspberry Pi may be very promising. With this suggested approach, the feedback would be audible to them as an announcement. The freedom of movement has always been restricted for those who are sight impaired. We hope that the suggested technique would be useful in giving visually impaired people a decent existence by allowing them to commute freely and without assistance. A camera was used to take the pictures, and a dataset including a variety of image types was used to verify the model. A Raspberry Pi was used to handle the data. The experimental findings demonstrate that the suggested model attains an accuracy of 70–100%. Moreover, the facial recognition system had a 95%–100% accuracy rate. In the future, we want to use the suggested paradigm in parallel settings.

REFERENCES

[1] World Health Organization, Blindness and Vision Impairment. http://www.who.int/en/news-room/fact-sheets/detail/blindness- and-visual-impairment, Access Date: 9/6/2019

[2] Duman, S., Elewi, A., Yetgin, Z. (2019). Design and implementation of an embedded real-time system for guiding visually impaired individuals. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) (pp. 1-5). IEEE.

[3] Guravaiah, Koppala, et al. "Third Eye: Object Recognition and Speech Generation for Visually Impaired." *Procedia Computer Science* 218 (2023): 1144-1155.

[4] Sharma, N. K., Gautam, D. K., Rathore, S., & Khan, M. R. (2021). WITHDRAWN: CNN Implementation for Detect Cheating in Online Exams During COVID-19 Pandemic: A CVRU Perspective.

[5] Mounir, A. J., Mallat, S., & Zrigui, M. (2021). Analyzing satellite images by apply deep learning instance segmentation of agricultural fields. Periodicals of Engineering and Natural Sciences, 9(4), 1056-1069.

[6] Roy, B., Nandy, S., Ghosh, D., Dutta, D., Biswas, P., & Das, T. (2020). MOXA: A deep learning based unmanned approach for real-time monitoring of people wearing medical masks. Transactions of the Indian National Academy of Engineering, 5, 509-518.

[7] Peeples, L. (2020). Face masks: what the data say. Nature, 586(7828), 186-189.

[8] Yi, Z., Yongliang, S., & Jun, Z. (2019). An improved tiny-yolov3 pedestrian detection algorithm. Optik, 183, 17-23.

[9] Duman, S., Elewi, A., & Yetgin, Z. (2019, September). Design and implementation of an embedded real-time system for guiding visually impaired individuals. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) (pp. 1-5). IEEE.

[10] Gatti, R., Avinash, J. L., Nataraja, N., Poornima, G. R., Kumar, S. S., & Kumar, K. S. (2020, November). Design and Implementation of Vision Module for Visually Impaired People. In 2020 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT) (pp. 373-377). IEEE.

[11] Sivaraman, K., Gopi, P., Karthik, K., & Reddy, K. V. U. (2024). Object detection system with voice output using Artificial Intelligence. In Artificial Intelligence, Blockchain, Computing and Security Volume 1 (pp. 181-186). CRC Press.

[12] Mahmood, R. R., Younus, M. D., & Khalaf, E. A. (2021). Real Time Object Detection for Visually Impaired Person. Annals of the Romanian Society for Cell Biology, 14725-14732.

[13] Shaikh, S., Karale, V., & Tawde, G. (2020). Assistive object recognition system for visually impaired. International Journal of Engineering Research & Technology (IJERT), 9(9), 736-740.

[14] Shaikh, S., Karale, V., & Tawde, G. (2020). Assistive object recognition system for visually impaired. International Journal of Engineering Research & Technology (IJERT), 9(9), 736-740.

[15] Kumar, A., Kalia, A., Verma, K., Sharma, A., & Kaushal, M. (2021). Scaling up face masks detection with YOLO on a novel dataset. Optik, 239, 166744.

[16] Huang, R., Pedoeem, J., & Chen, C. (2018). YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers. In 2018 IEEE international conference on big data (big data) (pp. 2503-2510). IEEE.

[17] Qing, Y., Liu, W., Feng, L., & Gao, W. (2021). Improved Yolo network for free-angle remote sensing target detection. Remote Sensing, 13(11), 2171.

[18] Zhu, X., Dai, J., Yuan, L., & Wei, Y. (2018). Towards high performance video object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7210-7218).

[19] Duman, S., Elewi, A., & Yetgin, Z. (2019, September). Design and implementation of an embedded real-time system for guiding visually impaired individuals. In 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) (pp. 1-5). IEEE.