Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 14725 - 14732
Received 05 March 2021; Accepted 01 April 2021.

Real Time Object Detection for Visually Impaired Person

Raghad Raied Mahmood[1], Dr. Majid Dherar Younus[2], Dr. Emad Atiya Khalaf[3]
[1,2,3]Department of Computer And Information Engineering, College of Electronics Engineering, Nineveh University

ABSTRACT

According to statistics from the World Health Organization (WHO), at least 285 million people are visually impaired or blindness. Blind people generally have to rely on white canes, guide dogs, screen-reading software, magnifiers, and glasses for navigation and surrounding object detection. Therefore, to help blind people, the visual world has to be transformed into the audio world with the potential to inform them about objects.

In this paper, we propose a real-time object detection system to help visually impaired people in their daily life. This system consists of a Raspberry Pi in which YOLO (You Only Look Once) deep learning algorithm is employed.

We will use YOLOv3 real-time Object Detection algorithm trained on the COCO dataset to identify the object present before the person. Then the label of the object is identified and then converted into audio by using Google Text to Speech (gTTS), which will be the expected output.

Keywords
Visually impaired, Object detection, Raspberry Pi, YOLO, Text to speech.

1. Introduction

According to the study conducted worldwide by the World Health Organization (WHO), about 285 million people suffer visually impaired, of whom 39 million were blind, 246 million had low vision. The number of visually impaired people is exploding with the growth of the newborn population, eye diseases, accidents, aging, and so on, and every year, this number grows by up to 2 million worldwide [1][2].

The abilities of the visually impaired for performing daily tasks are limited or influenced. For that reason, many visually impaired people will bring a sighted friend or family member to help navigate unknown environments. These social challenges limit a blind person's ability to meet people [3].
Previous research has suggested many strategies to overcome the issues of visually impaired people (VIPs) to live normally. These strategies have not been able to fully address the safety measures when VIPs walk on their own and the proposed ideas are generally high in complexity, and not cost-effective etc.[4].

We suggest a system based on image processing and machine learning breakthroughs. The system comprises a Raspberry Pi in which the YOLO (You Only Look Once) deep learning algorithm is employed, whereby the device includes a camera module and an audio jack. The camera will capture the object's image that is in front of the person. Thereafter, it gets processed using deep learning methods and, in turn, the output, which is the name of the object, will be converted into audio for the user through the audio jack. A system is proposed aids visually impaired people in dealing with day-to-day activities like walking, working, and doing house chores.

2. Related Work

To support the visually impaired, many technologies have been developed. Some relevant works connected to this segment are described below:

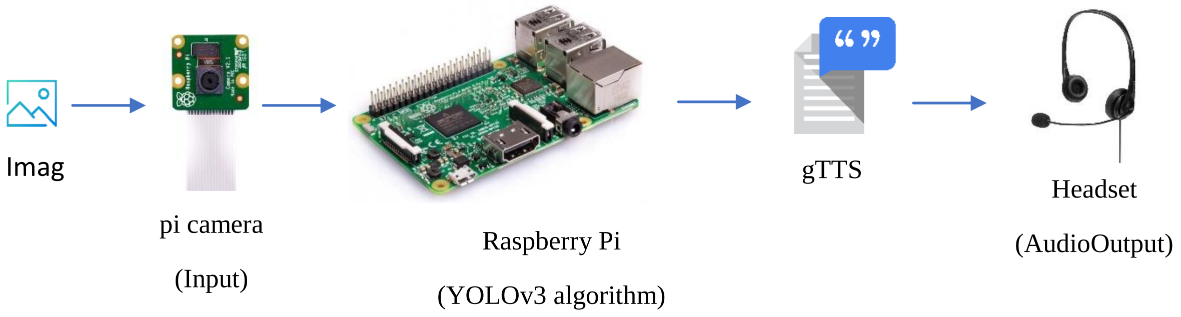
In 2019, Sanghyeon Lee and Moonsik Kang. [5], proposed an object detection system for the blind using deep learning technologies. The authors used voice recognition technology to know what objects a blind person wants, and then to find the objects via object recognition. The object recognition deep learning model utilizes the Single Shot MultiBox Detector (SSD) neural network architecture, and voice recognition is designed through speech-to-text (STT) technology. Also, a voice announcement is synthesized using text-to-speech (TTS) to make it easier for the blind to get information about objects. The control system is based on the Arduino microprocessor.

In 2019, Aswath Suresh et al. [6], investigated Smart Glass representing a potential aid for people who are visually impaired. The Smart glass consists of ultrasonic sensors to detect the object ahead in real-time and feeds the Raspberry for analysis. It had an added feature of GSM, which can assist the person to make a call during an emergency situation. It is developed using the ROS catkin workspace with necessary packages and nodes.

In 2019, Amruta Bhandari et al. [7], proposed a system provide a low-cost portable solution to help blind people. It consisted of shoes having ultrasonic sensors to survey the scene. Once the sensors detect the object in front of the person, the camera module gets activated. After detecting the object, the camera module clicks the image of an object which is in front of the people, and then it will be processed to detect the object type, according to which information will be sent as audio via a Bluetooth headset to the user. The visually impaired human will get the audio instructions accordingly about the obstacles in the dynamic environment.

3. Block Diagram

The block diagram of our system is depicted in the figure below.



Received 05 March 2021; Accepted 01 April 2021.

Figure -1 System Block Diagram

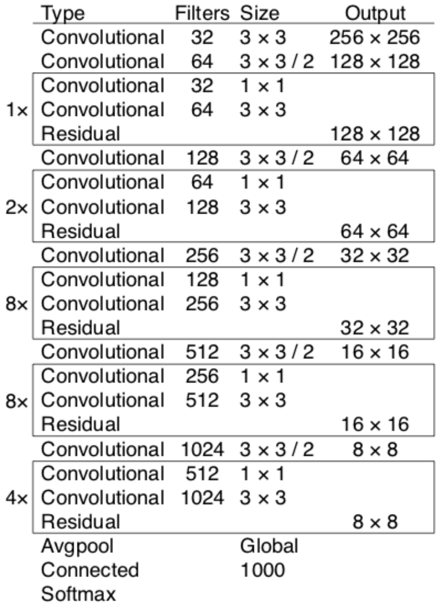
Our system's process begins with an image acquisition requirement, which is met by the Raspberry Pi camera connected to the Raspberry Pi port. A YOLOv3 algorithm is installed on the Raspberry Pi. A headset is connected to one of the Raspberry Pi's USB ports, as an output speech unit.

3.1 System Information

Raspberry Pi: is the core of our system. The Raspberry Pi is one of the most common single-board computers on the market. With OpenCV and TensorFlow on the Raspberry Pi, you can quickly implement any of the main image processing algorithms and operations. With our Raspberry Pi, we're using a 32 GB class 10 SD card and a Raspberry Pi 3. We're going to get the results in audio form, so we're using a headset and a Raspberry Pi camera for image acquisition.

YOLO:is an algorithm that uses convolutional neural networks for object detection. YOLO, is one of the faster object detection algorithms out there.

YOLO v3 deeper architecture of feature extractor called Darknet-53. which contains of 53 convolutional layers, each followed by batch normalization layer and Leaky ReLU activation[8].the figure below show the darknet-53 architecture.



	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure - 2 Darknet-53 Architecture

This Algorithm applies a single Neural network to the Full Image. It means that this network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities[9].

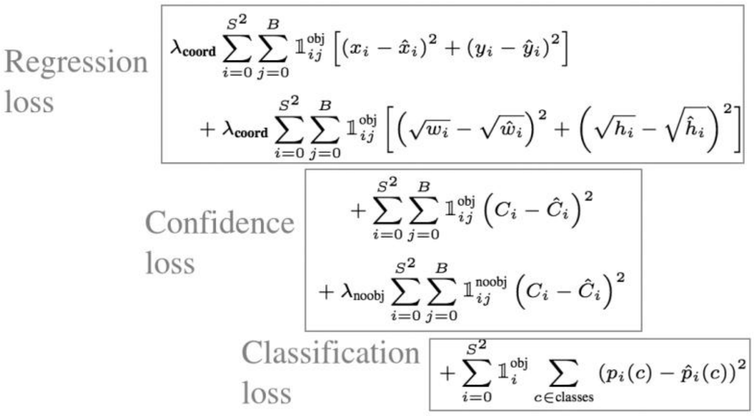
YOLO predicts multiple bounding boxes per grid cell. For this, we select only a few boxes based on:

- Score-thresholding: Boxes that have detected a class with a value less than the threshold should be discarded.

- Non-max suppression: Calculate the intersection over the union to avoid selecting boxes that are overlapping.

To measure the loss, YOLO uses the sum squared error between the predictions and the ground truth.
The loss function [8]of YOLO v3 can be summarized as follows:

- Confidence loss: determine whether there are objects in the prediction frame.
- Box regression loss: computed only when the prediction box contains objects.
- Classification loss: determine the class of the object in the prediction frame.



Regression loss

$$\lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2 \right] + \lambda_{\textbf{coord}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left[\left(\sqrt{w_i} - \sqrt{\hat{w}_i} \right)^2 + \left(\sqrt{h_i} - \sqrt{\hat{h}_i} \right)^2 \right]$$

Confidence loss

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{obj}} \left(C_i - \hat{C}_i \right)^2 + \lambda_{\text{noobj}} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \mathbb{1}_{ij}^{\text{noobj}} \left(C_i - \hat{C}_i \right)^2$$

Classification loss

$$+ \sum_{i=0}^{S^2} \mathbb{1}_{i}^{\text{obj}} \sum_{c \in \text{classes}} (p_i(c) - \hat{p}_i(c))^2$$

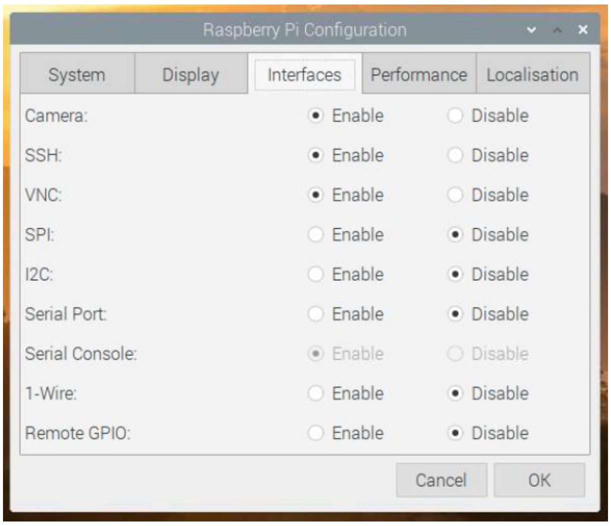
Figure -3 YOLO Loss function

OpenCV (Open Source Computer Vision Library): is an open source computer vision and machine learning software library. OpenCV was built to provide a common infrastructure for computer vision applications and to accelerate the use of machine perception in the commercial products[10]**.**

DNN module (Deep Neural Network): is the module in OpenCV responsible for all deep learning related concepts and enables the use of pre-trained models. A DNN based algorithm is more robust and accurate on a wide range of faces. In the DNN module of OpenCV, it requires your input to transform to a blob, or tensor[11].

gTTS (Google Text to Speech) : a Python library and CLI tool to interface with Google Translates text-to-speech API. Writes spoken mp3 data to a (stdout). It features flexible pre-processing and tokenizing[12].

Raspberry Pi Camera: for performing object detection on the Raspberry Pi, the Python script detects objects in live feeds from a Pi camera. Pi camera is enabled in the Raspberry Pi configuration menu:



Camera: ● Enable ○ Disable
SSH: ● Enable ○ Disable
VNC: ● Enable ○ Disable
SPI: ○ Enable ● Disable
I2C: ○ Enable ● Disable
Serial Port: ○ Enable ● Disable
Serial Console: ● Enable ○ Disable
1-Wire: ○ Enable ● Disable
Remote GPIO: ○ Enable ● Disable

Figure -4 Raspberry Pi configuration menu

4. Implementation

When the Raspberry Pi is turned on, the code implemented begins with importing required libraries which are OpenCV, gTTS, Time, and NumPy and reading COCO class names from the text file, YOLO weights, and configuration files, then the code will initialize the camera connected to Raspberry Pi.

After that, the camera will capture real-time frames, then the code will read the input frame and get its width and height. To obtain a correct predication, we will use the OpenCV function (**blobFromImage**) to get the BLOB of the input frame, then set it to the YOLO pre-trained model using dnn model.

Then the code performs a forward pass of the YOLO object detector For each detection from each output layer we will get the confidence, class ID, bounding box. Then ignore the object (confidence < 0.5), and apply non-max suppression to determine the detections remaining.

Our system is designed to provide audio output to visually impaired people. The Detected object labels are converted into speech using the gTTS library, which is a python library. Our system gives the spatial location and name of the object to the person. By using this information, the person can have a visualization of the surrounding objects. The flowchart below shows how our system works.

Received 05 March 2021; Accepted 01 April 2021.

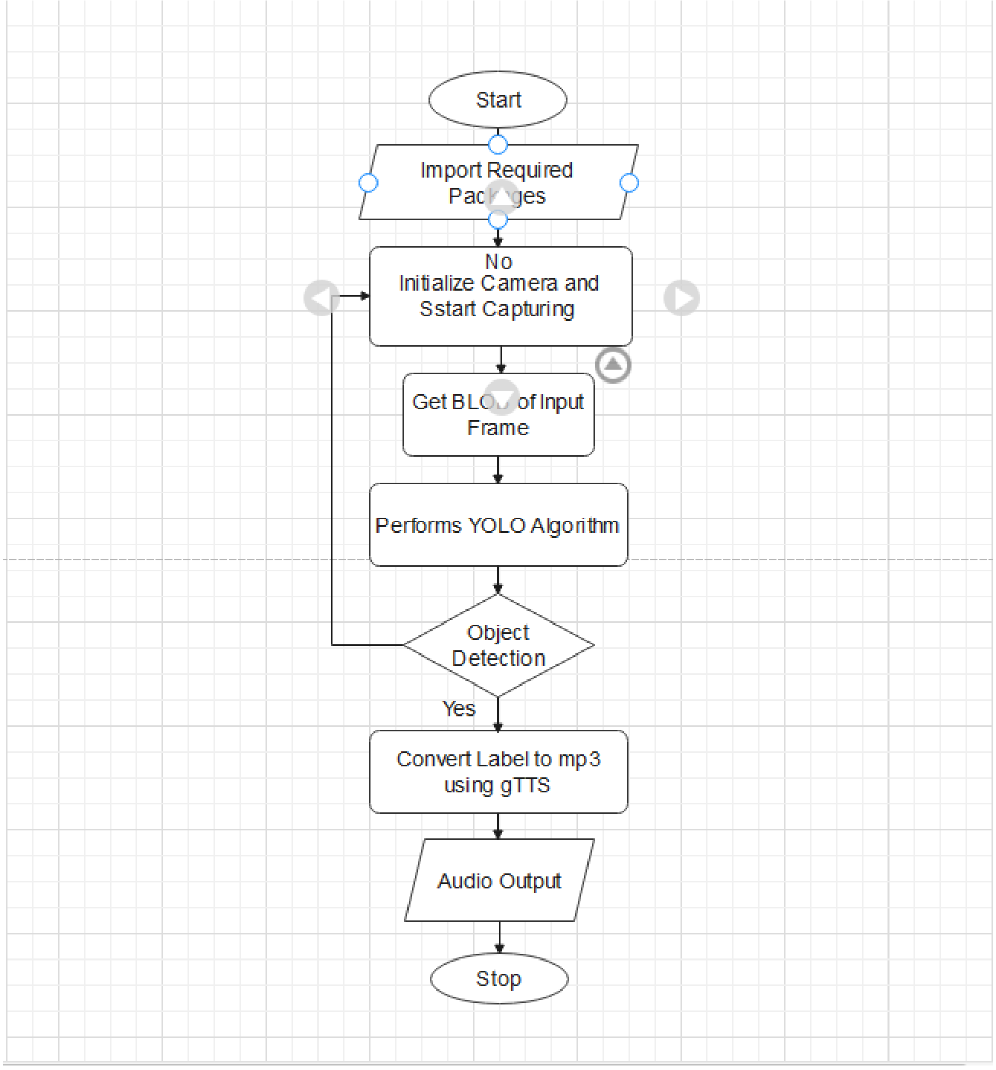


Figure -5 Flowchart of system

5. Result and Experiments

The proposed system will be able to identify the object in front of the camera and will convert it into mp3.

We performed an experiment on the proposed system, using multiple objects as seen in the figure below.

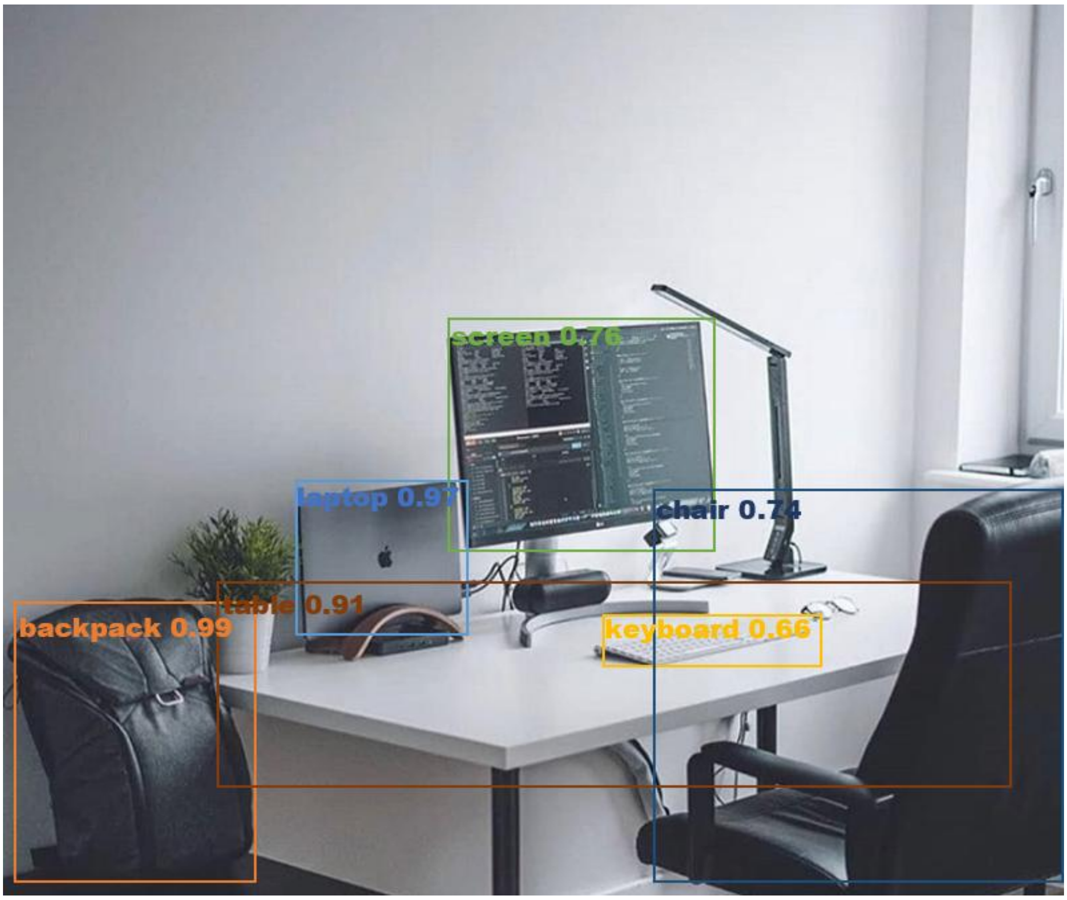


Figure -6 Real-time Object Detection

6. Conclusion

Deep Learning and Raspberry technology gave us the ability to develop these projects. Our technology will assist the visually impaired. We will be able to detect objects more reliably and independently identify objects with the fixed location of an object in the x-axis, y-axis image.

Later, utilizing gTTs, it is translated to mp3 and used for visually impaired people. In our project, we designed a low-cost system that can be extremely beneficial to individuals in need.

References

[1] “WHO | World Health Organization.” https://www.who.int/en (accessed Oct. 08, 2020).

[2] J. CHEN, “Research on Image Processing for Assisting the Visually Impaired to Access Visual Information,” no. September, 2015.

[3] R. Rajwani, D. Purswani, and P. Kalinani, “Proposed System on Object Detection for Visually Impaired People,” *Int. J. Inf. Technol.*, vol. 4, no. 1, pp. 1–6, 2018.

[4] F. Rahman, I. J. Ritun, and N. Farhin, “Assisting the visually impaired people using image processing,” no. July, 2018.

[5] S. Lee and M. Kang, “Object detection system for the blind with voice command and

guidance," *IEIE Trans. Smart Process. Comput.*, vol. 8, no. 5, pp. 373–379, 2019, doi: 10.5573/IEIESPC.2019.8.5.373.

[6] A. Suresh, C. Arora, D. Laha, D. Gaba, and S. Bhambri, "Intelligent smart glass for visually impaired using deep learning machine vision techniques and robot operating system (ROS)," *Adv. Intell. Syst. Comput.*, vol. 751, pp. 99–112, 2019, doi: 10.1007/978-3-319-78452-6_10.

[7] A. Bhandari, R. Gorad, S. Thakur, and J. Sangoi, "Charanatra : A smart assistive footwear for visually impaired," *Int. J. Adv. Res. Ideas Innov. Techn*, vol. 5, no. 2, pp. 850–852, 2019.

[8] J. Redmon and A. Farhadi, "YOLO v.3," *Tech Rep.*, pp. 1–6, 2018, [Online]. Available: https://pjreddie.com/media/files/papers/YOLOv3.pdf.

[9] A. F. Joseph Redmon∗, Santosh Divvala∗†, Ross Girshick¶, "You Only Look Once: Unified, Real-Time Object Detection Joseph," *J. Chem. Eng. Data*, vol. 27, no. 3, pp. 779–788, 2016, doi: 10.1021/je00029a022.

[10] "About - OpenCV." https://opencv.org/about/ (accessed Apr. 28, 2021).

[11] Shifa Shaikh, "Assistive Object Recognition System for Visually Impaired," *Int. J. Eng. Res.*, vol. V9, no. 09, pp. 736–740, 2020, doi: 10.17577/ijertv9is090382.

[12] "gTTS — gTTS documentation." https://gtts.readthedocs.io/en/latest/ (accessed Apr. 28, 2021).