A Deep Learning Based Assistant for the Visually Impaired

Ashwini Gaikwad[1], Dr. Vinaya V. Gohokar[2]

School of Electronics and Communication Engineering, MIT World Peace University,
Pune, India[1,2].

ABSTRACT

For visually impaired people to carry out basic tasks like recognizing objects, people in the background are very challenging. The paper presents work done in the field of object detection for visually impaired people. It mainly focuses to detect sharp, dangerous objects like a fork, knife, gas stove, stairs, and microwave using a pertained model. Google's open image dataset v6 is used for training the R- CNN.

Keywords: Gas stove, Knife, Stair's detection, Deep learning, Dataset, R-CNN.

1. INTRODUCTION

According to the world health organization (WHO) in 2012, 285 million peoples were visually impaired people in the world. Roughly 36 million individuals are blind among them and the rest 217 million individuals have different vision impairments [1].

Now a days dangerous and sharp object detection is very important in the research field. New technology is arriving every day which makes our living more comfortable. But the life of visually impaired people is still difficult. The visually impaired requires more help in their daily life. To make their life more comfortable like us we can use new technologies and develop models for their assistance. Some new models are already developed to make them independent like to solve a traveling problem of visually impaired people staircase detection is done using pertain model and sensors [1]. Smart Cap is developed for visually impaired people to interact with people with some commands. This smart cap includes features like face, text, and image captioning [12].

In our day to day lives we come across many dangerous and sharp objects even at our home. Such dangerous and sharp objects are mainly found in kitchen places. In this paper, some of the dangerous objects such as gas stove, sharp knife and stair case are used to for designing our object detection model that will help visually impaired people. Dangerous and sharp object detection is done using a pertained model. For the pertained model of a specific class of objects like a gas stove, knife and staircase are used. We have taken these three classes of object images from Google's open image dataset v6. These images have different categories of a gas stoves and stair case. We are detecting the ON and OFF condition of a gas stove, up and down stair and knife with high confidence score. These images are trained using regions with convolution neural networks (R-CNN) which is a good object detection model.

2. LITERATURE REVIEW

There are different proposed methods provided in application areas like object detection, face recognition, text recognition, and human activity recognition in recent years [11]. Literature review is described in Table 1.

Table 1: Literature review.

References	Description	Accuracy
9	In this paper, they developed a model for real-time surrounding identification using deep learning techniques. Transfer learning on pretrained network to detect signs. SSD (Single Shot Multi-Box Detector) object detector and MobilenetV2 architecture are used as a base network for their model. They used Restrooms, Pharmacies, and Metro Station signs for object detection.	90.99%

10	In this paper they developed object detector model to detect small objects in images, webcam and from video. To increase performance, they used faster R-CNN and single shot multi-box algorithm.	75%
11	Their main aim to get probability and an easy user interface. This model itself divides into major aspects like object detection, face recognition, and sign language recognition. For object detection they used COCO dataset for training and SSD (Single Shot Multi-Box Detector) is used to train the model.	92.1%
1	In this paper, they develop a hybrid system for the detection of staircase and ground using pertain model and sensors. They have collected 250 images for their research and 300 images of different buildings in the real world. They used faster R-CNN to train the model.	98.73%
12	The paper SMARTCAP is a deep learning and IOT based assistance for the visually impaired, they used real time multimodal system that uses audio commands like “who is in front of me”, “describe my surroundings” and this audio command is converted into text using “Google speech to text” library.	
14	In this paper, they developed indoor signage and door recognition system. For indoor signage, they used four types of signage: exit, wc, disabled exit, and confidence zone. The deep learning algorithm is used for this system and to develop classification system transfer learning technique used.	99.8%

3. PROPOSED WORK

The block diagram of the proposed system is shown in Figure 1.

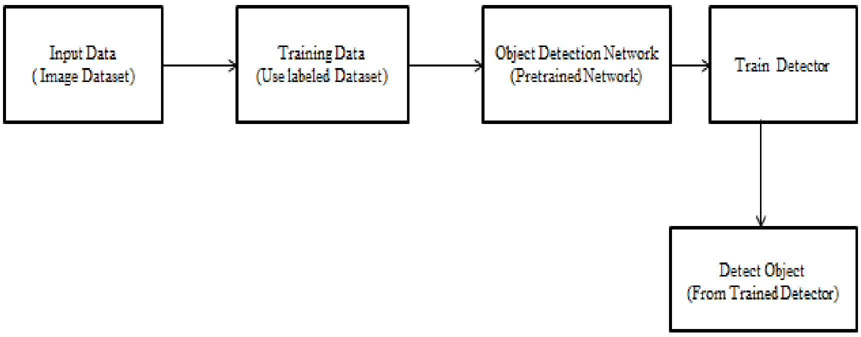

Figure 1 Block Diagram of proposed system.

The image dataset which contains images is given as an input to the system. Images are labeled using the image labeler application in MATLAB. R-CNN pretrained network is used for training. Inputs which are required for training R-CNN object detector are.

- Training data: Labeled Image dataset in input to detector. Dataset is in table format which contains grayscale or true color images. The table contains two or more columns. In this table, the first column must be an image filename and other columns are single object class.
- Network: Which network is used to train our detector is specified here. Some valid networks are listed alexnet’,'vgg16','vgg19','resnet18','resnet50','resnet101','inceptionv3','googlenet','inceptionresnetv2','squeezenet',' mobilenetv2’. Alexnet network is used in our system.
- Option: Training parameters of a neural network are defined in option.

After training the detector testing is done. For testing we have given one image to the network to verify whether the system detects the correct object in that image. And at the output stage we get the final object detected.

3.1 Dataset

Dataset includes a number of images of a number of classes. In this project common object in context (COCO2017) dataset is used. COCO2017 has a maximum number of dangerous objects are found. Some classes of objects are downloaded from Google open image Dataset V6 which comes with labels and annotation for each image. The dangerous and sharp object list is shown in Table 2.

Table 2: Dangerous object list.

Dangerous Object List	
Fork	Bicycles
Knife	Car
Microwave Oven	Bus
Edges and corner of tables	Train
Edges and corner of chairs	Traffic
Door	Wild animals
Scissors	Trees
Broken glass	Stair cases
Fungal food	Drainage
Footpaths	Gas Cylinder

3.2 Custom Dataset

We are detecting specific classes of object like gas stove, knife and stairs. Dataset for this object is not available in COCO 2017. The gas stove, knife and stair dataset are downloaded from Google's open image v6 dataset. Images are resized in 227x227 sizes for less processing time.

3.3 Dataset: Splits into Training and Testing

For gas stove class of object 118 images are used for training and 29 images used for testing. For knife class of object 25 images are used for training and 10 images used for testing. For staircase class of object 50 images are used for training and 30 images used for testing.

3.4 Labeling of Images

Labeling of images is done using the Image labeler app in MATLAB. Training Images contains gas stove in ON condition is labeled as 'ON' and for OFF condition is labeled 'OFF'. For knife class of object is labeled as 'knife' and for stair case class of object upstairs and downstairs are labeled as 'Up stair' and 'Downstair'. The illustration of labeling is shown in Figure 2. Figure 3 shows the sample of labeling of gas stove object.

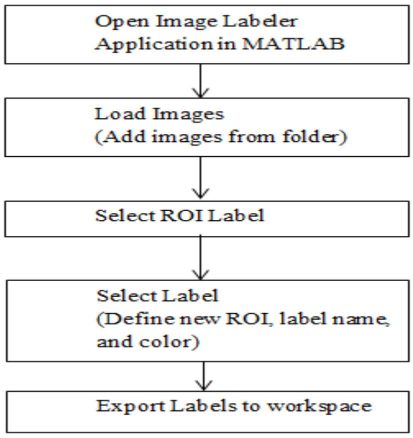

Figure 2 Illustration of Image Labeling.

Following images from Figure 3 Show samples of labeled images in MATLAB.

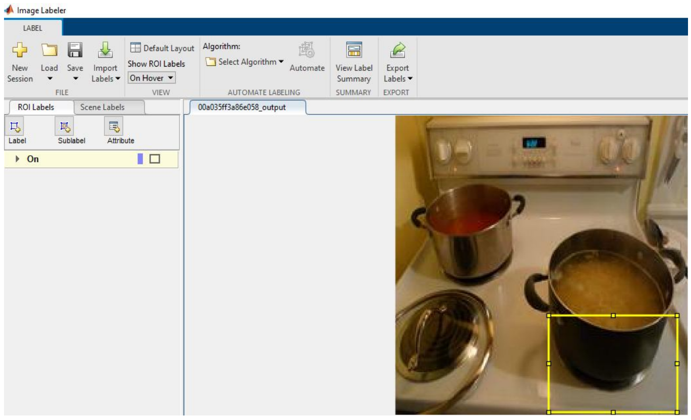

Figure 3(a). ON

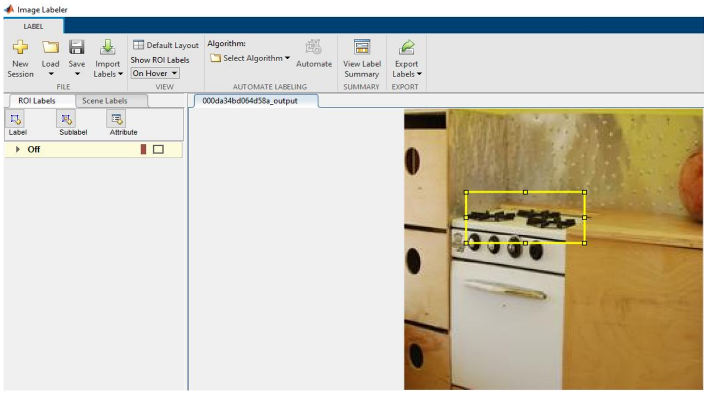

Figure 3(b). OFF

3.5 R-CNN Network

R-CNN is part of the machine learning model and it is mostly used for object detection and computer vision application. In object detection for R-CNN input is images. When an input is given, it starts the extracting region of interest called ROI. The boundary of the object in the image represents ROI that is in rectangle form. If the object is detected in an image region, then these regions are fed through CNN to extract features. Using these extracted features object is classified.

Using edge box algorithm region proposals are generated by RCNN detector. From the images these region proposals are resized and cropped. Support vector machine is used for refinement of region proposals bounding boxes. The function 'trainRCNNObjectDetector' is used to train detector. Detected object in an image is output of the detector. Figure 4 shows R-CNN detector.

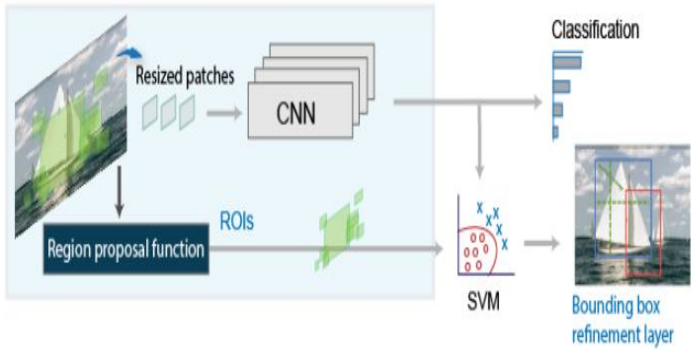

Figure 4 R-CNN Detector [16].

For our system Alexnet network is used. Using Alexnet network we have done comparative analysis on the training parameters. These analyses are explained in Table 3 below. Depending upon analysis done, we have used Sr. No 1 parameter in our system as it has less time duration is less and better accuracy compared to others.

Table 3: R-CNN Network parameters.

Sr. No	**Parameter**	**Time Elapsed**	**MiniBatch Accuracy**	**MinibBatch Loss**
1	Epoch-10	00.31.41	96.88%	0.2390
	Mini Batch size-32			
	Learn Rate-le-4			
2	Epoch-10	00.33.20	95.31%	0.1139
	Mini Batch size-64			
	Learn Rate-le-6			
3	Epoch-10	00.33.00	94.53%	0.1738
	Mini Batch size-128			
	Learn Rate-0.01			

Training progress of knife class of object is shown in below Figure 5. Figure 5 shows the training loss for each iteration to number of iterations.

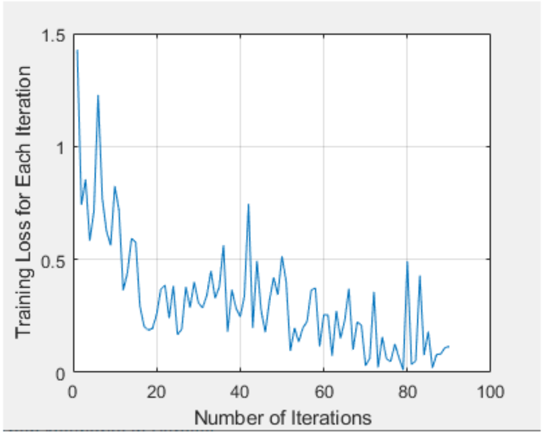

Figure 5 Training progress knife class of object.

3.6 Faster R-CNN Network

Instead of using an algorithm to add region proposal to network, Faster R-CNN adds region proposal directly to the network. Region proposals are generated faster in faster R-CNN. Faster R-CNN is more complex than R-CNN. Hence, we used R-CNN in our system. Faster R-CNN detector is shown in figure 6.

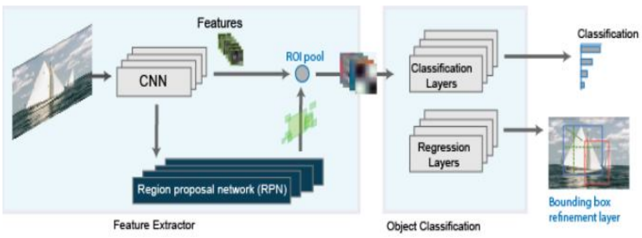

Figure 6 Faster R-CNN Detector [16].

Comparison between R-CNN and Faster R-CNN is explained in below Table 4.

Table 4: Comparison between R-CNN and Faster R-CNN

Network Type	Features
R-CNN	"Slow training and detection Allows custom region proposal" [16].
Faster R-CNN	"Optimal run time performance Does not support custom region proposal" [16]. Used for real time application

3.7 Alexnet Network

Alexnet network is already trained for large number of images. It can classify images in to 1000 object classes. Work flow of alexnet network is explained in figure 7.

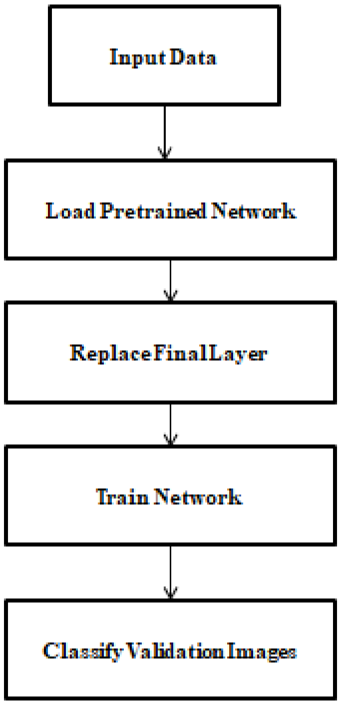

Figure 7 Work Flow of Alexnet Network

- Input Data: Load the image data. Image datastore automatically label the images and store in image datastore object. Split the dataset in to training and validating. For our system 60% images are used for training and 40% for validating. Store these split images in to new tow datastores. Figure 8 shows some sample of images from dataset.

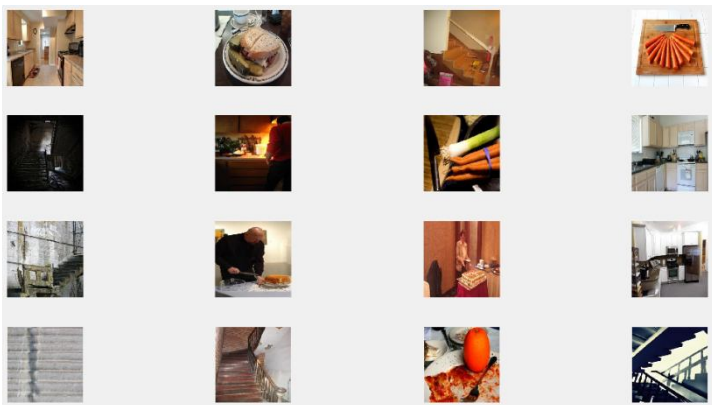

Figure 8 Classified Images from Dataset

- Load Pretrained Network: Load alexnet Pretrained network. We have to install deep learning toolbox model for alexnet network. First layer of this network is Data. Image's size of 227x227x3 is required for input data. Final layers are replaced to do fine tuning for new classification images. These layers are replaced with fully connected layer, softmax layer and classification output layer. In fully connected layer filter size is equal to number of object classes. In this model we use three classes of objects, so filter size kept three for the network.

- Train Network: Network is trained by giving training options. Training options used in this system is shown in table5. Augmented image datastore is used which resizes images during training. Training progress of the network is shown in figure 9.

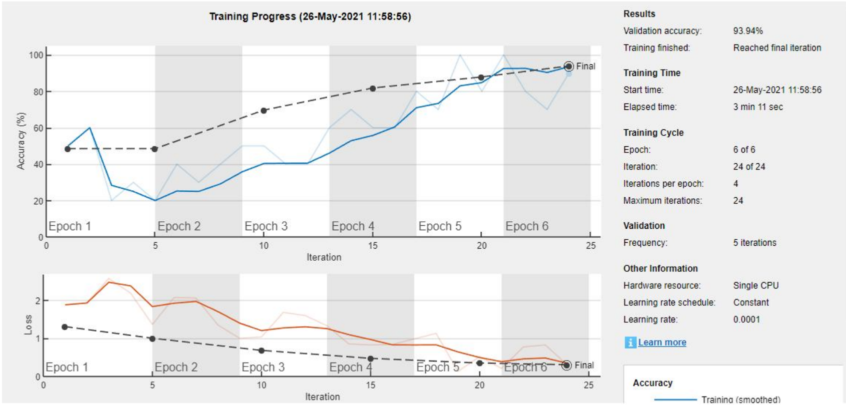

Figure 9 Training Progress the Network

Table 5: Training Options Used In the Network

Options	Values
MiniBatchSize	10
MaxEpochs	6
InitailLearnRate	1e-4
Validation Frequency	5
ValidationData	augimdsValidation

- Classify the validation images: “Fine-tuned networks are used to classify the validation images” [17]. Classified images are shown in figure 10.

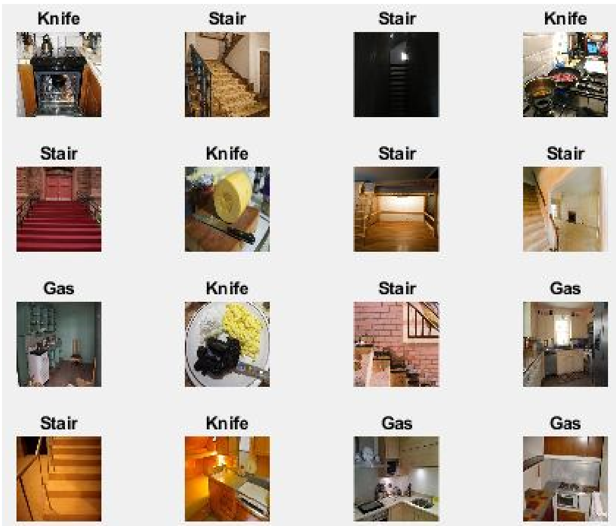

Figure 10 Classified Validation Images

4. RESULTS

Object detected with high confidence score of class knife, gas stove and stair are shown in Table 6. For Gas stove ‘ON’ and ‘Off’ condition of object is detected. Stair case ‘Upstair’ and ‘Downstair’ condition is detected. Using Alexnet

International Journal of Application or Innovation in Engineering & Management (IJAIEM)
Web Site: www.ijaiem.org Email: editor@ijaiem.org
Volume 10, Issue 5, May 2021 **ISSN 2319 - 4847**

pretrained network classification of these three classes of object is done. We have achieved 91.8% of accuracy rate for the classification of the images. Confusion matrix is shown in Figure 11.

Table 6: Results A) test images B) detected images

A. Test Image	**B. Detected Image**
	
	
	
	
	

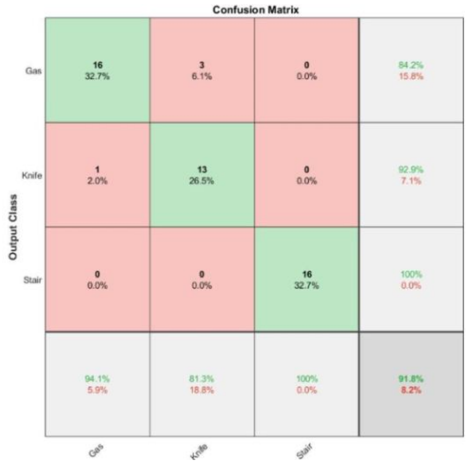

Figure 11 Confusion matrix for gas stove, knife and stair class of object

5. CONCLUSION

In this paper dangerous object detection is done using pertained R-CNN model. The gas stove is detected for both ON and OFF conditions. The stair case is detected for both Upstair and Downstair condition. Sharp object knife is also detected with high confidence score. We can develop an alert system for the same. For optimal run time performance, we can use a faster R-CNN network.

References

[1] Md. Ahsan Habib, Md. Milon Islam, Md. Milon Islam,Mahmudul Hasan,"Staircase Detection to Guide Visually Impaired People: A Hybrid Approach ", *International Journal of Computer Science and Information Security (IJCSIS)*,Vol. 16, No. 12, December 2018.

[2] S Prabakaran, Samanvya Tripathi, and Utkarsh Nagpal, " Navigational Aid for the Blind Using Deep Learning on Edge Device", *International Journal of Advanced Science and Technology* Vol. 29, No. 3, (2020), pp. 11421 – 11433.

[3] Junlong Zhou, Jianming Yan, TongquanWei, Kaijie Wu, Xiaodao Chen, and Shiyan Hu, "Sharp Corner/Edge Recognition in Domestic Environments Using RGB-D Camera Systems", *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—II*: VOL. 62, NO. 10, OCTOBER 2015.

[4] Yulong Wang , Hang Su, Bo Zhang, and Xiaolin Hu , *Senior Member*, *IEEE*, "Learning Reliable Visual Saliency For Model Explanations", *IEEE TRANSACTIONS ON MULTIMEDIA*, VOL. 22, NO. 7, JULY 2020.

[5] Salma kammoun jarraya 1, wafa saad al-shehri 2, and manar salamah ali 1, "Deep Multi-Layer Perceptron-Based Obstacle Classification Method From Partial Visual Information: Application to the Assistance of Visually Impaired People", *IEE Access*, VOLUME 8, 2020.

[6] Chanhum park , se woon cho , na rae baek , jiho choi ,and kang ryoung park , (Member, IEEE), "Deep Feature-Based Three-Stage Detection of Banknotes and Coins for Assisting Visually Impaired, People", *IEE Access*, VOLUME 8, 2020.

[7] Wan-jung chang , (member, ieee), liang-bi chen , (Senior Member, IEEE),chia-hao hsu , jheng-hao chen , tzu-chin yang , and cheng-pei lin, "MedGlasses: A Wearable Smart-Glasses-Based Drug Pill Recognition System Using Deep Learning for Visually Impaired Chronic Patients",*IEE Access*, VOLUME 8, 2020.

[8] Abdullah asim yilmaz , mehmet serdar guzel, erkan bostanci ,and iman askerzade, "A Novel Action Recognition Framework Based on "Deep-Learning and Genetic Algorithms", *IEE Access*, VOLUME 8, 2020.

[9] Hardik Gupta1, Dhruv Dahiya1, Malay Kishore Dutta1, Carlos M. Travieso2 and Jose Luis Vásquez-Nuñez3, "Real Time Surrounding Identification for Visually Impaired using Deep Learning Technique", *IEEE International Work Conference on Bioinspired Intelligence*, July 3-5, 2019.

[10] Ashwani Kumar, S S Sai Satyanarayana Reddy, Vivek Kulkarni, "An Object Detection Technique For Blind People in Real-Time Using Deep Neural Network", *2019 Fifth International Conference on Image Information Processing (ICIIP).*

[11] Jinesh, A Shah, Aashreen Raorane, Akash Ramani Hitanshu Rami, Narendra Shekokar,"EYERIS: A Virtual Eye to Aid the Visually Impaired", *3rd International Conference on Communication System, Computing and IT Applications (CSCITA)*, 2020.

[12] Amey Hengle, Atharva Kulkarni, Nachiket Bavadekar, Niraj Kulkarni, Rutuja Udyawar," Smart Cap: A Deep Learning and IoT Based Assistant for the Visually Impaired", *Proceedings of the Third International Conference on Smart Systems and Inventive Technology (ICSSIT 2020).*

[13] Ashwani Kumar, S S Sai Satyanarayana Reddy, Vivek Kulkarni, "An Object Detection Technique For Blind People in Real-Time Using Deep Neural Network", 2019 *Fifth International Conference on Image Information Processing (ICIIP).*

[14] Mouna Afif*1, Riadh ayachi1, Yahia Said1,2,, Edwige Pissaloux3, Mohamed Atri1,4,,"Recognizing signs and doors for Indoor Wayfinding for Blind and Visually Impaired Persons", *5th International Conference on Advanced Technologies For Signal and Image Processing, ATSIP*' 2020,September 02-05, 2020, Sfax, Tunisia

[15] Saleh Shadi, Saleh Hadi, Mohammad Amin Nazari, Wolfram Hardt, "Outdoor Navigation for Visually Impaired based on Deep Learning**"**.

[16] https://in.mathworks.com/help/vision/ug/getting-started-with-r-cnn-fast-r-cnn-and-faster-r-cnn.html#mw_5ad75928-8822-4277-a1f6-6a762a5bda32.

[17] https://in.mathworks.com/help/deeplearning/ref/alexnet.html.