Object Recognition and Classification System for Visually Impaired

Rashika Joshi, Meenakshi Tripathi., Amit Kumar and Manoj Singh Gaur

***Abstract*—There have been diverse accounts of research relating to navigation assistance for the visually impaired. Using the object recognition approach , we aim to assist the blind to travel independently with the ability to identify any obstacles the the path.The subject now is to identify a solution that is low powered, easily portable and still effective. The system consists of Jetson Nano ported with the trained deep learning model and is interfaced with camera, that acts as a easy-to-use platform for object recognition, speech processing and image classification. Tests indicate that system is accurate and functions as navigation assistant.**

***Index Terms*—Google colab, Jetson Nano, MobileNet, SSD, voice Assistant.**

I. Introduction

THIS system intends to concentrate on the world's blind people and assist them through technology at every step. Mobility of visually challenged people is confined by their incapability to perceive their surroundings.

As seen in Fig. 1 data from the National Blind Federation and World Health Organization (WHO), there are around 253 million visually impaired people worldwide, of whom 36 million are blind. [1].

India now has the largest number of the world's blind, according to the Times of India report. Over 15 million of the 36 million visually impaired across the globe are from India [2] . On the other hand, while India needs an annual donation of 2.5 lakh eye, the country's 109 eye banks (five in Delhi) [3] manage to raise a maximum of only 25,000eyes, 30 % of which cannot be used. India has become the country with large number of blind people [4].

Inability to see makes everyday life a major challenge for the blind. When they visit some new or crowded place they always have to rely on their family members or walking stick. Pedestrians are frequently forced off sidewalks cluttered with animals, vendors and other obstacles so they find it even more difficult to walk in busy roads. Of example, blind people typically use canes, but the conventional cane cannot identify objects that are higher than waist. On the top of that, the death toll statistics of visually disabled people are especially alarming as they navigate the roads across the globe. Therefore, in this age of technological advancement, a solution for their autonomous, uninterrupted movement is very much required [5]

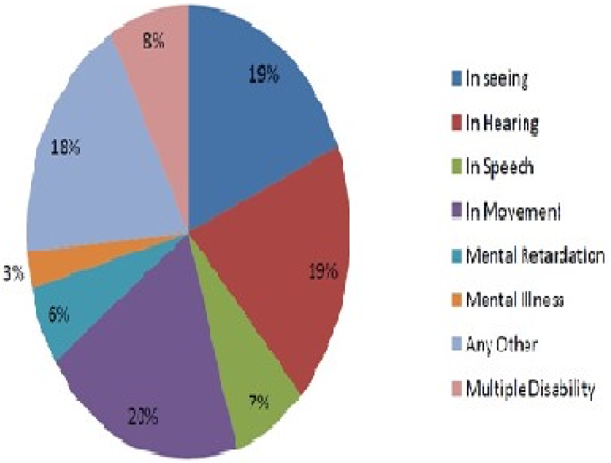


Fig. 1. Percentage of people with various disabilities as estimated in census 2011[1]

Using white cane or Guided dog is often common among visually challenged, but is not a permanent help to them. When using a guided dog the individual may go to the places where the dog is trained to go. Having to care for a dog is an additional burden for the visually impaired [6].

The main criterion we focus on is to strike a balance between the skill of the ETA's and cost-effectiveness and thus provide assistance to increasing number of visually impaired people. The main concept simply focuses on the using camera located on device to capture the object and identify the same and in return give feedback to user in form of audio and haptic feedback. The user can be made familiar to the surrounding by knowing the class of object in path. [7]

Deep learning approach for object recognition such as image classification and segmentation combined with machine learning concept for recognizing the objects in the image.

Rashika Joshi, JRF, Department of computer science, MNIT Jaipur (email: er.rashika@gmail.com).

Meenakshi Tripathi, Associate Professor, Department of computer science, MNIT Jaipur (email: er.rashika@gmail.com, mtripathi.cse@mnit.ac.in).

Amit Kumar, Assistant Professor, Department of computer science, IIIT Kota (email: amit@iiitk.ac.in).

Manoj Singh Gaur, Director, IIT Jammu (email: gaurms@gmail.com).

978-1-7281-4988-2/20/$31.00 ©2020 IEEE

Instead of designing new algorithms, object recognition and classification system focuses on using existing deep learning algorithms and tailoring them to suit application requirements. We identified the algorithms for implementing the requirements. We also performed tests on real-time environment for the testing of the system. We had overcome many challenges associated with porting the algorithms on an embedded platform and having the live system to work.

The rest of the paper is arranged as follows: Section II describes the literature survey of existing work; Section III features the system description while results are covered in section IV. Finally section V wraps up the paper with conclusion and future scope.

II. Literature Review

Blindness common worldwide can occur as a result of illness, injury, birth, and other difficulties. Many efforts have been made so far to make life better for visually impaired people. Many special equipments for visually impaired people have been created, such as speaking calculators, speaking computer terminals, speaking calipers, etc. Berry carried out a research to make web access easier [8]. In [9] authors suggested an Audio-based GPS software design for autonomous city navigation. Siemens conducted a research in which they sought to make walking with their mobile phones easier for visually disabled people [10].

Mobility Assistant for Visually Impaired (MAVI) [11] is a tool for assisting visually impaired users in independent mobility. The input sensors include an optical RGB camera,IMU sensor and GPS. The framework is equipped with four image processing modules: Signboard Detection (SBD),Texture Detection (TD),Animal Detection (AD) and Face Detection (FD). The VGA image captured by RGB camera is transmitted to these blocks and further actions are triggered based on their outputs. For example, if SBD detects a signboard, then an Optical Character Recognition system is called upon to read the contents of the signboard. The modules are controlled by a central controller enabling each of them as required. Included is a cloud-based repository for storing landmarks and other relevant information.

For object recognition, most of the techniques that are CNN based for instance R-CNN [12], begin with suggesting various scales and locations in a test image as a input to the object classifiers, for purpose of training and return the proposed region classifiers for object detection.

After classification, post-processing is performed for re scoring the boxes along with rectification of the bounding boxes based on other objects in that frame.

Followed by which, we have some of the advanced version of RCNN, like Fast- RCNN [13] and Faster-RCNN [14], which use a lot of policies to minimize regional proposal manipulation and hit a detection speed of about 5 FPS on a K40 GPU system. Faster RCNN gives satisfactory outcomes when detecting of objects over existing dataset, but fails to perform well in terms of detection speed for real-life data. This gives a clear indication that a lot more work needs to be done in enhancing the inference speed for real life data.

However the problem relating to the refinement of inference speed for real life data was overcome by YOLO[15] algorithm, using the approach of combining the area proposal with that of classification to create a unique regression problem directly from the image pixel to the bounding box coordinates along with class probabilities and access entire image in a one run. The entire detection pipeline being a unique network, it can improve the efficiency of direct end-to-end detection.

YOLO was the only system that could provide 45 FPS (on GPU) and 63.4% mAP value on VOC2007 (real-time data). However, it still faces problems in identifying smaller objects in the frame. SSD [16] rectified the problem using policy of combining faster-RCNN anchor box proposal framework and uses multi-scale features for detection layer efficiency. By maintaining the same detection speed as that of YOLO, the mAP value on VOC2007 was increased to 73.9 %.

Mobile Net [17] is based on a model of depth-based convolution that implies single input for each filter. Mobile Net's architecture model can either be thinner, or shallower. To render Mobile Net lightweight, there must be 5 layers of separable filters with a feature size of 14 *14 * 512. The Mobile Net model should therefore be thinner, offering 3% better performance than the shallow model.

III. System Description

This object recognition and classification (OCR) is a smart system for guiding the visually challenged to be aware of the surroundings. It is intended to be used along with normal walking cane. Fig. 2 shows block diagram of the system:

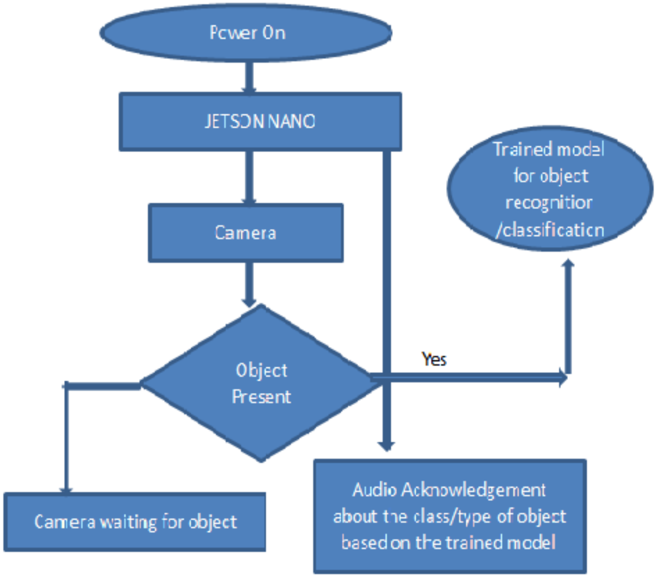

Power On
JETSON NANO
Trained model for object recognition /classification
Camera
Object Present
Yes
Camera waiting for object
Audio Acknowledgement about the class/type of object based on the trained model

Fig. 2. Object Recognition and Classification Model for Visually Impaired

A. Hardware description

Jetson Nano: This is a low-powered, small computer development kit from Nvidia that is easy to use. It is an AI computer that allows Artificial Intelligence's ability to be brought to resource-constrained devices and capable of running in parallel numerous neural networks [18]. Being a GPU enabled board, we thought this is the ideal choice to run deep learning models with high processing speed and it is also possible to train models on this, as this is best suited for low powered AI systems which is reported in Table I.

TABLE I
COMPARISON OF HARDWARE PLATFORMS

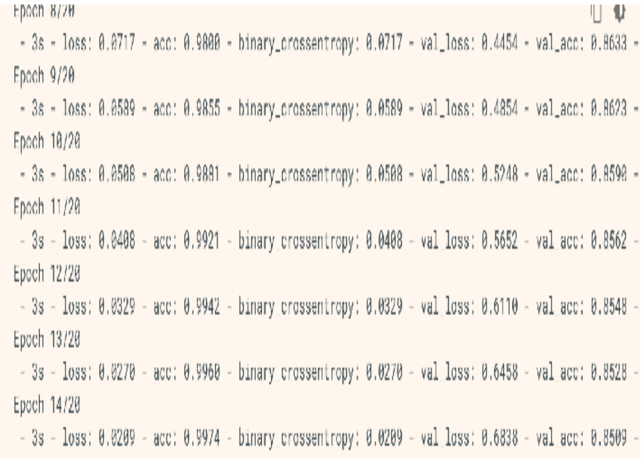
	Jetson Nano Dev Board|	RaspberryPi 3A+	Raspberry Pi 3B+,
AI Performance	472 GFLOPS	21.5 GFLOPs(est.)	21.4 GFLOPs(est.)
CPU	1.4 GHz 64-bit Quad Core ARM cortex-A57 MPCore	1.4 GHz 64-bit Quad Core ARM cortex-A53	1.4 GHz 64-bit Quad Core ARM cortex- A53
GPU	128 Core Nvidia Maxwell	Broadcomm video core IV	Broadcomm video core IV
RAM	4GB LPDDR4	512MB LPDDR2 SDRAM	1GB LPDDR2 SDRAM
Multimedia	2160p30 (H.264)	1080p30 (H.264)	1080p30 (H.264)
Video Output	HDMI, Display Port(4K)	HDMI, Display Serial Interface(DSI)	HDMI, Display Serial Interface(DSI)
Ports	4 USB 3.0, Wired Ethernet 10\100\1000 Mbps	1 USB 2.0	4 USB 2.0, Wired Ethernet up to 330 Mbps
M.2 KEY E slot	Yes	No	No

Camera: The camera module we took into use is Raspberry pi camera module v2. This module has a great still resolution of 8 mega pixels and a high video resolution up to 1080p as well. This is easy to integrate with any hardware module although best suited with raspberry pi. Using the pi camera module we get a pure python interface for this camera.

B. Software description

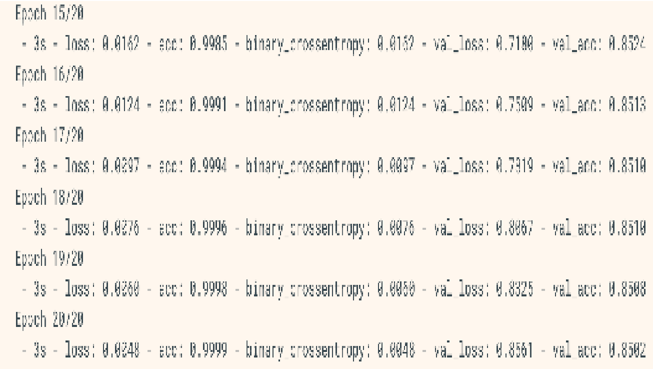
Machine learning is highly dependent on data, in fact dataset is the most crucial aspect to make any kind of training possible. We have collected dataset from various sources, like:
– Web browser, initially we searched images of a class using Google search. Further, we downloaded those images by installing bulk image downloader software and tapping on the icon and opening it in a separate tab. This helps to download multiple images of the same class and later on filter them as per our criteria.
– And for some classes for which we had issues collecting images online, we captured images manually via camera. Also, we kept image quality low to reduce the image processing time and speed up the training and dataset uploading on the cloud.

We kept 1000 images per class for training purposes with a total of 12 classes including- Background, bicycle, motorbike, bus, car, chair, dog, person, bottle, horse, train, TV monitor. Then we performed the training over Google colab that provides free gpu for training the model as seen in Fig. 3 and Fig. 4 and came up with the final trained MobileNetSSD model to perform object recognition and classification.



```
Epoch 8/20
 - 3s - loss: 0.0717 - acc: 0.9800 - binary_crossentropy: 0.0717 - val_loss: 0.4454 - val_acc: 0.8633 -
Epoch 9/20
 - 3s - loss: 0.0589 - acc: 0.9855 - binary_crossentropy: 0.0589 - val_loss: 0.4854 - val_acc: 0.8623 -
Epoch 10/20
 - 3s - loss: 0.0508 - acc: 0.9881 - binary_crossentropy: 0.0508 - val_loss: 0.5248 - val_acc: 0.8590 -
Epoch 11/20
 - 3s - loss: 0.0408 - acc: 0.9921 - binary crossentropy: 0.0408 - val loss: 0.5652 - val acc: 0.8562 -
Epoch 12/20
 - 3s - loss: 0.0329 - acc: 0.9942 - binary crossentropy: 0.0329 - val loss: 0.6110 - val acc: 0.8548 -
Epoch 13/20
 - 3s - loss: 0.0270 - acc: 0.9960 - binary crossentropy: 0.0270 - val loss: 0.6458 - val acc: 0.8528 -
Epoch 14/20
 - 3s - loss: 0.0209 - acc: 0.9974 - binary crossentropy: 0.0209 - val loss: 0.6838 - val acc: 0.8509 -
```

Fig. 3. Screenshots for training



```
Epoch 15/20
 - 3s - loss: 0.0162 - acc: 0.9985 - binary_crossentropy: 0.0162 - val_loss: 0.7100 - val_acc: 0.8524
Epoch 16/20
 - 3s - loss: 0.0124 - acc: 0.9991 - binary_crossentropy: 0.0124 - val_loss: 0.7509 - val_acc: 0.8513
Epoch 17/20
 - 3s - loss: 0.0297 - acc: 0.9994 - binary_crossentropy: 0.0097 - val_loss: 0.7919 - val_acc: 0.8510
Epoch 18/20
 - 3s - loss: 0.0276 - acc: 0.9996 - binary crossentropy: 0.0076 - val loss: 0.8067 - val acc: 0.8510
Epoch 19/20
 - 3s - loss: 0.0260 - acc: 0.9998 - binary crossentropy: 0.0060 - val loss: 0.8325 - val acc: 0.8508
Epoch 20/20
 - 3s - loss: 0.0248 - acc: 0.9999 - binary crossentropy: 0.0048 - val loss: 0.8561 - val acc: 0.8502
```

Fig. 4. Screenshot for training

C. Hardware and Software integration

Mobile Net SSD model that we obtained after training is ported into jetson nano via sd card. Then this module is integrated with the rpi v2 camera. The Jetson Nano Developer Kit has a connector which is compatible with RPi camera. To install lift up the piece of plastic on the Jetson Nano J13 Camera Connector that will keep the ribbon cable in place. When free, the camera ribbon cable is inserted, with the cable contacts pointing inwards towards the Nano module. Press the plastic tab to capture the ribbon cable as shown in Fig. 5.

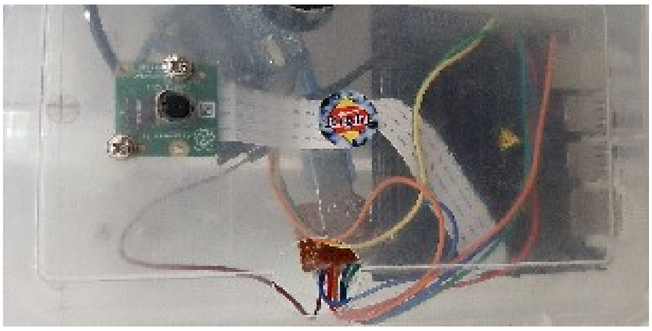


Fig. 5. Integrating Jetson Nano and PiV2 camera

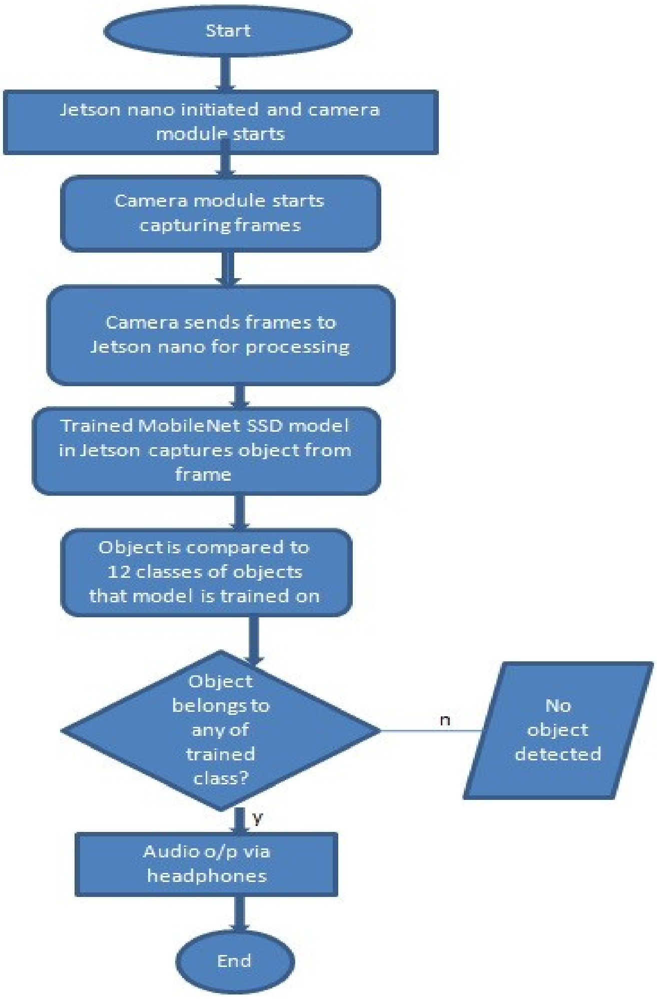

Start
Jetson nano initiated and camera module starts
Camera module starts capturing frames
Camera sends frames to Jetson nano for processing
Trained MobileNet SSD model in Jetson captures object from frame
Object is compared to 12 classes of objects that model is trained on
Object belongs to any of trained class?
n
No object detected
y
Audio o/p via headphones
End

Fig. 6. Flow of object recognition and classification system

Now once the camera is connected to the gpu enabled jetson nano, we power on the device by connecting it with the help of power adapter. Once powered on, we run the module as start up application and perform object recognition and detection which is shown in Fig. 6.

IV. Results

After the integration of hardware and software, we proceeded with the testing of the OCRS and came up with the results as seen in Fig. 7, Fig. 8 and Fig.9.

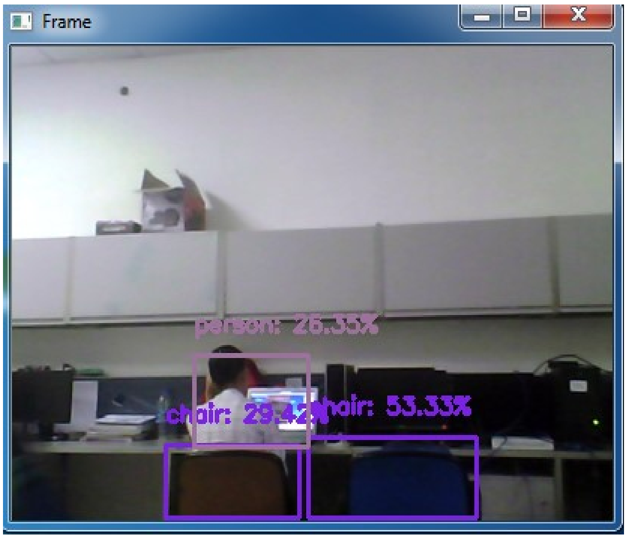


Fig. 7. ORCS screenshot1

We can clearly see that the system is able to perform object recognition as well as is classifying objects based on their class.

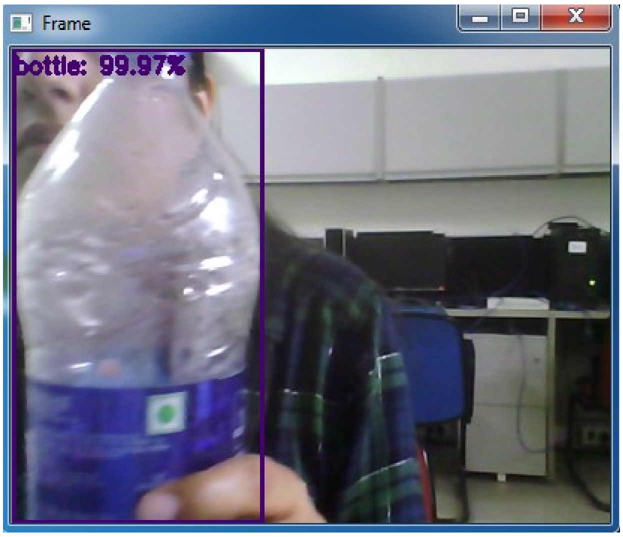


Fig. 8. ORCS screenshot 2

Also it can be seen that, once the object is in complete frame and near to the system the accuracy values are more high rather than that for objects slightly out of frame or being far from the system. For example, we can clearly see in fig8. that accuracy value for a person sitting far but completely in frame is 94.87% while for the one near the frame but not completely in frame is 75.07%.

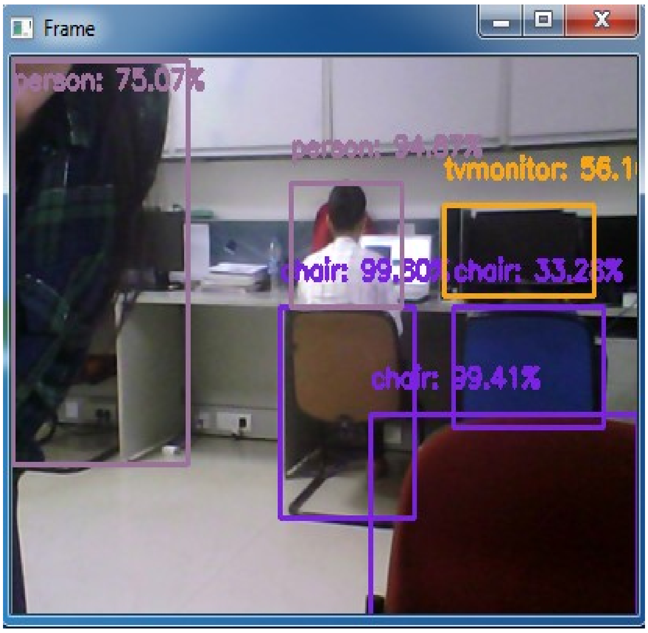


Fig. 9. ORCS screenshot 3

V. CONCLUSION

It is worth mentioning at this point that the aim of this system which is the object recognition and classification has been fully achieved. The proposed system provides low power consumption, a stable and portable navigation solution. The characteristic of this program is that it helps identify the object before the user and know what the class of object is with very fast response time. This device recognizes the object and sends information about the object to the user through headset, rather than simply detecting the object. When creating such an inspiring approach, visually impaired and blind people were high on our agendas in all developing countries. This initial state of the system is sub concept of a substantial project currently under implementation. Various other modules are also being worked on and integrated to the system .Our ongoing research on the exploration of systematic design for these complex systems will aid in better analysis leading to optimal design decisions. Also we are focusing towards the cost-effectiveness of the system in order to be implemented on a large scale.

REFERENCES

[1] Global data on visual impairments ,World Health Organization, 2010.

[2] https://timesofindia.indiatimes.com/india/India-has-largest-blindpopulation/articleshow/2447603.cms

[3] http://www.myeyeworld.com/files/eyebanks.htm

[4] "Analysis of Blind Pedestrian Deaths and Injuries from Motor Vehicle Crashes," http://files.meetup.com/211111/analysis-blind-pedestrian-deaths.pdf.

[5] Nabila Shahnaz Khan, Shusmoy Kundu, Sazid Al Ahsan, Moumita Sarker and Muhammad Nazrul Islam "An Assistive System of Walking for Visually Impaired"ICCCMEE-2018.

[6] T.Miura, Y.Ebihara, M.Sakajiri and T.Ifukube, "Utilization of auditory perceptions of sounds and silent objects for orientation and mobility by visually-impaired people," IEEE Trans. Systems, Man and Cybernetics, Part A:Systems and Humans, , pp.1080-1082, Oct 2011.

[7] Samir Patel, Amit Kumar, Pradeep Yadav, Jay Desai and Deepali Patil "Smartphone Based Obstacle Detection for Visually Impaired People"ICIIECS-2017.

[8] J. Berry, "Apart or a Part? Access to the Internet by Visually Impaired and Blind People, with Particular Emphasis on Assistive Enabling Technology and User Perceptions," Information technology and disabilities, vol. 6, no. 3-4, pp. 1–16, 1999.

[9] J. S´anchez and N. de la Torre, "Autonomous Navigation through the City for the Blind," in Proc of the 12th international ACM SIGACCESS conf on Computers and accessibility. ACM, 2010,pp. 195–202.

[10] "App Improves the Safety of Blind Pedestrians in Cities," https://phys.org/news/2015-01-app-safety-pedestrians-cities.html, 2015.

[11] R. Kedia, K. K. Yoosuf, P. Dedeepya, M. Fazal, C. Arora and M. Balakrishnan, "MAVI: An Embedded Device to Assist Mobility of Visually Impaired," 2017 30th International Conference on VLSI Design and 2017 16th International Conference on Embedded Systems (VLSID), Hyderabad, 2017, pp. 213-218.

[12] T. D. R. Girshick, J. Donahue and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation". Computer Vision and Pattern Recognition (CVPR),2014, 2014.

[13] R. Girshick. "Fast r-cnn", In Proceedings of the IEEE International Conference on Computer Vision, pages 1440–1448, 2015.

[14] S. Ren, K. He, R. Girshick, and J. Sun. "Faster r-cnn: Towards real-time object detection with region proposal networks". In Advances in neural information processing systems, pages 91–99, 2015.

[15] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. "You only look once: Unified, real-time object detection", In Pro of the IEEE Conf on Computer Vision and Pattern Recognition, pages 779–788,2016.

[16] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C.Berg. Ssd: "Single shot multibox detector",.In European Conference on Computer Vision, pages21–37. Springer, 2016.

[17] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam. "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications". In ArXiv , 17 Apr 2017.

[18] JetsonNanoDeveloperKit,https://developer.nvidia.com/embedded/jetson-nano-developer-kit.