Real Time Object Detection for Visually Challenged Persons

Anitha.J, Subalaxmi.A, Vijayalakshmi.G

Abstract: ***The visually impaired and blind people face various challenges in their day to day life. The objective of the proposed work is to develop an application for visually challenged persons based on the Android smart phone. It will eliminate the need for dedicated devices and other wearable devices to assist them to recognize objects as they move around. The Android application helps the visually impaired to navigate independently using real-time object detection and identification technology. The application makes use of the image processing technique to detect the object and speech synthesis to produce the voice output. The objective of the system is to detect real time objects which are scanned through the mobile camera and notify the blind persons about the object through audio or vocal information. The detection of images on moving objects has been a significant research area in computer vision which has been highly worked upon, and integrated with residential, commercial and industrial environments. Due to lack of data analysis of the trained data, and dependence of the motion of the objects, inability to differentiate one object from the other has led to various limitations in the existing techniques which include less accuracy and performance. Hence, Fast R-CNN (Region-based Convolutional Neural Networks) algorithm has been implemented to detect the object with high accuracy and processing speed. The detected image information is provided as a voice output using a speech synthesizer to the visually challenged persons to assist them in their mobility.***

Index Terms: Computer Vision, Convolution Neural Networks, Deep Learning, Mobile Application, Object Detection, Tensorflow

I. INTRODUCTION

Millions of people suffer from vision impairment in one or other way. One of the important senses that is very essential for a human being to lead a normal life is the vision. Many People suffer from blindness face quite difficulties while moving around the surrounding environment. This condition leads to the need for guidance or assistance for every action of the handicapped person. Blindness makes the normal, professional and social life of the people very difficult. Human vision has the extraordinary capacities of storing billion of images in the brain and realizing the images by comparing with the preimages. But still some people are not blessed with the vision and some suffer from the retinal diseases. By using computer vision techniques, the quality of life of blind people could be improved. One of the major research areas in computer vision is the object detection and identification technology which is used for object recognition and prediction of moving objects[7][8][9]. The features of objects like intensity, edge, and shape are used to recognize objects from the input image by using object detection technology. The Fast R-CNN algorithm and the trained Tensorflow models are used to detect objects in the image.

The computer vision application is deployed in the Android platform due to the wide popularity and usage of the Android-based devices. According to Gartner's survey, the Android mobile platform has gained 70% share of the smartphone market at the end of 2012[2]. The application is easy to use and it is equipped with speech synthesis so that the detected object is communicated to the blind people as voice output.

H. Mao, S. Yao, T. Tang, B. Li, J. Yao and Y. Wang[1] has proposed the Convolution Neural Network based for object detection in embedded systems. The Fast R-CNN algorithm has been used for the object detection and recognition task which helped in overcoming the various limitations while integrating the object detection technique into the embedded system. High energy efficient and close to real-time application results is achieved through this implementation of fast R-CNN which is much better than the other convolution neural network algorithms.

Jae Sung Cha, Dong Kyun Lim, and Yong-Nyuo Shin[3] have discussed the main features of software modules developed for Android smartphones that are dedicated for the blind users. The main module can recognize and match scanned objects to a database of objects, e.g. food or medicine containers. The two other modules are capable of detecting major colors and locate the direction of the maximum brightness regions in the captured scenes

Aishwarya Sarkale, Kaiwant shah, and Anandji Chaudhary[4] have discussed the machine learning approach for object detection using Artificial Neural Networks for recognition of static objects.

II. PROPOSED SYSTEM

The visually impaired individuals face many difficulties in performing their mundane tasks, especially in self-navigating at an environment which is strange for them. They are unable to detect the object around them and they are unable to identify the object around them accurately in most cases. This application will help them to overcome their difficulties in self-navigation which help them to lead their life like a normal person.

The detection of an object cannot be achieved with high accuracy in normal cases, as most of the application provides the detection of objects captured as an image rather than detection in a live streaming video.

Revised Manuscript Received on June 8, 2019

Anitha.J, Department of Information Technology, Sri Ramakrishna Engineering College, Coimbatore, India.
Subalaxmi.A, Department of Information Technology, Sri Ramakrishna Engineering College, Coimbatore, India.
Vijayalakshmi.G, Department of Information Technology, Sri Ramakrishna Engineering College, Coimbatore, India.

Smartphones are commonly used by all people and it has been turned into one of the daily needs of human beings. Even though there are sophisticated techniques to address the above-mentioned issues, they are not cost-efficient and not easily accessible by the visually challenged persons. The proposed Android application is cost effective and easily accessible by blind people. The fast R-CNN algorithm has been used for classification of the real-time object from continuous streaming through the mobile camera and the detected object is given as a voice output to the blind people. The object is detected initially and the produces output in the text format which is then the text is converted into speech.

The proposed system captures the image through the continuous video stream rather than taking the picture of each object every time. The object detection and recognition have become easier when it is detecting from the image but in the case of the video stream, the processing speed should be high to detect all the object in a frame. The functional process is to capture the images through the camera of the Android smartphone and process it through image processing algorithms.

The object detection module and the voice output module produce the required output of the system. In object detection module the scanned objects are first segmented and then form multiple regions. Then the images in the multiple regions are classified and identified using the Fast R-CNN algorithm. The voice output module acquires a 3D real image of any text constraints area and to convert this image into text and providing audio output using speech processing. The system uses the Google API for conversion of text to speech conversion.

III. METHODOLOGY

The object detection and recognition have to be accurate and the detecting speed of the object should be high so that the navigation for the blind people would be easier. For such high-speed detection, fast R-CNN algorithm has been implemented[10]. The system does not work on a massive number of regions, the R-CNN algorithm process the image into a bundle of boxes and checks if any of these boxes contain any object. The selective search methodology has been used by the fast R-CNN algorithm to extract these boxes from an image. The varying scales, colors, textures, and enclosure are the four regions that form an object. The selective search identifies the patterns in the image and based on those patterns they split the image into various regions.

The first step is to capture the objects using the camera of the smartphone for this the application should get the camera permission of the device. Then the image is sub-segmented so that multiple regions can be formed from a single image. The algorithm then combines similar regions to form a larger region and finally produce the region of interest.

The pre-trained convolutional neural network is trained again based on the number of classes that has to be detected then the region of interest is identified. Based on the region of interest, the objects and the backgrounds are classified. For each identified object in the image, tighter bounding boxes are generated based on the linear regression model. The input image to the CNN, generate the convolution feature maps. The regions of proposals are extracted using those maps. The object is detected within 2 seconds. The recognized information is in the text format but to the blind people, the information has to be provided in the vocal format.

The CNN model deployed for training is the MobileNet[5], which is a low-latency parametrized models which meet the resource constraints for various use cases.

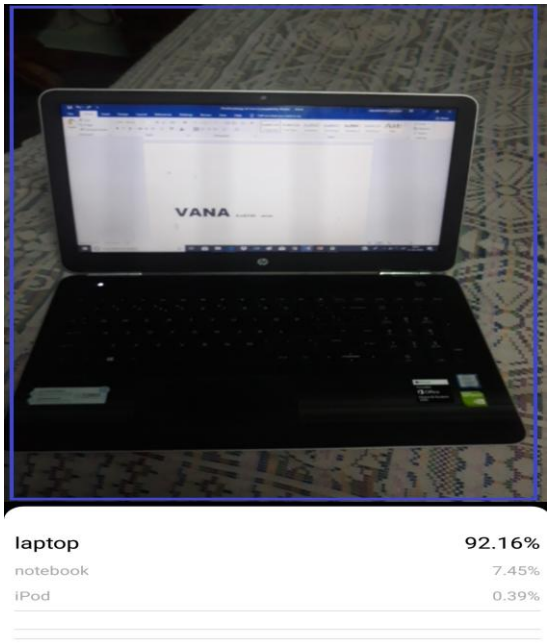

Fig.1: Detection of Laptop

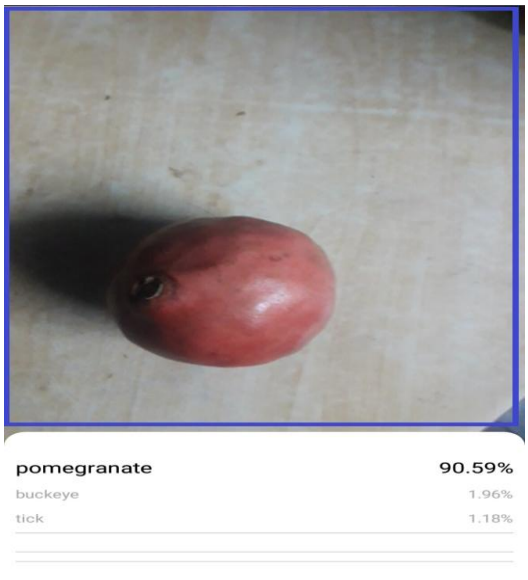

Fig.2: Detection of Pomegranate

Fig.1 and Fig.2 illustrates the detection of objects in real time captured through the smartphone camera with an accuracy of 92.16% and 90.59% respectively.

IV. RESULTS AND DISCUSSION

The proposed system uses Google's Tensorflow object detection API[11]. The dataset acquired has been trained with the pre-trained SSD MobileNetV1 and SSD MobileNetV2 models [6] and the accuracy obtained from both the models are compared. The results reveal that SSDMobileNetV2 model provides better accuracy with low latency in detecting the objects. For detection when paired with the Tensorflow Lite the new model is about 35% faster with the same accuracy than MobileNetV1.

V. CONCLUSION AND FUTURE WORK

The proposed system is a simple, economical guidance system which provides constructive assistance and support for blind and visually impaired persons. The results show that the system is efficient and unique in identifying the object that the blind person may encounter. It also resolves the limitations of other systems that are related to mobility-oriented problems that influence the blind people in their environment.

The future scope of the project determines to recognize multiple objects in a view with better accuracy and less detection time. The extension of this system identifies any kind of entity with faster frame rate. The text to speech module has also been developed according to the futuristic pace. Instead of using the pre-trained models, self-trained models can be used. The model can be trained to recognize objects which are frequently encountered by the user. Thus, it can be customized for the specific needs of the user and ensures safer navigation. The addition of a face recognition feature, the application can be trained to store the information about the people closely related to the person, which would help them to differentiate between peers and strangers.

REFERENCES

1. H. Mao, S. Yao, T. Tang, B. Li, J. Yao and Y. Wang, "Towards Real-Time Object Detection on Embedded Systems," in IEEE Transactions on Emerging Topics in Computing, vol. 6, no. 3, pp. 417-431, 1 July-Sept. 2018. doi: 10.1109/TETC.2016.2593643
2. Gartner Says Worldwide Sales of Mobile Phone [Available: http:/www.gartner.com/newsroom/id/2237315]
3. Jae Sung Cha, Dong Kyun Lim and Yong-Nyuo Shin "Design and Implementation of a Voice Based Navigation for Visually Impaired Persons" International Journal of Bio-Science and Bio-Technology, vol. 5, No. 3, June, 2013
4. A. Sarkale, K. Shah, A. Chaudhary and T. Nagarhalli, "An Innovative Machine Learning Approach for Object Detection and Recognition," 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), Coimbatore, 2018, pp. 1008-1010.doi: 10.1109/ICICCT.2018.8473221
5. MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications, Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, Andreetto M, Adam H, arXiv:1704.04861, 2017.
6. MobileNetV2: Inverted Residuals and Linear Bottlenecks, Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. arXiv preprint. arXiv:1801.04381, 2018.
7. K. Matusiak, P. Skulimowski and P. Strumiłło, "Object recognition in a mobile phone application for visually impaired users," 6th International Conference on Human System Interactions (HSI), Sopot, 2013, pp. 479-484.doi: 10.1109/HSI.2013.6577868
8. Y. Zhong, P. J. Garrigues, and J. P. Bigham, "Real time object scanning using a mobile phone and cloud-based visual search engine," presented at the Proceedings of the 1 5th International ACM SIGACCESS Conference on Computers and Accessibility, Bellevue, Washington, 2013.
9. Yi, Chucai & Tian, Yingli & Arditi, Aries. (2014). Portable Camera-Based Assistive Text and Product Label Reading from Hand-Held Objects for Blind Persons. Mechatronics, IEEE/ASME Transactions on. 19. 808-817. 10.1109/TMECH.2013.2261083.
10. R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, 2015, pp. 1440-1448. doi: 10.1109/ICCV.2015.169
11. Real Time Object Detection with Tensorflow Detection Model [Available:https://towardsdatascience.com/real-time-object-detection-with-tensorflow-detection-model-e7fd20421d5d]

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication