IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

REAL TIME OBJECT DETECTION AND RECOGNITION SYSTEM TO ASSIST THE VISUALLY IMPAIRED

[1]DSOUZA ELSTON RONALD, [2]DHANYA BHAT [3]JOHN ALSTEN TAURO, [4]S HARSHITH, [5]MERLYN MELITA MATHIAS

[1]Student, [2]Student, [3]Student, [4]Student, [5]Assistant Professor

[1,2,3,4,5] Department of Computer Science and Engineering

[1,2,3,4,5] Alva's Institute of Engineering and Technology, Moodbidri, India

Abstract: The human visual system is fast and accurate and can perform complex tasks like identifying multiple objects and detect obstacles with minute precision. Also humans can easily detect and identify objects present in the surroundings or in an image. Now with developments in web and computer vision technologies, we are detecting objects in images and videos with high speed and accuracy using machines. We are using the same technology for helping visually impaired in their daily routine.

Index Terms **– ReactJs , TensorflowJs , Object Detection , COCOSSD.**

I. INTRODUCTION

In the present world, there are lakhs of people who have the problem of visual disability. There are approximately as many as 28.5 crore people have the problem of visual disorder as per the report of WHO, of which 3.9 crore people are blind and 24.6 crore people can see things at a very low scale. People with visual impairment have a variety of options for managing their daily routines, including some assistant to take care of them, but often have significant difficulties moving independently without assistance. There are various adverse effects of having poor eye sight. Some of which are Reading/Writing, Sightseeing, Daily chores, etc. There has been far too much progress in modern science and computer vision technologies to justify conducting research in order to build an ideal tool to assist the visually impaired.

In this study, we have devised a method to assist visually impaired people in detecting objects in their environment. We have proposed to create a web application that would detect the objects and various things in the surrounding within the help of certain machine learning techniques. The system with detect and recognize the objects and present the same to the user in an audio format.

II. PROBLEM STATEMENT

- To develop a real-time object detection system to help the visually impaired persons. Objects in the immediate environment will be identified in real-time and displayed to the visually impaired user..
- To generate the results obtained from the detection to the person in the form of audio. The audio results generated will be given to the user in the order in which the objects occur in the video frame.

III. LITERATURE SURVEY

In order to create an object detection model, it is very much necessary to inculcate machine learning techniques. We need to create certain data models for the computer system to detect the objects. In order to achieve , we need to capture labelled images of certain classes and store it in individual datasets. Larger the data set, the greater is the accuracy of the trained model, which is very much mandatory to achieve better object detection.

Deep learning techniques such as Convolutional Neural Networks (CNN) are being used for object detection as shown in figure 1, are used on a large scale nowadays. [1] The Convolutional Neural networks consist of multi-layered architecture including an input layer, many hidden layers and an output layer. The image is extracted from the video and given to the input layer. In general, the hidden layers in a CNN operate as filters, receiving input from the higher layer, processing it with a certain pattern or feature, and then sending it to the next layer. The output layer classifies the image based on the input from the hidden convolutional layers.

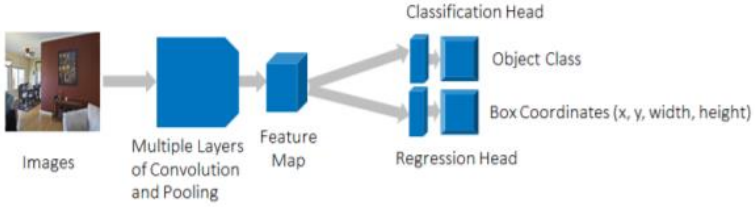

Figure 1. Convolutional Neural Network

Following are implementations of CNN for object detection:

3.1. R-CNN:

It is basically a region based approach which uses Convolutional Neural Networks as shown in figure 2. Ross Girshick, Jeff Donahue, Trevor Darrell and Jitendra Malik [2] in their research were able to achieve a mean average precision (mAP) of 53.7% with more than 30% improvement [2] as compared to the previous research..

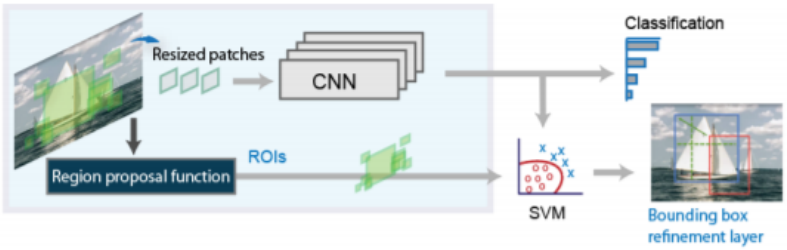

Figure 2. R-CNN: Regions with CNN features [2]

3.2. FAST R-CNN:

Ross Girshick [3] developed Fast R-CNN which is an improvement over R-CNN. In Fast R-CNN as shown in figure 3, the entire image is processed to detect feature maps which are furthermore fed to the Region of Interest(ROI) layer to calculate region of interest in the given map. The output thus obtained are probabilities of existence of the object in the frame and its final output is calculated based on the threshold value after filtration of values obtained.

All network layers are handled in a single stage during the training. With more appropriate training schemes, it leads to increased accuracy and efficiency, as well as more storage. Fast R-CNN has the disadvantage that, while it improved object detection performance, it is still too slow to be used in real-time object detection applications.

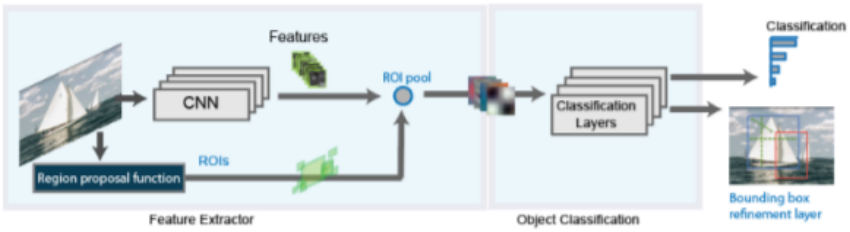

Figure 3. Fast R-CNN [3]

3.3. FASTER R-CNN:

Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun [4] developed an improvement on the Fast R-CNN algorithm called Faster R-CNN shown in figure 4. It consists of two modules: a deep convolutional network based on regions and a Fast R-CNN detector that use these areas.

For object detection, this system creates a uniform network. This algorithm is incredibly fast because the time spent on each image is decreased to 10ms. This strategy, however, requires a lot of computation, making it unsuitable for applications that must run on CPUs with limited processing capability.

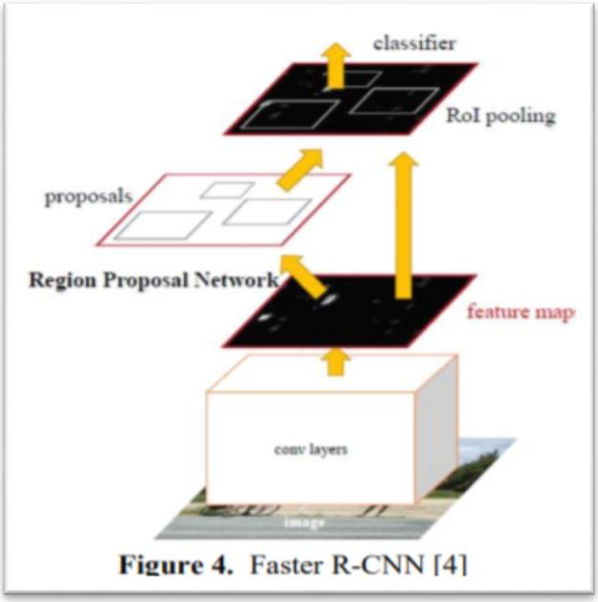

Figure 4. Faster R-CNN [4]

3.4. YOLO:

To overcome the drawbacks of the above algorithms, it is better to use a regression/classification First International Conference on Advances in Physical Sciences and Materials Journal of Physics: Conference Series 1706 (2020) 012149 IOP Publishing doi:10.1088/1742-6596/1706/1/012149 4 based framework rather than the region proposed based frameworks mentioned above.

Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi [5] in their paper titled 'You Only Look Once: Unified, Real-Time Object Detection' defined a framework called (You Only Look Once) or YOLO. A human user just requires one glance at an object to know its properties such as which object it is, its dimensions, its location, depth, etc. The same methodology is used in this framework to detect objects with a single look. The overall detection of object takes place as shown in figure 5.

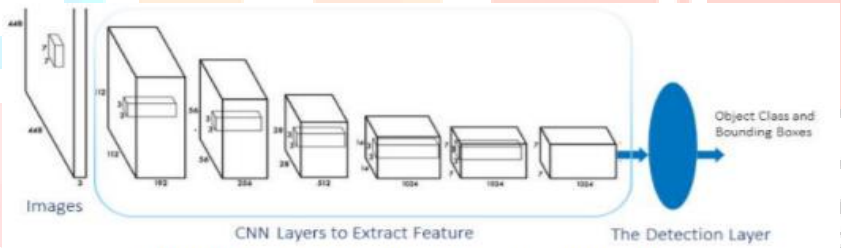

Figure 5. YOLO Network Architecture [5]

IV. EXISTING SYSTEM

In the present world, there are many computer vision systems to help the visually impaired in various stages of their life. These mainly include AI and GPS based navigation systems, Wearable Goggles for visually impaired based on Augmented Reality, video calling apps for the visually impaired to ask for assistance and many other technologies. These systems are built to work only in specific cases or conditions, and cannot be used on a large scale. There are cases wherein the people with visual impairment have to realize about their surroundings, which is not easily possible with the existing systems.

Drawbacks of existing system:

- They are expensive which makes it difficult for the visually impaired people to afford it.
- These systems have functionality which is very complex, making it difficult to be used by the visually impaired.
- No personalised user experience and the systems are not real-time.

V. PROPOSED SYSTEM

We propose to build an React based web application which can detect the objects existing in the surrounding. This becomes very easy for the visually impaired user to move around freely without the fear of being crashed.

Advantages of the proposed system:

- Provides real-time results based on the videos captured .
- Reliability: The visually challenged user may rely on the system to provide accurate results. Differences between distinct items, such as a chair and a table, can be easily identified and studied depending on the video quality.
- Simple User Interface makes it simple to use.

VI. METHODOLOGY

6.1.TOOLS AND TECHNOLOGIES USED:

ReactJS: ReactJS is a free and open-source front-end JavaScript library[3] for building user interfaces or UI components. Facebook and a community of individual developers and businesses maintain it. [4] [5] [6] React can be used to create single-page or mobile applications as a foundation.

TensorFlow: An open source software library and framework used for neural network and machine learning applications. It has a flexible architecture and can be deployed on servers, laptops, mobile phones, edge devices, etc.

COCOSSD: COCO-SSD is an object detection model powered by the TensorFlow object detection API. Single-Shot MultiBox Detection (SSD) is an acronym for Single-Shot MultiBox Detection. In the COCO Dataset, this model can detect 90 different classes.

OpenCV: Used for performing real time computer vision tasks.

YOLO: Provides a framework that allows detection of objects in near real time speeds. For deployment in a mobile device we are using Tiny YOLO, which is a lightweight YOLO framework for mobile and edge devices.

6.2. DEVELOPMENT ENVIRONMENT:

Visual Studio Code

6.3. SYSTEM FLOW DIAGRAM:

The Flow of Execution of our project is as shown in figure 6.

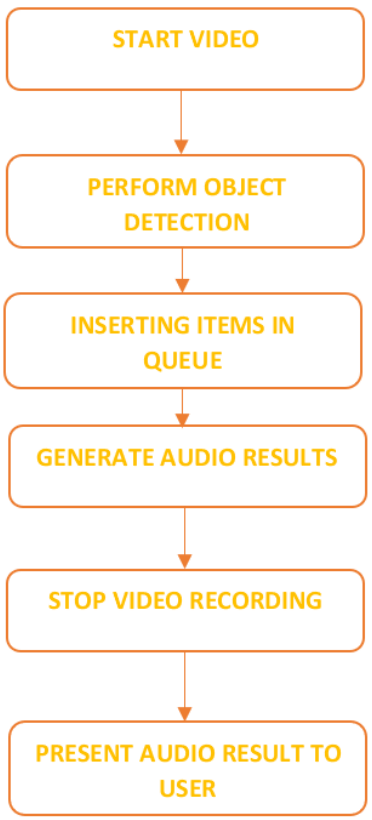

Figure 6. System Flow Diagram

6.4. IMPLEMENTATION:

In the initial we created a React Application in order to build the front-end of our Application or in other words the User Application(UI). In order to get In the object detection , we used TensorFlow module and imported the required modules into our project. In the next step, we created a promise to get the detection as a promise and we created a rectangular box around the detected object.

After getting the detections , we used the SpeechSynthesisUtterence to generate the audio results. This Web Speech API interface represents a speech request. It contains the content the speech service should read and information about how to read it. To get the audio results , we create an array containing the detections and pass the class in the speech synthesizer in order to generate the audio voice alert to assist the visually impaired.

VII. TESTING:

In this proposed project, System testing was performed to ensure that the Web Application was performing object detection as per our requirements. Various test cases were developed for various classes of objects to ensure that the system was performing object detection in a satisfactory way and producing acceptable output. The result is shown in Table 1.

Sample Input	Expected Output	Actual Output	Accuracy Score	Detection Status
Person	Person should be detected	Person Detected	0.92	Successfully Detected
Bottle	Bottle should be detected	Bottle Detected	0.90	Successfully Detected
Chair	Chair should be detected	Chair Detected	0.87	Successfully Detected
Clock	Clock should be detected	Clock Detected	0.91	Successfully Detected
Cell Phone	Cell Phone should be detected	Cell Phone Detected	0.92	Successfully Detected

Table 1. Testing of Object Detection model with accuracy score.

VIII. RESULTS:

The sample output showing detected objects with bounding boxes, respective labels are shown in figure 7. These photographs were captured using a personal computer and the object detection web application.

Figure 7: Sample Outputs

IX. CONCLUSION:

One of the main issues involving object detection is object classification and object localisation within a scene. The application of deep neural networks has aided in the detection of objects. However, deploying such techniques on Computer devices requires high computational and memory resources. Hence using small deep neural network architectures for object detection such as TensorFlow are giving good results and show that they can be used for real time object detection which can help the visually challenged.

REFERENCES

[1] Zhao Z, Zheng P, Xu S and Wu X Object Detection withDeep Learning: A Review2019IEEE Transactions on Neural Networks and Learning Systems 30 3212
[2] Girshick R, Donahue J, Darrell T and Malik J Rich feature hierarchies for accurate object detection and semantic segmentation 2014 IEEE Conference on Computer Vision and Pattern Recognition(Columbus, OH)pp 580-587
[3] Girshik R Fast R-CNN 2015IEEE International Conference on Computer Vision(Santiago) pp 1440-1448
[4] Ren S, He K, Girshick R and Sun J Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks 2017 IEEE Transactions on Pattern Analysis and Machine Intelligence 39 1137
[5] Redmon J, Divvala S, Girshick R and Farhadi A You Only Look Once: Unified, Real-Time Object Detection 2016IEEE Conference on Computer Vision and Pattern Recognition (Las Vegas) pp 779-788
[6] Redmon J and Farhadi A YOLO9000: Better, Faster, Stronger 2017IEEE Conference on Computer Vision and Pattern Recognition (Honolulu) pp 6517-6525
[7] Redmon J and Farhadi A YOLOv3: An Incremental Improvement 2018
[8] Schroff F, Kalenichenko D and Philbin J FaceNet: A Unified Embedding for face recognition and clustering 2015IEEE Conference on Computer Vision and Pattern Recognition (Boston) pp 815-823