Android-based object recognition application for visually impaired

Akilesh Salunkhe[1], Manthan Raut[1], Shayantan Santra[1] and Mrs. Sumedha Bhagwat[2]

[1] Department of Information Technology, Ramrao Adik Institute of Technology, Nerul, India
[2] Mentor from Department of Information Technology, Ramrao Adik Institute of Technology, Nerul, India

Abstract - Detecting objects in real-time and converting them into an audio output was a challenging task. Recent advancement in computer vision has allowed the development of various real-time object detection applications. This paper describes a simple android app that would help the visually impaired people in understanding their surroundings. The information about the surrounding environment was captured through a phone's camera where real-time object recognition through tensorflow's object detection API was done. The detected objects were then converted into an audio output by using android's text-to-speech library. Tensorflow lite made the offline processing of complex algorithms simple. The overall accuracy of the proposed system was found to be approximately 90%.

Keywords - Tensorflow, TensorFlow Lite, Machine learning, Computer vision, Object recognition, and Android

1. Introduction

Out of all the human senses, eyesight is the most important sense. It allows a person to analyse and understand their surrounding environment. At least more than 285 million people are facing eyesight challenges or are visually impaires as per the data gathered from WHO. Eyesight issues can cause disturbance in the daily activities of a person. Identifying objects in day-to-day lives, reading text, crossing a road are a few examples of such problems. The proposed system is a simple android app based object detection application named "digital eyes" to help the visually impaires. This application tries to replicate the human eye with the help of a smartphone camera using object detection. The normal life of the people can be improved by using modern computer vision techniques. Object detection is one of the methods of computer vision which have had many broader applications over recent years. Object detection technology [3] uses the contrasting features intensity, edge, and shape to recognize the object from the input image. The advancement in object detection algorithms has enabled us to incorporate complex algorithms into an android application. The SSD algorithm and the trained tensorflow models are used for object detection in our android application. Image processing techniques are now currently being used in object detection domain for various applications. [4], which are used for social and many other applications. The purpose of this project is object detection for the visually impaired by using speech feedback and extracting features from live camera feeds. The application is easy to use and it is equipped with speech synthesizing so that the detected object is

communicated to the blind people as voice output. In section ii, describes a comparative study of several object detection methods and their statistics in tabular format. Section iii describes the system description along with the technology stack used in the system. In sections iv and v, this paper explains about the advantages and disadvantages of the proposed system in this paper. In section vi it explains the proposed system technically giving an idea about the important components in the system followed by a conclusion and future scope.

2. Related work of different algorithms

Table no. 1. Related work of different algorithms [13]

Mode Type	Latency of model	(Accuracy)mAP	FPS	Real-Time
R-CNN	High	~60	<1	NO
Fast R- CNN	Medium	~70	<1	NO
Faster R-CNN	Medium	~70	7	NO
YOLO	Low	~60	46	YES
SSD	Low	~74	59	YES

The table above shows the comparison of the various models regarding latency, average accuracy (map), images per second (FPS), and whether they could be used in real-time or non real-time applications. Analysis of the outcome of a system or algorithm was based on certain parameters. The most common parameters were efficiency, time, resources, accuracy, etc. that were undertaken in almost all analyses. On applying the general parameters over the R-CNN method of object detection, the results showed that it was much faster than the old methods based on the classification methods. Instead of a huge number of regions, RCNN used the selected search to retrieve only 2000 regions per frame. So the feature extraction would run over only 2000 regions. A new version of R-CNN called fast R-CNN was far better than R-CNN as it did not transmit 2000 regional proposals to CNN each time. Instead, the CNN operation was carried out once per frame. The implementation of the new method was similar to the previous methods but instead of a selective search algorithm, an independent network was used to anticipate the proposed regions. A new method You Only Look Once (YOLO) was proposed for the recognition of the objects. As the above methods used the suggested regions to identify the objects in the image, it never considers the entire image. Regions with a

high likelihood of having objects were instead passed through the system for object detection. But in YOLO there was only a single convolutional network and the entire image was analysed by that network [13]. The SSD was very close to R-CNN in terms of accuracy. This made the SSD the best algorithm that balances speed and accuracy. Due to this, the SSD algo was used broadly in object based detection systems.

3. System description

3.1 Object Detection

Object detection is a computer technology related to computer vision and image processing that deals with detecting the presence of objects with a limitative box and types or classes of objects located in an image in digital images and videos [5]. By Using object detection, visually impaired users can understand their surrounding environment without any challenges and remain independent of others

Input: a picture with one or several objects, like a photograph.

Output: one or more limiting boxes (e.g., defined by a dot, width and height) and a class label for each limiting box.

3.2 Tensorflow

Tensorflow is an open-source software library framework, which was used to implement object detection and recognition. This consists of a pre-trained object detection model, which uses an SSD algorithm to detect objects more efficiently and accurately. This method of object detection uses the COCO mobile net SSD v1 model, which also consists of datasets of 80 object categories, which are commonly found around us.

3.3 Android Studio

Android SDK was being used to make the android application, which can be easily used by visually impaired users for detecting objects and understanding their surrounding environment. The application's front end and backend are implemented using this platform. This platform provides all the libraries and packages required for implementing this system.

3.4 Mobile device-based object recognition

With the ever-increasing advance in smartphone technology, many have tried to implement identification of objects on smartphones [14]. Thanks to smartphones, applications adapted to the blind can be made user-friendly, portable and widely available, eliminating the need for special equipment to do the processing. However, because of a mobile phone's limited processing power, some such applications rely on a client-server architecture [19]. One such well-known application is google goggles, which requires an internet connection and cannot add new images to the application

systems for visually impaired users that utilize only local processing bases on the calculation resources of a smartphone like in this document an application was developed for android that performs all the processes locally giving the result in the form of auditory feedback [5]. This implementation uses the functionalities of the SIFT. However, the work proposes in this document is not dedicates to real-time processing. Commercial object recognition applications are available to blind individuals. Looktel developed two purpose-based apps with a particular focus on the visually impaired and they are: Looktel recognizer and Looktel money reader, which was designed for IOS-enabled devices. Both applications perform in real time and do not require an internet connection to operate. Looktel recognition [17] works by pronouncing object names when they are paired to a database that is normally pre-built for the user by a blind person.

4. Proposed system

The system was being implemented on an android app that detects diverse objects in live video feed along with a real-time text reader. In proposed system an object recognition android app was developed used google's tensorflow object detection API model which implemented using SSD algorithm and real-time text reader feature which was using google's TTS engine and google played services mobile vision API which describes the used of text recognizer class to detect text from a real-time video feed. SSD algorithm based object detection model was used for real-time and offline object detection.

4.1. System Outline

The system uses a mobile phone to capture incoming data in live video feed. The application gives two options to the user for detecting objects and reading the text. The camera of the application is automatically accessible and it begins to capture the surrounding objects and texts. Data is sent to the TensorFlow object detection model for processing and later it identifies the class of the objects detected and returns the output as spoken feedback. In the case of reading text, it uses Google play services mobile vision API which consists of TextRecognizer class to detect texts from real time video feed and sends it to google's TTS (text-to-speech) engine for converting text to speech and thus reads out the text detected by the phone's camera.
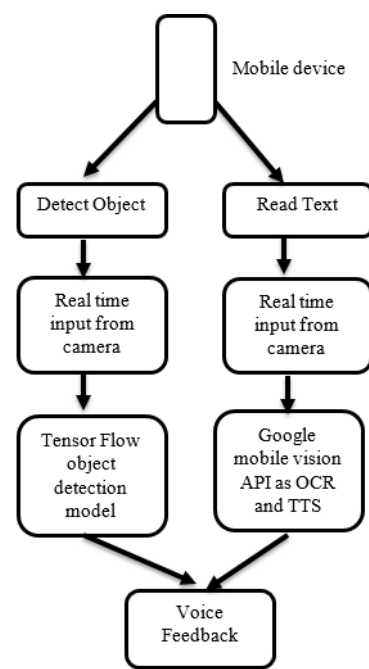

Fig no.1 Flowchart of the proposed system

4.2. Implementation

The system was developed by integrating various technologies, which are mentioned below. Android SDK was used for developing the application because it is the official Integrated development environment (IDE) designed specifically for developing Android applications. [8]. The Android framework supports capturing images and video through the android.hardware.camera2 API or camera Intent. It is a package used for capturing real-time video for object detection and reading text.

Tensorflow library is used for implementing object detection models inside the android application. It provides high performance numerical computing. It has a flexible architecture, making it easy to deploy the calculation through a variety of all possible platforms[9]. SSD-Mobile-Net-COCO model was being used for real time processing. The SSD architecture is a single convolution network that learns to predict bounding box locations and predicts the detected object in the form of limitation boxes. The system uses two object detector modules and real time text reader.

4.2.1 Object Detection: The app is using the SSD-MobileNet-COCO model detecting objects. It utilizes only one neural network for the entire input image. The network model then separates the real time input img into various different regions and predicts the objects using a quadrilateral surrounding the object along with its probability score [12].

4.2.2. Text Reader: The text reader is using Google mobile vision API for detecting texts in real time just like an OCR and then converts the text into speech by using Googles TTS library which was available in android SDK while making application.Using this a user can easily read the menu cards in restaurants, texts on objects(medicines, food,etc.), hotel room no., or even read a paper document, etc.

TextRecognizer: This object processes the images and determines the text contained therein. Once initialized, it can be used to detect text in all picture types.

Reading text feature was implemented using Google Text-to-Speech, which speaks the detected text and acknowledged objects.

4.3. Dataset

In this project, the Common-Object-in-Context (COCO) dataset was used for training the model i.e. SSD MobileNet model, which was able to recognize 81 different categories [22].

5. Analysis of system

In this project, Tensorflow's Object detection model was used which used an SSD algorithm in the backend, and it was able to work by balancing between accuracy and speed. This model successfully detects approximately 81 objects. This model has 74.3 mAP (Mean Average Precision) value, which is highest among the models targeted for real-time processing. After implementing this project, it was expected for a speech feedback for the object which was being detected. But the same object was getting called out multiple times as it got detected. But it will be undesirable to speak out the same object name even if the detection result is the same. Also, it was undesirable if two object names spoken are overlapping or very closely that the user would not be able to distinguish. To solve this problem, if one object was getting detected in the first frame and was speaking out. Then the program will not speak out its class for next five seconds, even if it gets detected. By this the problem of detecting a single object multiple times was being solved.

Upon testing on several objects we found that the results may vary sometimes and the accuracy for detecting objects depends on several parameters. In order to improve the accuracy the model needs to be trained by using objects under different scenarios and test cases. Different cases may include under different light, distance, state of object, direction of object.,etc.

Following are some of the results which show the prediction value of detecting an object accurately. Thus giving us an idea about the accurate performance from the model while detecting objects correctly. Possible Objects can be detected at a time but only objects which have precision value higher than fixed threshold value will be told to visually impaired users using voice output feedback. Multiple objects can even be accurately detected at a single time.

Read text example
(a)

Detecting teddy bear
(b)
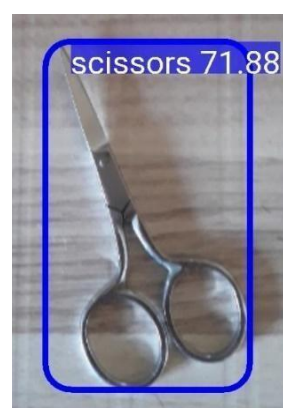
Detecting scissors
(c)

Detecting bottle
(d)

Detecting TV
(e)

Detecting a Laptop
(f)

Table no. 2. Confidence results from object detection of some objects under different cases.

Objects	Normal	Dark	Shadowed object	Long dist.
Mobile	95%	81%	78%	66%
Person	93%	78%	71%	73%
Dog	89%	77%	-	69%
Orange	92%	82%	67%	65%

Laptop	96%	88%	75%	69%
car	91%	-	-	74%

6. Conclusion

In this paper, a model, which was using the SSD algorithm, was made use for creating an application for object detection, which uses TensorFlow object detection API for working offline, and giving maximum accuracy as possible. An object detection API was used for the purpose of detecting objects. The future work includes further enhancing the efficiency of the model by training a large number of images, working on live stream image capturing and recognition, and training the model a higher number of steps for better results. This system's voice synthesis provides convenience features for the visually impaired. Tensorflow lite module was used to create a mobile compatible object recognition model for easy use by visually impaired users. The Android application can be further improved on its stability and functionality.

7. Future scope

For security reasons, wired serial communications were used instead of a wireless server. If the information is linked to a server, it could be leaked onto the Internet. Since the information in question contains a lot of privacy and camera-based observations, such leaks could create critical security issues for users. However, a wired connection can secure the information by keeping it offline [13]. Continuous research is expected to solve server security problems, eliminate blind spots in observations by connecting Internet of Things (IOT) cameras to a secure network, and increase precision in object recognition [18]. This study can be used widely to provide the blind with privacy and convenience in everyday life. With the addition of a face recognition feature, the application can be trained to store information on people closely associated with the person, which would help them to distinguish between peers and outsiders.

References

[1] D. Choi, and M. Kim, "Trends on Object Detection Techniques Based on Deep Learning," Electronics and Telecommunications Trends, Vol. **33**, No. 4, pp. 23-32, Aug. 2018, Article

[2] World Vision Disability Data, WHO, https://www.who.int/blindness/publications/globaldata/en/.

[3] Tom M. Mitchell "Machine Learning". McGraw Hill Education 2017.

[4] Raphael C. Gonzalez & Richard E. Woods "Digital Image Processing," Pearson 2018..

[5] Salman TOSUN, Eni KARARSLAN "RealTime Object Detection App for Visually Impaied:Third Eye". IEEE Conferences 2018.

[6] "Jayshre Punsare, Adit Gaikwad, Vaishnavi Ankarm, Priyanka Kare & Sikha Sharma "Real time Text Reader" International Journal of Computer Applications 182(34):42-45, December 2018

[7] Anirban sarkar, ayush goyal, david hicks and saikat hazra "Android Application Development: A Brief Overview of Android Platforms and Evolution of Security Systems" 2019 Third International conference on I-SMAC.

[8] Tensorflow,https://www.tensorflow.org/ https://www.tensorflow.org/lite

[9] Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi "(YOLO) You Only Look Once: Unified, Live Object Detection". Cornell University, Jun 2015

[10] W. Liu et al., "SSD: Single Shot MultiBox Detector," European Conference on Computer Vision, Vol**. 9905**, pp. 21-37, Sept. 2016, Article (CrossRef Link)

[11] Prince Kumar, Vaibhav Garg, Pavan Somvanshi and Pathanjali C "A Comparative Study of Object Detection Algorithms in A Scene" in IJERT published on 20-05-2019, ISSN (Online) 2278-0181.

[12] A.Quattoni, and A.Torralba. Recognizing Indoor Scenes. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2009.

[13] K. Jung, "Object recognition on mobile devices," in Consumer Electronics - Berlin (ICCE-Berlin), 2012 IEEE International Conference on, 2012, pp. 258–262.

[14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Object Detection Techniques: Detection and comparison," Proceedings of IEEE Computer Society's conference on computer vision and pattern recognition, 11 2013.

[15] R. Girshick, "Fast R-CNN," in 2015 IEEE International Conference on Computer Vision (ICCV), Dec 2015, pp. 1440–1448.

[16] "Looktel." [Online]. Available: www.looktel.com [Accessed: August 2013] Sarkale, Shah, Chaudhary, & Nagarhalli.

[17] "A novel machine learning approach for object detection and recognition.," 2018 Second International Conference on Creative Communication and Information Technology.

[18] K. Matusiak, P. Skulimowski and P. Strurniłło, "Object detection in a mobile phone app for blind users," Sixth International. Conference on Human System Interactions (HSI), Sopot.

[19] Live object detection using TensorFlow Detection Model. [Available: https://towarddatascience.com/live-object-detection-using-tensorflow-detection-model-e7fd20421d5d]

[20] Microsoft "COCO Dataset", http://cocodataset.org