OBJECT DETECTION, RECOGNITION AND DISTANCE TRACKING TO AID IN NAVIGATION FOR VISUALLY CHALLENGED

Divya Bharambe[1], Mandar Chaudhari[2], Soham Patil[3]

[1-3]*Students, Dept. of Computer Science, Smt. Indira Gandhi College of Engineering, Navi Mumbai, Maharashtra, India.*

Abstract: Object detection is one of the most relevant and difficult branches of computer vision, and it has been commonly used in people's lives, such as security surveillance, autopilot driving, and so on, to identify instances of semantic objects of a certain class. The performance of object detectors has improved due to the fast creation of deep learning networks for detection tasks. Of all of the human senses, vision is one of the most important, and it plays an important role in comprehending the natural world. Visually challenged people need help to walk around the streets or even in their homes. As a result, this paper is an effort to create an object recognition method for people with low vision. This necessitates the use of a camera, an application, and an audio system, among other things. We created and implemented an Android application that uses the phone's camera to identify objects in the environment of a visually impaired person. In addition, the application can warn the user of the object's location as well as the user's distance from it. Audio devices like earphones, birds or the phone speaker can be used in the application to alert the visually challenged user about the object's name & status. As a result, our aim is to present a visual replacement system that will assist visually challenged people in their daily lives by educating them of the different objects around them through the use of an object detection system.

Keywords: CNN, TensorFlow Lite, YOLO, Text-To-Speech

I. INTRODUCTION

Object detection has gotten a lot of press in recent years because of its wide variety of applications and recent technological advancements. This role is being researched extensively in academia as well as in real-world applications such as security monitoring, autonomous driving, transportation surveillance, drone scene analysis, and robotic vision.

Among the many factors and efforts that have contributed to the rapid evolution of object detection techniques, the growth of deep convolutional neural networks and GPU computing power should be credited. The deep learning model is now commonly used in the entire field of computer vision, for both general and domain-specific object detection. Millions of people around the world suffer from visual impairments, making it impossible for them to comprehend their surroundings. One of the most difficult tasks for visually challenged people is getting around. They find it difficult to navigate independently since they are unable to assess the location of objects and people around them. Visually challenged people need someone to accompany them outside to get around. One of the most popular aids for the visually impaired is the white can. While it aids navigation, it does not alert the user to the presence of numerous obstacles until they are very close to them. As a result of the flaws in these traditional solutions, a lot of research is being undertaken. As a result of the shortcomings of these traditional solutions, more study is being conducted in order to create better and more sophisticated aids to help the visually impaired.An Android application for object detection will be designed for this device, which will use the phone's camera to identify objects in front of the user. The application will use TensorFlow's object detection API to detect the objects and provide the user with an audio message that includes the object's name and position. The position and distance of the object from the user are included in the location. An audio system, such as headphones or the phone's speaker, would be used to deliver the audio message alerts to the visually impaired person. The device does not need an external camera since it can perform the tasks above using the phone's camera. This paper aims to show how object recognition, a computer vision technology, can assist visually impaired people in supporting independent travel by providing an overview of object detection applications for visually impaired people, as well as their modalities and functionalities.

II. OBJECT DETECTION

Object recognition is a computer vision technique for identifying and locating objects in images and videos. Bounding box is created around the Objects detected allowing us to predictit or see the move of the object in frame.

Since object detection and image recognition are often confused, it's important to understand the differences between the two before moving forward.

A picture is labeled using image recognition. The word "dog" is applied to a photograph of a dog. The label "dog" is still applied to a photograph of two dogs. Object detection, on the other hand, creates a box around a dog that is labeled "dog." The model forecasts the location of each item and the mark that should be added. Object identification, in this sense, offers more detail about an image than recognition.

III. TECHNOLOGIES

Computer vision(Open CV)

Computer vision is an interdisciplinary field of study that looks at how computers can be designed to view visual images and videos at a high level. From an engineering standpoint, it helps to simplify functions that the human visual system can do. Methods for collecting, encoding, analyzing, and interpreting digital images, as well as the retrieval of high-dimensional data from the physical world to obtain numerical or symbolic knowledge, such as both are examples of computer vision functions in the form of judgments. In this case, understanding refers to the conversion of visual representations (the retina's input) into interpretations of the environment that can be used to communicate with other thought processes and evoke suitable action. The disentangling of symbolic knowledge from image data using models built with the help of geometry, physics, statistics, and learning theory can be seen as image comprehension. Computer vision is a research discipline that studies the philosophy behind artificial systems that derive knowledge from images. Video loops, multiple camera images, or multidimensional imagery from a medical scanner are also examples of image data. Computer vision, as a scientific discipline, aims to apply its ideas and models to the creation of computer vision systems. Scene reconstruction, incident detecting, video tracking, object recognition, 3D pose prediction, learning, indexing, motion estimation, and image restoration are all subdomains of computer vision.

The classic problem in computer vision, image processing, and machine vision is deciding whether image data includes a particular entity, function, or operation.

Neural networks

A neural network is a network or circuit of neurons, or in today's terms, an artificial neural network made up of artificial neurons or nodes. Thus, a neural network can be either a biological neural network (made up of real biological neurons) or an artificial neural network (made up of artificial biological neurons) for solving artificial intelligence (AI) problems. The biological neuron's interactions are represented as weights. An excitatory relation has a positive weight, whereas inhibitory connections have a negative weight. A weight is applied to all inputs before they are summed. A linear combination is a name for this operation. Finally, the output's amplitude is regulated by an activation function. For instance, an appropriate performance range is typically between 0 and 1, but it may also be between 1 and 1. Artificial neural networks, unlike von Neumann model computations, do not distinguish memory and processing and instead run by signal flow through the network connections, similar to biological networks.

These artificial networks can be used for predictive modeling, adaptive control and fields where they can be trained & built via a dataset.

Yolo

You Only Look Once (YOLO) is an object detection algorithm. The object detection task entails locating and classifying specific objects on an image. Previous approaches, such as R-CNN and its variants, used a pipeline to execute this process in several stages. Since the person is unique, this can be sluggish to run and difficult to refine.

An S x S grid of cells is created from the input image. One grid cell is claimed to be “responsible” for predicting each object that appears within the picture. It is the cell in which the object's center is located.

B bounding boxes and C class probabilities are predicted for each grid cell. There are five sections of the bounding box prediction: (x, y, w, h, confidence). The (x, y) coordinates indicate the position of the box's center about to the grid cell (remember that, if the center of the box does not fall inside the grid cell, then this cell isn't liable for it). These coordinates have been standardized to be between a and b.

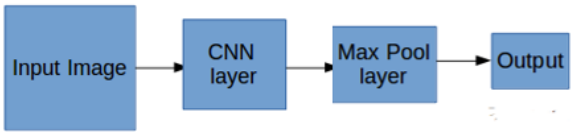

Fig No 1 YOLO Layer

How Does Yolo Work

The “You Only Look Once,” or YOLO, a family of models is a collection of end-to-end deep learning models for quick object detection created by Joseph Redmon et al. and first presented in the paper “You Only Look Once: Unified, Real-Time Object Detection” published in 2015.

A single deep convolutional neural network (originally a variant of GoogLeNet, later revised and named DarkNet based on VGG) splits the input into a grid of cells, each of which predicts a bounding box and object classification directly. After a post-processing stage, a large number of candidate bounding boxes are consolidated into a final prediction.

At the time of publication, there are three major versions of the approach: YOLOv1, YOLOv2, and YOLOv3. The first version proposed a general architecture, while the second version refined the concept and used predefined anchor boxes to improve the bounding box proposal, and the third version improved the model architecture and training mechanism even further.

TensorFlow

TensorFlow is an open-source machine learning tool that runs from start to finish. It has a large, scalable ecosystem of software, databases, and community resources that enable researchers to advance the state-of-the-art in machine learning and developers to quickly create and deploy ML applications. TensorFlow provides several layers of abstraction, allowing you to choose the one that best suits your needs. Create and train your models.

Text-To-Speech

Text to speech engine is a service that will help to deliver the texts/result in the form of speech to the user. The Language can be set to a US tone for easy understanding of the user.

Distance detection

Distance detection is very important in object detection for visually challenged people to guide them perfectly, rather to alert them quickly about the action of the object in the frame. There are many techniques for distance detection. Here in this project the bounding box is taken in consideration to calculate the distance i.e by calculating the perimeter of the bounding box(square/rectangle) distance/location is predicted.

IV. METHODOLOGY

The methodology we are using in our project is based on detecting and recognizing objects. We have used the YOLO model with TensorFlow Lite. We have converted the TensorFlow models and weights for YOLO to TensorFlow Lite format so that the YOLO can run on andriod. All the libraries and TensorFlow Lite API are initialized and the YOLO model is loaded. Then the objects are detected and a bounding box is formed for each object. The perimeter is calculated for each bounding box. For a particular object if the perimeter of the bounding box increases from the previous perimeter then the object is said to be moving closer. If the perimeter is decreasing from their previous perimeter then the object is said to be moving away. This will be given output as a voice command to the visually challenged. And If the perimeter is found to be steady then there is no moment of the object and output will say object ahead.

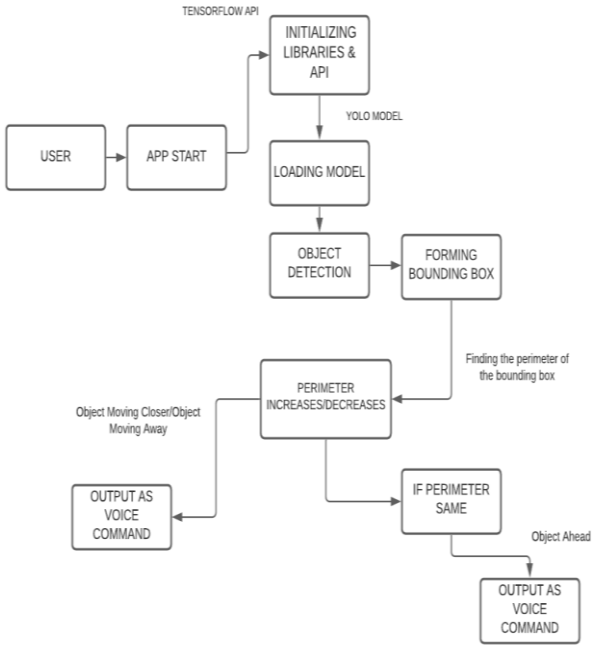

Fig No 2 Flow Chart for Object Detection

V. RESULT

Computer Vision

Some Problem with CV are described below:

• Object recognition (also called object classification) – One or more pre-specified or learned objects or object classes may be identified, normally along with their 2D image positions or 3D scene poses. Standalone applications such as Blippar, Google Goggles, and Like That demonstrate this feature.

• Identification – An object's instance is remembered. Identification of a particular person's face or fingerprint, identification of handwritten numbers, or identification of a specific vehicle are also examples of identification.

• Detection – The image data is scanned to see if it meets those criteria.

Convolutional neural networks are now the best architectures for such activities.

Yolo

- It is extremely fast compared to other real time detectors which came before it as it uses a Unified. Model where the detection is seen as a single regression problem and there is no complex pipeline, just a neural network run on the image.
- It makes less errors than Fast R-CNN as it can see the bigger context because YOLO, unlike Fast R-CNN, can globally reason the image when making predictions. YOLO sees the entire image and encodes some of the contextual information about all classes and their appearance.
- YOLO has learnt generalized representations of objects.YOLO successfully differentiates natural images against artwork.

Distance Calculation

Perimeter of the bounding box is calculated and is compared with the previous perimeter to check the movement of the object. If the perimeter of the current one is greater than the older one than the object is coming close to the user. And if the perimeter is getting smaller than the older one then it is going far. But while comparing the perimeter the class of objects should be the same.

VI. CONCLUSION

The project conducted more rigorous detection under the use of YOLO to provide suitable outcomes and help the project culminate into being assistance to impacted users, out of all the detection algorithms tested. The tensorFlowLite model ran seamlessly and successfully on the mobile device, offering a cost-effective and reliable medium for harnessing and instilling the benefits of machine learning.

VII. REFERENCES

[1] J. Redmon and A. Angelova, "Real-time grasp detection using convolutional neural networks," 2015 IEEE International Conference on Robotics and Automation (ICRA), 2015.

[2] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation," 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[3] R. Girshick, "Fast R-CNN," 2015 IEEE International Conference on Computer Vision (ICCV), 2015.

[4] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, 2017.

[5] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.

[6] H. Caesar, J. Uijlings, and V. Ferrari, "COCO-Stuff: Thing and Stuff Classes in Context," 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2018.

[7] S. Gupta, R. Girshick, P. Arbeláez, and J. Malik, "Learning Rich Features from RGB-D Images for Object Detection and Segmentation," Computer Vision – ECCV 2014 Lecture Notes in Computer Science, pp. 345–360, 2014.

[8] A. Karpathy and L. Fei-Fei, "Deep visual-semantic alignments for generating image descriptions," 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015.

[9] J. Xiao, K. Ramdath, M. Iosilevish, D. Sigh, and A. Takacs, "A low cost outdoor assistive navigation system for blind people," 2013 IEEE 8th Conference on Industrial Electronics and Applications (ICIEA), 2013.

[10] "TensorFlow Lite | TensorFlow," TensorFlow. [Online]. Available: https://www.tensorflow.org/lite. [Accessed: 24-Mar-2019].

[11] "An introduction to Text-To-Speech in Android," Android Developers Blog, 23-Sep-2009. [Online]. Available: https://android-developers.googleblog.com/2009/09/introduction-totext-to-speech-in.html. [Accessed: 24-Mar-2019].