Drishyam - Virtual Eye for Blind

Manasi Bhagwat[1], Ashwika Ghumate[2], Siddhi Kharawade[3], Prof. Mahendra Patil[4]

[1],[2],[3]Student, Department of Computer Engineering, Atharva College of Engineering, Mumbai, Maharashtra,India
[4]Professor, Department of Computer Engineering, Atharva College of Engineering, Mumbai, Maharashtra, India

---

Abstract - *In our society, we do find some people who are visually impaired. They may face difficulties while doing their daily activities and hence need someone to rely on for help. Till date many technologies for assistance of visually blind people have been developed. Among the various technologies, this project made use of Artificial Intelligence technology and proposed a helpful system for visually impaired people. The project performs classification on real-time images or videos using the webcam. With the use of such emerging advanced technology, blind people can interact with their surroundings. We are able to utilize the facilities of technology for the betterment of the society by developing helpful, affordable, transportable and economical systems.*

Key Words*: *Image Recognition, Visually imapired, Object Detection, YOLO, Tensorflow, Python.

1. INTRODUCTION

Artificial Intelligence systems have various applications and one such application is the Blind Assistance System which is built for helping the visually impaired. It not only assists them through voice commands but also does image recognition of the photographs captured or with the help of a camera it recognizes the object and describes it in audio. Such systems have made navigation easier for visually impaired people. Even if the person is blind, partially sighted or with progressive loss of vision, this system can help him to travel through familiar or unfamiliar places assisting him with voice commands.

The fundamental concept is to develop a valuable framework for the disabled individuals. The framework assists them through the YOLO (You Only Look Once) algorithm which detects the objects in the surrounding and lets the user know through speech recognition. Sometimes, it happens that the blind person fails to judge the distance of the obstacle ahead of him resulting in falling down and getting hurt. So, such a person needs an assistance or device that can inform him about the position of the objects situated ahead[1]. Thus, the object detection tool becomes the most used blind assistance tool for visually impaired people. Also, they need to be assisted through voice commands for proper navigation.

2. BASIC CONCEPT

The main idea is to develop a useful and efficient framework for the disabled individuals. The framework assists them by using the YOLO (You Only Look Once) algorithm which detects the nearby objects in the surrounding and informs the user via speech recognition.These features can be combined into a single device or system that is user-friendly and simple to operate. For proper navigation, they require assistance from voice instructions.

Applications:

1. Indoor navigation can be done efficiently.
2. Foundations and NGOs can make use of this system for the visually impaired.
3. Can be used by individuals when they need to travel to unknown places alone.

3. OVERVIEW OF ALGORITHM

3.1 YOLO ALGORITHM

YOLO algorithm is an algorithm that uses convolution neural networks to detect objects in real-time. This algorithm is based on a regression problem where it detects the object and provides the class probabilities of the detected objects. The prediction of objects in an image takes place in a single run of an algorithm. Among the various different variants of YOLO, YOLOv3 and tiny YOLO are the most commonly used[2]. The pre-trained Common Objects in Context(COCO) dataset is used in the implementation[3].

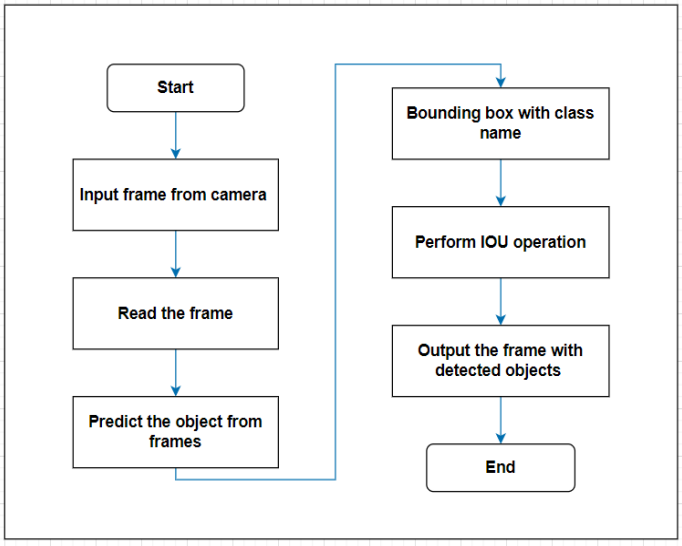

Fig 3.1: YOLO Algorithm

Advantages of YOLO algorithm:

- Speed: As this algorithm is fast, it is ideal for real-time processing.
- High accuracy: Predictions are made from one single network.
- Range: It is better than any other methods when generalizing from natural images to even artworks. Thus, it is more generalized.

4. PROPOSED SYSTEM

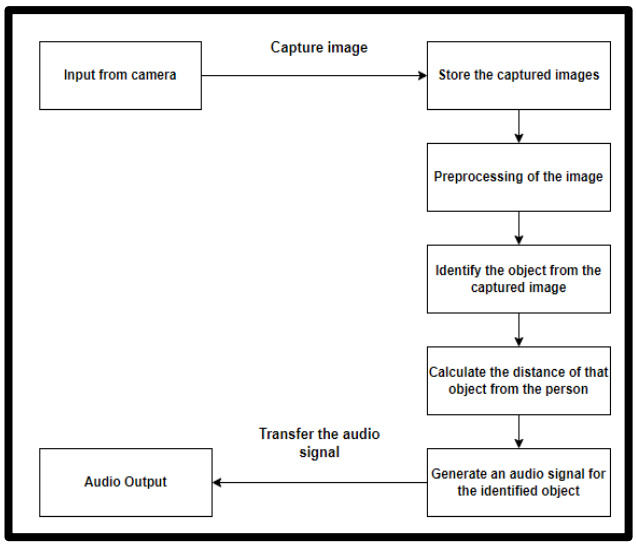

Fig 4.1: Proposed System

1. The aim is to build a system that can detect, classify and identify objects that are captured by the camera.
2. The camera module takes in the image as input from the ongoing video and stores the captured image for further preprocessing.
3. After that, feature extraction is done on the image and is further compared with the already trained dataset. It detects the object from the frame and recognizes the object to which category it belongs to.
4. After recognition, it then calculates the distance between the object and user. And also alerts whether it is at a safe distance or not.
5. It makes use of an audio synthesizer to generate speech commands thereby alerting the user about the object and its distance.

With such a framework being able to detect objects and being able to provide the correct distance, the user will feel more secure and will be more aware of his/her surroundings. The postponement between the detected object and the audio output to the user is made less and the audio output is clarified enough for the user to perceive[4]. Also the audio will be in the language understood by the user.

4.1 PSEUDO CODE:

1. Import required dependencies like opencv, numpy
2. Import training dataset (COCO)
3. Import darknet framework
4. Import suitable convolutional weights
5. Input data(image or video) is fetched from the webcam that captures the object.
6. Feature Extraction of the image is done in the next step, where the splitting of frames into 13x13 cells takes place along with Blob generation (stores images as bytes) and Bounding box construction.
7. The next step involves Object Classification where the confidence (X, Y, W, H) is updated and the relative position is found – Mid, Top, Bottom.
8. Once the features have been extracted and the objects are classified, the corresponding text descriptions are sent to the gTTS API where the text is analyzed and the letter to sound conversion takes place.
9. The output is predicted correctly from the results of the classification and the object is recognised with the audio.

5. RESULT ANALYSIS

Objects in the environment have been recognized and detected with more than 90% test accuracy. Below Figures represents the snapshot of the system. It displays the name of the object and its probability. whilst the application also tells the class label and how far the object from the person through the device's speakers.

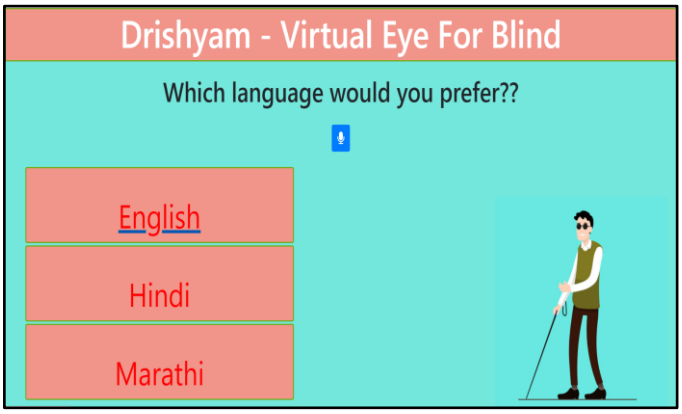

Fig 5.1: User Interface

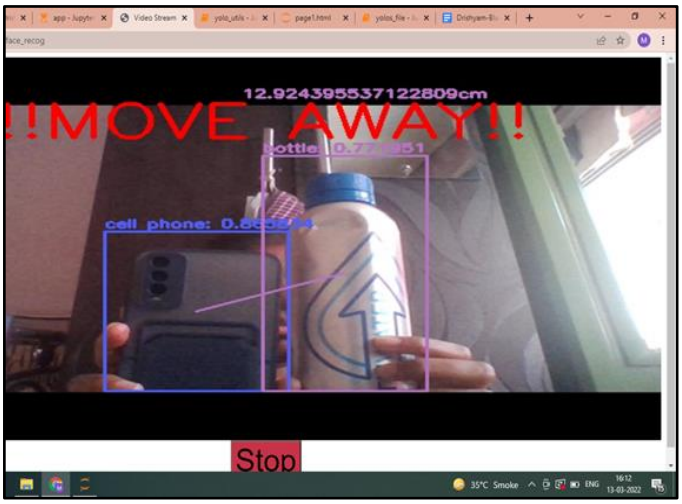

Fig 5.2: Result 1

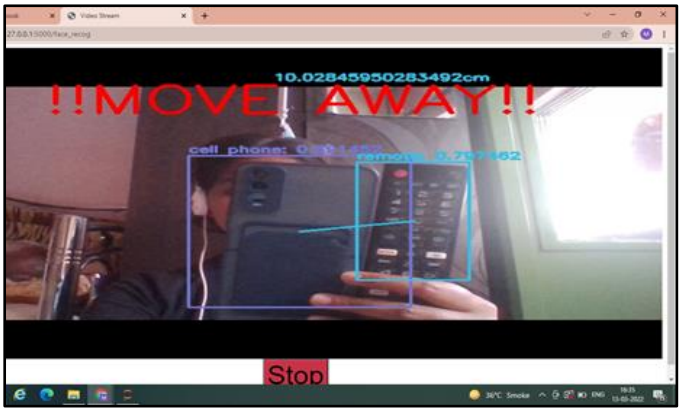

Fig 5.3: Result 2

Advantages

- The system makes use of text-to-speech technology which provides audio descriptions of the surroundings and helps users to travel with confidence.
- The system is mobile, robust, user-friendly and efficient.
- As we made use of the YOLOv3 algorithm, it is easier to detect even the small objects hence, making the system more effective.
- The user will be able to interact with the system in the language known to him and thus will feel safe and confident using it.
- The system is affordable and reliable.
- Real-time interface.

Limitations

- The system requires internet connectivity for text-to-speech conversion and hence, it cannot work well in offline mode.
- If the system compatibility is poor, there may be a delay in displaying the frames on the screen.

6. FUTURE SCOPE

In future we may also try to evaluate with other algorithms and with more parameters and images . We will try to implement this analysis on our own created dataset which is more effective and objective. Currently our system is made considering only three languages, but in future we will add more languages.

7. CONCLUSION

The project and the proposed system successfully reached a stage where a user-friendly, portable and accessible system using image processing technologies is in a position to assist visually imapired people in navigating their way and helping them in handling their daily activities with ease. Thus, this system gives them a sense of visualization because it helps them in visualizing their nearby environment based on their voice commands. This system consists of an easy architecture which makes it complexity-free and user-friendly.

ACKNOWLEDGEMENT

We sincerely thank our Atharva College of Engineering for providing us with a platform to prepare a project on the topic **"Drishyam - Virtual Eye For Blind"** . We want to

appreciate all the help and support we have received from the organization and would like to thank them for instilling in us the need for this research and to give us the opportunity and time to conduct and present research on the topic. We are also grateful for having **Prof. Mahendra Patil** as our guide and **Dr. Suvarna Pansambal**, Head of Computer Engineering Department for the motivation and constant support. Moreover, completion of the project would have been impossible without the cooperation, suggestions and help of our family and friends.

REFERENCES

[1] Leo Abraham, Nikita Sara Mathew, Liza George, Shebin Sam Sajan, "VISION- Wearable Speech Based Feedback System for the Visually Impaired using Computer Vision", IEEE Xplore Part Number: CFP20J32-ART; ISBN: 978-1-7281-5518-0

[2] "You Only Look Once: Unified, Real-Time Object Detection." J Redmon, S Divvala, R Girshick, A Farhadi, IEEE transactions, May 2016.

[3] COCO - Common Objects in Context (cocodataset.org)

[4] M. Mahendru and S. K. Dubey, "Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 734-740, doi: 10.1109/Confluence51648.2021.9377064.

[5] T. Guo, J. Dong, H. Li and Y. Gao, "Simple convolutional neural network on image classification," 2017 IEEE 2nd International Conference on Big Data Analysis (ICBDA), Beijing, 2017, pp. 721-724.

[6] Sumitra A. Jakhete, Avanti Dorle, Piyush Pimplikar, "Object Recognition App for Visually Impaired", 2019 IEEE Pune Section International Conference (PuneCon) MIT World Peace University, Pune, India. Dec 18-20, 2019

[7] Rahman, Ferdousi & Jahan, Israt & Farhin, Nafisa & Uddin, Jia. (2019). An assistive model for visually impaired people using YOLO and MTCNN. ICCSP '19: Proceedings of the 3rd International Conference on Cryptography, Security and Privacy. 225-230. 10.1145/3309074.3309114.

[8] Handalage, Upulie & Kuganandamurthy, Lakshini. (2021). Real-Time Object Detection Using YOLO: A Review. 10.13140/RG.2.2.24367.66723.

[9] Shifa Shaikh & Vrushali Karale & Gaurav Tawde,"Assistive Object Recognition System for Visually Impaired "International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181 IJERTV9IS090382 Vol. 9 Issue 09, September-2020

[10] A. Dionisi, E. Sardini and M. Serpelloni, "Wearable object detection system for the blind," 2012 IEEE International Instrumentation and Measurement Technology Conference Proceedings, 2012, pp. 1255-1258, doi: 10.1109/I2MTC.2012.6229180.

[11] Z. Shi, "Object Detection Models and Research Directions," 2021 IEEE International Conference on Consumer Electronics and Computer Engineering (ICCECE), 2021, pp.546-550, doi: 10.1109/ICCECE51280.2021.9342049.

[12] D. P. Khairnar, R. B. Karad, A. Kapse, G. Kale and P. Jadhav, "PARTHA: A Visually Impaired Assistance System," 2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA), 2020, pp. 32-37, doi: 10.1109/CSCITA47329.2020.9137791.

[13] S.Mahalakshmi, Veena N, Anisha Kumari ,"Visual Assistance for Blind using Image Processing", International Research Journal of Engineering and Technology (IRJET) , Volume: 07 Issue: 07 | July 2020

[14] H. Jeong, K. Park and YHa, "Image Preprocessing for Efficient Training of YOLO Deep Learning Networks," 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, 2018, pp. 635-637.