iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.53378

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Object Detection for Blind People Using Yolov3

Prof. Pranjali Deshmukh[1], Ajinkya Khedkar[2], Shubham Kulkarni[3], Shriram Morkhandikar[4]

Abstract: ***Vision constitutes one of the most important senses that humans utilise when interacting with their surroundings. There are over 200 visually challenged persons in the globe, while being visual impaired makes many daily activities difficult. As a consequence, it is critical to blind people to grasp their surroundings and the items with which they interact. In this project, we designed a website that assists blind persons to recognise various things in their surroundings by utilising the YOLO Version 3 algorithm. This combines many technologies to develop a rich website that not just helps individuals with vision impairments recognise different objects in their surroundings in real time, but also directs them through an auditory output. The You Only Look Once (YOLO) algorithms is employed.***

In accordance with World Health Organisation (WHO) data, at least 285 million individuals are visually impaired or blind. Blind persons have to depend 1 on whit canes, assistance dogs, screen-reading programmes, magnifying glass, and eyeglasses for movement and identifying objects. As a consequence, in order to help blind individuals, the world of sight must be turned into an aural world capable of informing them about items. In this research, we offer an ongoing item detection system to assist those with visual impairments in their daily lives. This system 1 makes use of the You Only Look Once(YOLO) deep learning technique. We will utilise the a version of actual time detection of objects method developed on th COCO dataset to detect the object in front of the person.

Keywords: YOLOV3, Object localization, Object detection, Image classification, Text to speech

I. INTRODUCTION

Vision constitutes one among the most crucial senses on which every individual in this world depends in order to interact with numerous things and individuals in the actual world. Normal people look around and immediately know which objects are nearby, how far away they are, and how to interact with them. It is not difficult for folks who can see.to perform their daily tasks because they can see all the surrounding things, any other individuals they encounter, and any barriers in their way, making it easy to engage with them. At the exact same time, visually impaired people must work hard to interact with real life because of their regular chores. According to a World Health Organisation (WHO) survey, around 286 million individuals are visually impaired, with 39 million being blind and 246 million having low vision. With the expansion of the newborn population, eye disorders, accidents, ageing, and other factors, the number of visually impaired persons is expanding, and this figure grows by up to 2 million people worldwide each year. The visually impaired's skills to do daily duties are hindered or impacted. For a result, many visually impaired people would bring a sighted friend or family member to assist them in navigating unfamiliar situations. These social difficulties impede a blind individual's capacity to meet new people.

Previous research has proposed numerous techniques to help those with visual impairments (VIPs) live normally. These solutions have not been able to adequately meet the safety measures when VIPs walk alone, and the presented concepts are often complex and inefficient. We propose a system that utilises the latest developments in image processing and machine learning. The system employs a YOLO(You Only Look Once) deep learning method, with the gadget including a camera module and an audio connector. The picture of the thing in proximity to the person will be captured by the camera. Following this, data is processed using deep learning methods, and the result, which is the name for the thing, is transformed into audio for the user via the audio jack. A system is proposed to assist visually impaired persons with day-to-day tasks such as walking, working, and performing housework.

Blind people rely on others to guide them constantly throughout the day, even for seemingly simple tasks like crossing the street or catching a bus. The primary goal in creating this website was to help those who are blind. This website attempts to assist blind persons in becoming aware of nearby objects that may be simple everyday items or pose a barrier to their normal activities. The website is designed to recognise or detect certain objects, such as people, motorbikes, potted plants, cars, and other exterior objects, as well as certain objects inside a home, such as tables, chairs, beds, computers, and refrigerators.

The web application will use the camera of the computer on which this page is loaded to capture real-time images of the environment's items while continuing to take still pictures for the video that is currently playing. These frames will then be forwarded to the following module, where the YOLO algorithm will generate bounding boxes around the items in the frame and classify the objects into the specified categories.

©IJRASET: All Rights are Reserved

The recognised object in the frame with the highest confidence score among all other objects in the frame will ultimately be the one for which the web application generates an audio output.

II. LITERATURE SURVEY

1) Paper Name- A Survey on "Object Detection Algorithms for Visually Challenged People"

Authors- Aishwarya Kumkar, Yash jagtap, Aishwarya Kumkar, Aditya Pole,

Abstract- Real-time item recognition and dimensioning is a significant challenge in many industrial sectors nowadays. This is an important area of computer vision issues.

This work offers an improved method for quickly measuring objects from video streams and detecting them. We proposed an object measurement method for real-time video that makes use of the canny edge recognition, dilation, and erosion algorithms with the OpenCV libraries. The suggested approach comprises of four steps: finding an object to be measured using a clever edge detection algorithm,filling in gaps between edges using morphological operators such as erosion and dilation algorithms, locating and sorting contours, and measuring the dimensions of objects. In order to execute the suggested strategy, we created a

2) Paper Name- Object Detection using Machine Learning Technique"

Authors- Amin, P., Anushree, B.S., Shetty, B.B., Kavya, K. and Shetty

Abstract-This research suggests a technique for using a stereo camera system and structural light to automatically and accurately assess object size. Before calculating size, the method goes through four steps: preprocessing, object detection, key point extraction, and depth interpolation. The initial step of preprocessing is aligning the depth and RGB frames. The object is then discovered using a depth threshold and key points are extracted using our suggested key point extraction algorithm in conjunction with the ShiTomasi corner detector. A depth interpolation algorithm is created to address inaccurate depth at object edges. Last but not least, using the 3D coordinates of the important spots, Euclidean distance is used to derive the object dimensions. Three-object measuring experimental findings suggest that average accuracy

3) Paper Name- Deep Hash Assisted Network for Object Detection in Remote Sensing Images

Authors- Rafael MIN WANG, ZEPEI SUN, GUANGYING XU, HONGBIN MA

Abstract- Images from remote sensing (RSIs) frequently feature a lot of terrain and a very broad width. In this work, a Deep Hashes Assisted Network (DHAN) is created through the use of a hashing encoding in order to quickly find objects in huge RSIs. in a two-stage deep neural network of photos. In contrast to conventional detection networks, DHAN first identifies potential object regions before sending the learnt features to a different Region Proposal Network (RPN) to do detection. One advantage is that calculations on background data that is irrelevant to the objects can be avoided. However, DHAN's built-in hash encoding layer can speed up detection using binary hash characteristics. To differentiate relatively small object regions, a self attention layer is created

4) Paper Name- Visual-LiDAR Based 3D Object Detection and Tracking for Embeded Systems

Authors- MUHAMMAD SUALEH AND GON-WOO KIM

Abstract- The idea that level 5 vehicular autonomy is feasible has been bolstered in recent years by consistent news updates on autonomous vehicles and the claims of companies joining the field.

just a short distance away. However, the primary barrier to establishing full autonomy still comes down to how the environment is perceived, which has an impact on the decisions made by the autonomous system.

Redundancy in sensor modalities that can operate in a variety of environmental circumstances is necessary for an efficient perceptual system. This system must also be able to provide accurate information with a minimum amount of computational resources. In this study, the major sensors used to detect and track 3D objects in the environment around the cars are a

5) Paper Name-.The development of image-based distance measurement system."

Authors- Lawrence Y.Deng.

Abstract- The method quickly calculates the real horizontal or vertical distance using two parallel laser lights as a scale. Although the fundamental concept of using ambient light to reduce image noise worked flawlessly in many circumstances, there were significant pitfalls when there isn't much light that proved difficult when developing Night Sight. In low light, auto white balancing (AWB) fails.) Tone mapping for difficult-to-see settings.

.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

III. PROPOSED SYSTEM

The proposed system was put into practise using Darknet-53, which serves as the foundation for YOLOv3 to find and categorise the objects. Data collecting, data modelling, training and testing the model, and performance analysis are the steps involved in the suggested system.

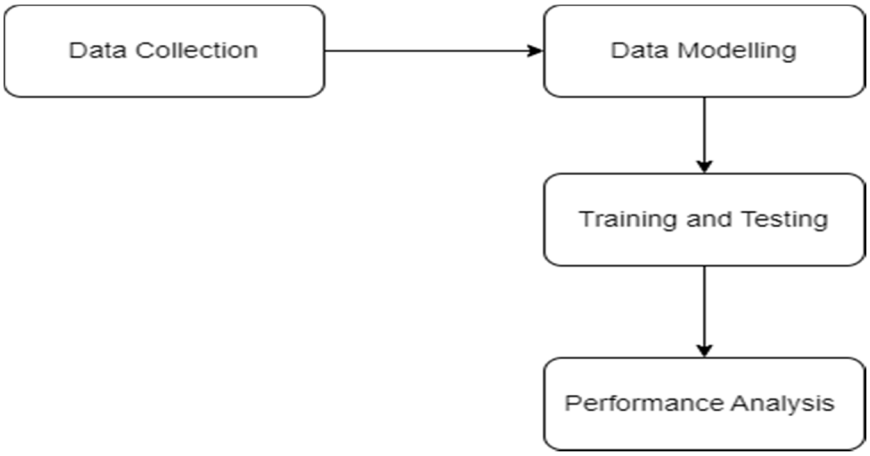


Fig. Proposed System

IV. SYSTEM ARCHITECTURE

The system architecture of object detection using YOLOv3 (You Only Look Once v3) consists of several components that work together to detect objects in an image. Here's an overview of the architecture:

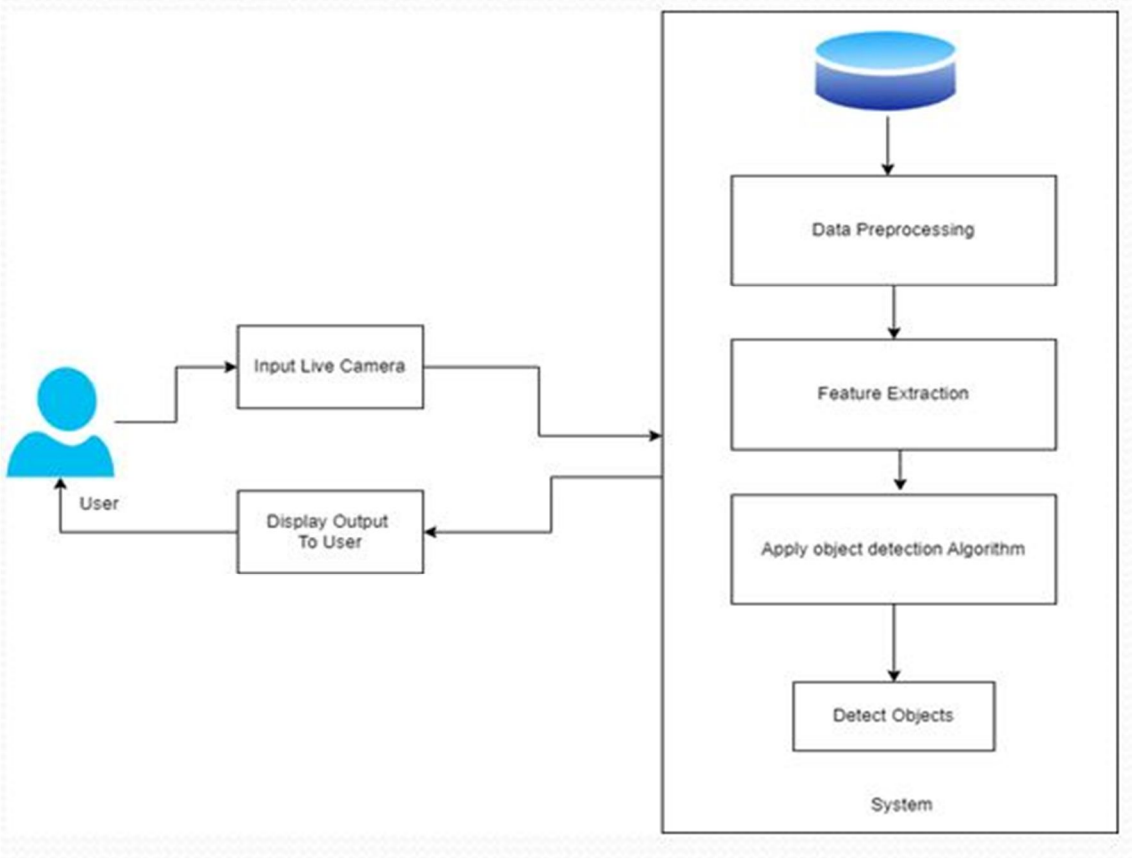


Fig. System Architecture

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

1) *Input Live Camera:* The system takes an input image in a specific format (typically RGB) on which object detection needs to be performed.
2) *Preprocessing:* The input image undergoes preprocessing steps to prepare it for the detection process. This may include resizing the image to a fixed size and normalizing pixel values.
3) *Feature Extraction:* The backbone network extracts high-level features from the input image, capturing both low-level and high-level visual information.
4) *Yolov3:* YOLOv3 has multiple detection layers at different scales to detect objects at various sizes. Each detection layer is responsible for detecting objects of a specific scale. These detection layers are usually implemented as convolutional layers followed by some additional layers for prediction.
5) *Output:* The system provides the final set of detected objects along with their class labels, bounding box coordinates, and confidence scores.

V. METHOD

A. YOLOV3

You Only Looked Once is also known as YOLO. This technique is used to instantly find and identify various things in an image. In YOLO, object detection is accomplished as a regression issue that provides the likelihood of classes for the observed photos. Convolutional neural networks (CNN) are used by the YOLO algorithm to recognise different objects in instantaneously. As the name implies, YOLO uses a neural network and only forward propagation process to recognise objects.This indicates that the ingle algorithm is used for identifying objects throughout the entire image. The CNN is used to simultaneously detect various boundaries and probabilities of classes. The YOLO algorithm comes in a variety of forms.Some popular ones include YOLO v1, v2, and v3.

How YOLO Algorithm Works:-

1) Yolo starts by requesting a picture.
2) Next, n X n grids are created from these input photos.
3) Each grid is used for image localization and categorization. Then, YOLO finds the various bounding boxes and associated class probabilities for each object in the image.

The model must get the labelled data in order to be trained. Assume that there are n classes in all that the objects in the image can be classified into, and that the image is divided into a grid of size n X n. The classes are, in order, chair, car, person, and motorcycle. A label y, an eight-dimensional vector, will now be generated for each grid cell:

Where Pc is the class probability, Y = pc, by, bx, bw, bh, c1, c2,..., cn.

A bounding box for an object is specified by the parameters bx, by, bw, and bh. The object's classes are c1, c2, c3,..., and cn.

For identifying objects, this technique use neural networks with convolution. YOLO is one of the fastest object detecting algorithms available.

YOLO v3 deeper feature extractor architecture named Darknet-53, which consists of 53, layers of convolution, each preceding a batch normalisation layer and Leaky ReLU activation. This Algorithm uses just one neural network to process the entire image. This network separates the picture in parts and forecasts boundaries and probability for every area. Those boundaries are assigned weights by the predicted probabilities.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Table 1. **Darknet-53.**

Opencv- OpenCV, which stands for "Open Source Computer Vision Library," is a free and open source software library for artificial intelligence and machine learning. OpenCV was created to provide a standard foundation for computer vision applications and to speed up the incorporation of artificial observation into commercial products.

gTTS (Google Text to Speech)-is a library for Python and command-line interface for interacting with Google's Translate's texts-to-speech API. This function writes spoken mp3 data to (stdout). It has flexible preprocessing and tokenization.

VI. IMPLEMENTATION

Dataset- Neural Networks are necessary for our web application to recognise objects. An picture dataset of objects is required to train the classifier. We used the COCO (Common Objects in Context) Dataset, which contains 81 distinct objects, in our application. The following are a few of the objects:

1) Person- person
2) Other- umbrella, handbag,tie,suitcase,frisbee. etc
3) Vehicle: bicycle, car, motorbike, aeroplane, bus, train, truck, boat
4) Animal: zebra, bear, elephant, cat, dog ,horse, sheep, cow etc
5) Indoor: chair, sofa, bed, dining table, toilet, tv monitor, laptop, cell phone

Data Preparation- Cocodataset.org is used to download the COCO dataset.

Data Labeling- The photos are labelled using the Label Img programme. For some photos, the annotations file is downloaded along with the dataset. Every single image in the annotation file has specific properties objectd_id and object_class,

Train-Test Split:- Once the dataset was collected and annotated, the dataset was shuffled. 80% of the data was used to train the model and the remaining 20% was dataset, which was not known to the model was used for testing purpose.

Object Detection:- The resulting bounding boxes are predicted using the degree of certainty, or confidence score. Knowing this score enables us to conclude that the bounding box contains some sort of object. Because each cell predicts an object's class for each node, a distribution probability is provided for each of the classes that are possible in the current model. The user can determine how likely it is miles that the bounding box contains some particular object by combining the self-assurance score with the chance that was just calculated.

VII. RESULT AND EVALUATION

On the grounds that there was quite a few studies and experiments in pc imaginative and prescient and mainly its area object Detection and popularity, there exists many unique algorithms and models that attempt to get almost correct results. There are following models to be had:

Haarcascade- A well-known method for object detection in computer vision is the Haar cascade algorithm. It works by using positive and negative examples of an object to train a classifier, where positive samples have the object and negative ones don't. The object in an image or video stream can then be found using this classifier.

For identifying objects with distinct, well-defined features, like faces, eyes, or particular items, the Haar cascade technique works well. On devices with limited resources, it can operate in real-time and is computationally effective. Using Haar Cascade Algorithm with YOLOv3 for Blind People: To assist blind people in object detection using YOLOv3, the Haar cascade algorithm can be used as a pre-processing step to detect specific objects that are critical for blind users, such as crosswalks, traffic lights, or pedestrian crossings

You Only Look Once (YOLO)- YOLO (You Only Look Once) is a deep learning-based object detection algorithm that takes a different approach compared to traditional methods like Haar cascades. YOLO divides an image into a grid and predicts bounding boxes and class probabilities for objects within each grid cell. It uses a single neural network to directly predict the obje ct classes and bounding boxes, making it efficient and capable of detecting multiple objects simultaneously. By combining the strengths of both algorithms, you can create a more comprehensive and efficient object detection system for blind people. The Haar cascade algorithm would handle critical objects with specific features, while YOLOv3 would provide a broader detection capability for a wide range of objects.

We accrued the COCO 2014 facts for diverse gadgets such as eighty lessons. at the beginning we determined to use educated YOLO model. We taken into consideration the YOLO implementation to get an in depth perception of the way item detection receives accomplished in diverse models.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

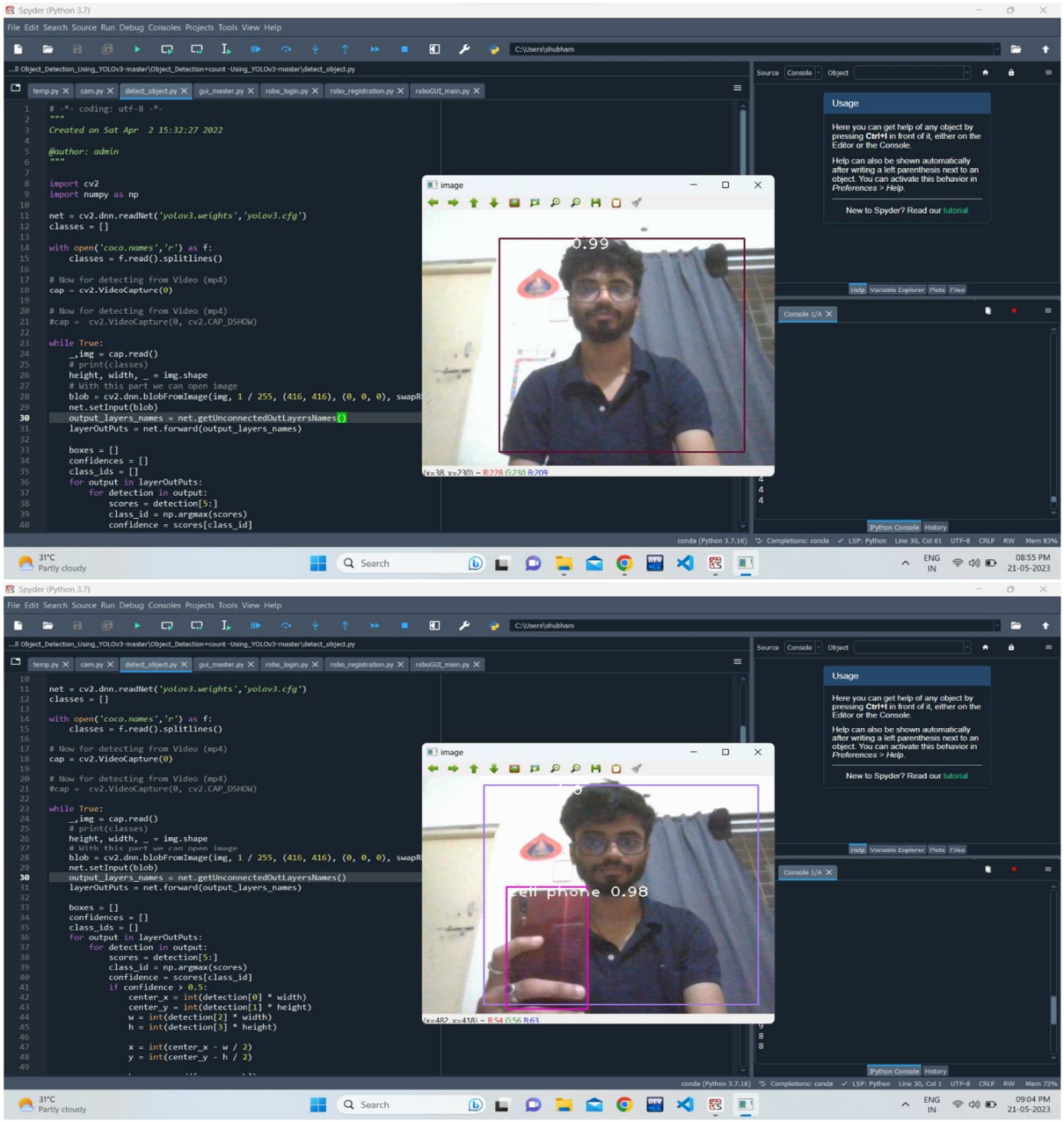


Fig.Detect Object Output

A. UML Diagram

Use case diagram

At its most basic level, a use case diagram is an illustration of a user's engagement with the system that demonstrates the connection between the user and the many use cases in which the user is engaged. A use case diagram can show the various kinds of Users of a system and the many use cases are frequently shown alongside other kinds of graphics. Either circles or ellipses are used to depict the use cases.

Use case diagrams were employed during the requirements analysis process to help understand the core characteristics and usage scenarios related to the established requirements. A use case illustration only displays a high-level view of the system as seen by an outsider (such as a client). It is viewed as a "black box," and the only thing that is known about it is what it does. A use case diagram includes actors, use cases, associations, and the system boundaries.

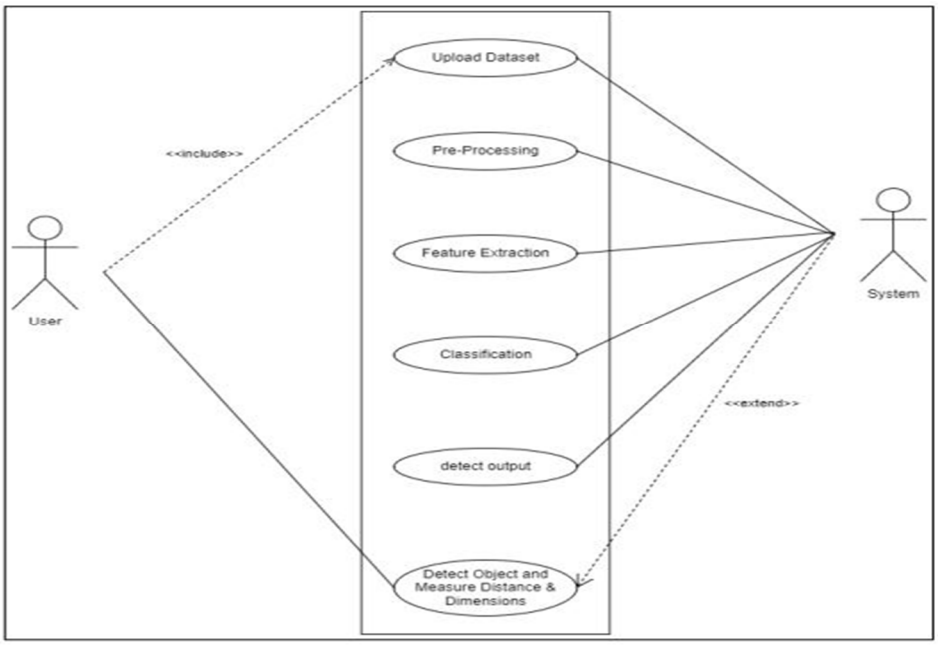


Fig.Usecase Diagram

B. Sequence Diagram

Object interactions are arranged in temporal sequence in a sequence diagram. It shows the objects involved in the scenario and the flow of messages that must be exchanged for the objects to work as intended. In the Logical View of the system being developed, sequence diagrams are often connected to use case realisations. Event diagrams and event scenarios are other names for sequence diagrams.

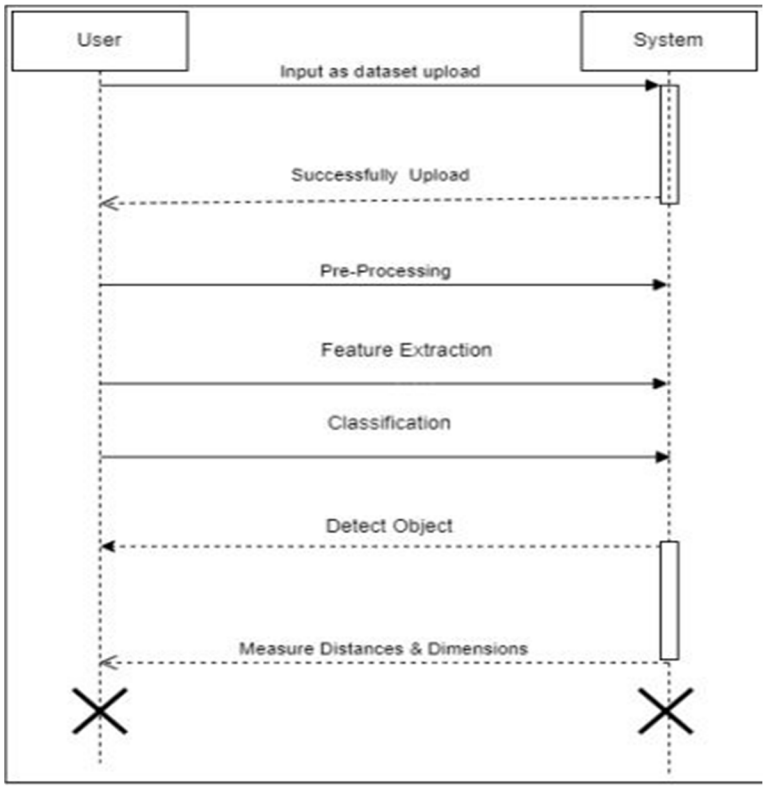


Fig. Sequnece Diagram

C. Advantages

For blind persons, object detection can have a number of benefits that improve their capacity to navigate and engage with their environment. Some of the main benefits are as follows:

Object detection can assist blind people in identifying and avoiding obstacles in their environment, such as walls, furniture, and other items. Objects can be recognised and their locations can be communicated to the user through audio signals or haptic feedback using computer vision algorithms, enabling them to travel securely.

Object detection technology can help the blind find their way by identifying and providing information about landmarks or particular objects of interest. For instance, it can proclaim the existence of a certain store, bus stop, or public amenity, assisting the user in navigating.

D. Limitations

Lack of Scene grasp: YOLO concentrates on item detection but may not offer a comprehensive grasp of the scene. Beyond simple object identification, efficient navigation for blind people requires a comprehension of context and the scene.

Limited Object Recognition: YOLO is capable of detecting a variety of items, but it may have trouble correctly identifying certain objects, especially in complicated or congested settings. This constraint may make it more difficult for blind people to interact and recognise objects in their environment. Difficulty identifying Small or Intricate items: YOLO may have trouble identifying small or intricate items, such as text on signs or facial expressions.

VIII. CONCLUSION

The potential for and results of using YOLOv3 for object detection to assist the blind are noteworthy. A well-liked and effective real-time object detection technique, YOLOv3 (You Only Look Once version 3), can swiftly find and identify objects in pictures or video streams. The purpose of this work is to identify things in photographs of traffic scenes. Along with the label identifying the class to which the object belongs, bounding boxes are formed around the discovered objects.

In conclusion, adopting YOLOv3 for object identification has proven to be a promising strategy for helping blind people. Environments that benefit individuals with visual impairments are made more inclusive and accessible thanks to its real-time processing, precision, and adaptability. The capabilities and usability of such systems will be further improved in the future thanks to ongoing developments in computer vision and machine learning techniques, user input, and iterative design.

IX. ACKNOWLEDGEMENTS

I would like to thank Prof. V.P. Patil for helping me out in selecting the topic and contents, giving valuable suggestions in preparation of Seminar report and pre-sentation 'OBJECT DETECTION FOR BLIND PEOPLE' I am grateful to Prof. Pranjali V. Deshmukh of Computer Engineering, for providing healthy environment and facilities in the department. She allowed us to raise our concern and worked to solve it by extending his co-operation time to time. Goal makes us to do work. Vision is more important than goal which makes us to do work in the best way to make work equally the best.

Thanks to Principal, Dr..R. V. Bhortake his support and vision. Consistent achievement requires boost on consistent interval basis. Management has given full support and boosted us to be consistent and achieve the target.

Thanks to management for their support. Thanks to all the colleagues for their extended support and valuable guidance. I would like to be grateful to all my friends for their consistent support, help and guidance.

REFERENCES

[1] Aishwarya Kumkar, Yash jagtap, Aishwarya Kumkar, Aditya Pole, "A Survey on "Object Detection Algorithms for Visually Challenged People"", in 2021.
[2] Amin, P., Anushree, B.S., Shetty, B.B., Kavya, K. and Shetty, L., 2019. Object Detection using Machine Learning Technique.
[3] Geethapriya. S, N. Duraimurugan, S.P. Chokkalingam, "Real-Time Object Detection with Yolo", in 2019.
[4] Avanti Dorle, Piyush Pimplikar, Pranit Bagmar, Atharva Rajkuvar, "Object Recognition App for Visually Impaired" in 2019.
[5] Sunit Vaidya, Niti shah, nishia shah,"Real-Time Object Detection for Visually Challenged People", in 2020.
[6] Jonathan Shen, Ruoming Pang, Ron J. Weiss, Mike Schuster, Navdeep Jaitly, Zongheng Yang, ARXIV, "Object Detection and Distance Estimation Tool for Blind People Using Convolutional Methods with Stereovision", in 2019.
[7] Juan du, "Understanding Object detetction based on CNN Family and YOLO", in 2018.
[8] Rui Li , Jun Yang, "Improved YOLOv2 Object Detection Model" in 2018.
[9] Melek, C.G., Sonmez, E.B. and Albayrak, S., 2019, July. Object detection in shelf images with YOLO. In IEEE EUROCON 2019-18th International Conference on Smart Technologies (pp. 1-5). IEEE.
[10] Zhou, X., Gong, W., Fu, W. and Du, F., 2017, May. Application of deep learning in object detection. In 2017 IEEE/ACIS 16th International Conference on Computer and Information Science (ICIS) (pp. 631-634). IEEE.

ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Author Information

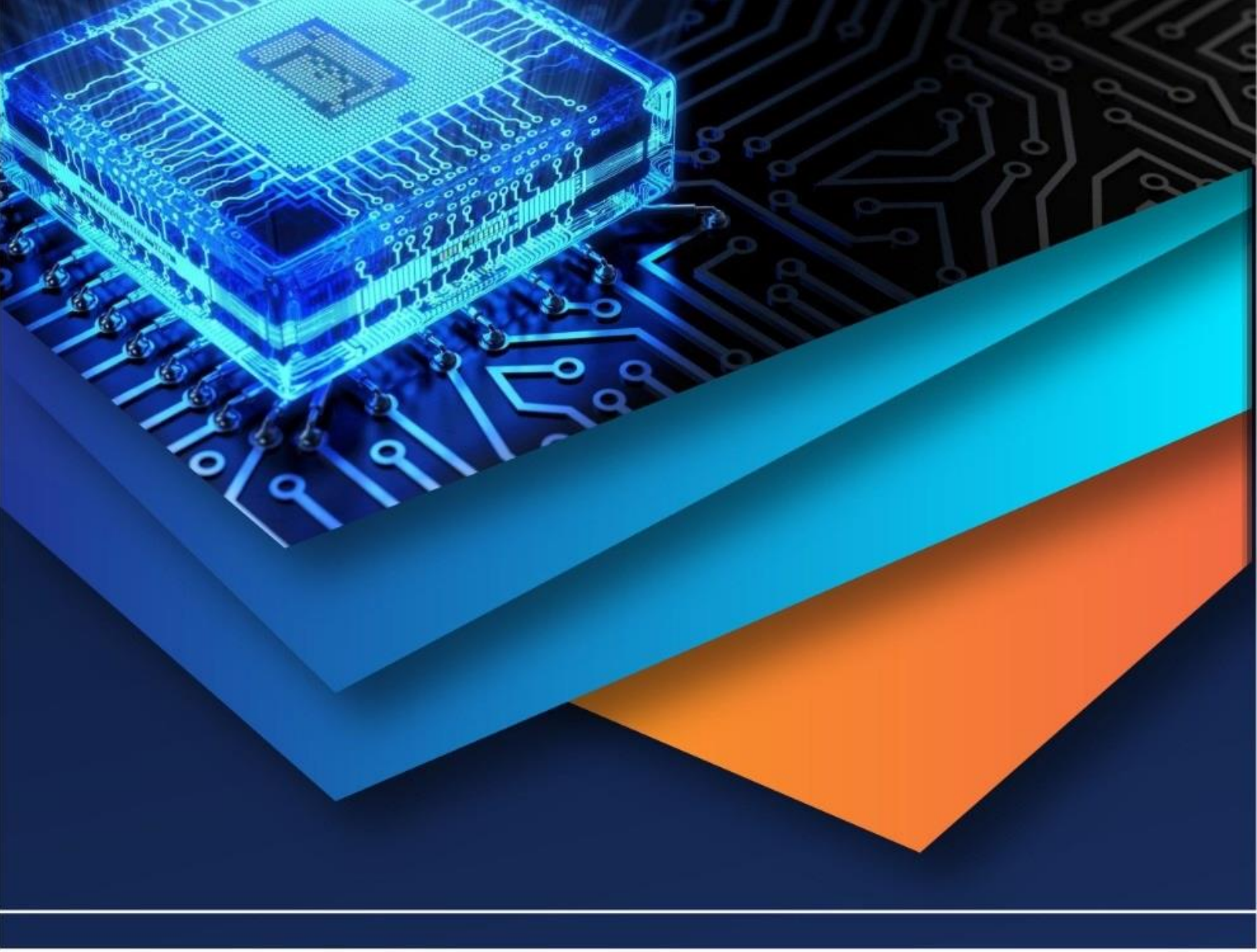
Prof.Pranjali Deshmukh
https://orcid.org/XXXX-XXXX-XXXX-XXXX
Marathwada Mitramandal's Institute of Technology
Sr.No. 35, Plot No. 5/6, Lohgaon, Pune – 411047
India
Contact e-mail: *pranjali.deshmukh@mmit.edu.in*

Shubham Kulkarni
https://orcid.org/XXXX-XXXX-XXXX-XXXX
Marathwada Mitramandal's Institute of Technology
Sr.No. 35, Plot No. 5/6, Lohgaon, Pune – 411047
India
Contact e-mail: shubham.kulkarni@mmit.edu.in

Ajinkya Khedkar
https://orcid.org/XXXX-XXXX-XXXX-XXXX
Marathwada Mitramandal's Institute of Technology
Sr.No. 35, Plot No. 5/6, Lohgaon, Pune – 411047
India
Contact e-mail: ajinkya.khedkar@mmit.edu.in

Shriram Morkhandikar
https://orcid.org/XXXX-XXXX-XXXX-XXXX
Marathwada Mitramandal's Institute of Technology
Sr.No. 35, Plot No. 5/6, Lohgaon, Pune – 411047
India
Contact e-mail: shriram.morkhandikar@mmit.edu.in

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)