IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

ARTIFICIAL VISION – YOLOv8 ALGORITHM FOR OBJECTS AND IMPEDIMENTS DETECTION

[1]Ashath Thauth S, [2] Karishma M

[1] Professor, [2] PG Scholar

[1]Department of Science, [2]Department of Computer Science and Engineering

[1]CS Academy, Coimbatore, TamilNadu, India, [2]Akshaya College of Engineering and Technology, Coimbatore, TamilNadu, India

Abstract: Blindness is a form of disability or inability to see. Lack of vision is only a disadvantage if the activity being pursued by the person who is blind is one that requires sight. Blind people may face troubles only due to inaccessible infrastructure and social challenges. Navigating around places and using pedestrian access signs are one of the biggest challenges for a blind person, especially the one with zero vision. For a physically challenged person, gaining independence may the most valuable asset and a person with visual impairment can lead an independent life in an environment which is designed with required adaptations for them. This paper presents a real-time object detection system using YOLO (You Only Look Once) based deep learning algorithm to help visually impaired people in their daily life. A Cobotic Spectacle is a cutting-edge product designed specifically for visually impaired individuals. The system seamlessly integrates face motion recognition, object identification and text detection functionalities using YOLO Algorithm. YOLO uses convolution neural networks (CNN) and single forward propagation through a neural network for object-detection in real–time. In this proposed system, YOLOv8, a real-time Object detection algorithm trained on the Common Object in Context (COCO) dataset is used to identify the object present before the person which produce high rate of accuracy in object identification. The system delivers the output through synthesized voice, enhancing visual perception. The face motion recognition module promotes improved social interaction, while object identification increases environmental awareness. This integrated solution combines edge processing and cloud-based analysis for portability, real-time processing, and computational efficiency. A Cobotic Spectacle empowers visually impaired individuals, providing a state-of-the-art smart spectacle system that enhances visual perception, promotes independence, and improves their overall quality of life.

***Index Terms* -** Visually impaired, Object detection, YOLOv8, Text to Speech, CNN

I. Introduction

Over the past decade, there has been a growth in the number of visually impaired people within our globe. The world health organization (WHO) has estimated and reported about 2.2 billion people worldwide to be visually impaired. Though visual impairment can affect people of all ages, the majority among them are over 50 years of age. When such visual impairment occurs in young children, it results in lower levels of educational achievement and stunts their further growth. Mean while in adults it often affects the quality of life lowering their productivity, decreases workforce participation and not to mention do results in high rates of depression [1]. Wearable devices are found to be the most useful solution among all assistive devices as they are hands free and require minimum use of hands. One such wearable, portable device for artificial vision is the "Cobotic glasses". These glasses allow visually impaired people to understand the text and identify objects through audio feedback. It can describe what they are unable to see as audio. These wearable devices can be enhanced with the latest techniques like Artificial Intelligence, Convolution Neural Network (CNN) using Machine Learning, Natural Language Processing, etc. YOLO algorithm integrated with wearable technology to detect objects and CNN may be used to train the model to detect the faces of the individual and surrounding objects. These glasses have an integrated wireless smart camera attached outside eyeglass frames which can read and verbalize text and further translate in desirable language. This verbalized text information can be converted to spoken words directly audible to the user's ear. One of the most essential growing up and useful feature set of Cobotic glasses is facial recognition using which the AI can be trained as the virtual roadside assistant and can be used to update the database of the device for storing new faces they meet, capture images, record videos, etc.

1.1 Basics of CNN

Among deep learning techniques, visual information like images, texts, videos and media can be evaluated using artificial neural networks such as a convolutional neural network(CNN). Convolutional neural networks can learn complicated objects and patterns from such visual information taking them as an input layer and generate an output layer through numerous hidden layers, and millions of parameters. CNN can sub sample the given input information before applying activation function and use convolution and pooling processes to partially connect all hidden layers with the completely connected layer at the end resulting in the output layer. Therefore CNN can be an efficient technique to recognise even partially occluded image information.

The recent trends in deep learning technology shows remarkable performance in various computer vision tasks such as image classification. A digital image contains a series of pixels arranged in a grid with values denoting the colour or brightness of each pixel. Therefore CNN has a big impact especially in the image classification, and it is an effective network model for learning filters that capture the shapes that repeatedly appear in images.

Due to this feature, CNNs can be used with recurrent neural networks not only to write captions for images and videos but also for character recognition and object detection. This feature can be enhanced and integrated with wearable devices and used for applications such as activity recognition or for video and image description for the visually impaired.

1.2 YOLO based Object Detection

YOLO is one of the most popular state-of-the-art, real-time object detection used for object detection. Due to its speed, accuracy, generalization and learning ability YOLO based object detection techniques are more researched in the recent decade. YOLO algorithm involves identification of objects in the form of processed image inputs. This technique involves a sequence of steps. The process starts with the image processing techniques such as noise removal, followed by feature extraction to locate lines, regions and possibly areas with certain textures which can be compared with the training data set for object detection.

In computer vision, object detection is a technique used for the identification and localization of objects within an input image and image localization is the process of locating one or multiple objects using bounding boxes, which correspond to rectangular shapes around the various objects within an input image. Various approaches such as fast R-CNN, Retina-Net and Single-shot Multibox Detector (SSD) have been used for object detection in past research works. Although these approaches have solved the challenges in object detection modelling, they are not able to detect objects in single algorithm run. YOLO algorithm has gained popularity among novel researchers because of its superior performance over the above mentioned object detection techniques. It involves instance segmentation used to find the exact boundaries of required objects with greater speed and accuracy.

In YOLO, object detection is performed as a regression problem which can provide various class probabilities of the detected images. This algorithm employs CNN to detect object in real-time which requires only a single forward propagation through a neural network to detect objects.

1.3 Deciphering Text in Visual Content

Reading text, understanding sign boards and pedestrian access materials with limited resources can be challenging for those who are blind or visually impaired in absence of an ideal circumstances. The goal of our work was to overcome such a challenge through artificial vision. This work is expected to make it possible for blind users to interact with printed text and hear real-time speech output in accordance with the digitally acquired image. Development of such a digital solution can enhance visually impaired people by providing an artificial vision with integrated optical character recognition (OCR) systems.

In order to design such a solution, two technologies are essential. They are ‘Text Information Extraction’ (TIE) and machine learning. The basic fundamentals of any assistive reading system are text information extraction. In general OCR systems are employed to identify words can accurately recognize letters, words, and sentences. OCR involves electronic conversion of captured images of typewritten or printed text into computer readable text that has a high rate of recognition. In OCR, TIE acts as a crucial component because it affects quality of output and how understandable the output can be.

The rest of the article is organised as follows: we have discussed all related works that led us to our proposed model in Section II. Section III presents the details about YOLOv8. Section IV presents the proposed system model and, the results obtained are discussed in section V. Section VI deals with evaluation metrics and their performance and finally conclusion and future work discussed in Section VII.

II. RELATED WORK

F. Rahman et al.[2] discussed about the earlier researches which suggests the various strategies used to overcome the issues of visually impaired people to have normal life. These strategies have not been able to fully address the safety measures when they walk on their own and the proposed ideas from the previous work are generally high in complexity and not cost-effective.

The study conducted by Anu Arora et al[3] presents the various hurdles faced by visually impaired individuals during ambulation, while using public transports & while working in organised sectors. This work shows the results of survey conducted through predefined questionnaire on different blind subjects working in organised sectors and concludes the various challenges faced by visually impaired people how improving technologies should help them.

The study by Y. R. S. Kumar et al. [4] focuses on the design and development of smart glasses based on Raspberry Pi, intended to enhance the well-being of visually impaired individuals. These glasses incorporate face recognition technology and a voice assistant to assist users in recognizing people around them, even when those individuals are not speaking. An assistive technology based on Computer Vision, Machine Learning and TensorFlow to support visually impaired people was developed by A. Ghosh et al in their article [5]. The proposed system explained in their paper allows the users to navigate independently using real-time object detection and identification. It also includes hardware implementation which is done to test the performance of the system, and the performance was tracked using a monitoring server. The system was developed on Raspberry pi 4 and a dedicated server with NVIDIA Titan X graphics where Google coral USB accelerator is used to boost processing power.

The significance of Facial Expression Recognition(FER) in diverse fields and acknowledging the challenges involved in accurately recognizing facial expressions are discussed by A.S.Vyas et al in their work [6]. In their article they present a survey of CNN-based face expression recognition techniques, including state-of-the-art methods proposed by different researchers. Their work also briefs the outline steps required for utilizing CNNs in FER and addresses key considerations and issues when selecting CNNs for solving FER. A Region Proposal Network (RPN) that shares full-image convolutional features with the detection network, that enables cost-free region proposals was introduced by S.Ren et al[7] and Jifeng Dai et al[8]. The RPN introduced in these work was trained end-to-end to generate high-quality region proposals, which were used by Fast R-CNN for detection. This technique merged RPN and Fast R-CNN into a single network by sharing their convolutional features using the attention mechanisms and identifies the area of interest and guides the system where to look. This method enables a unified, deep-learning-based object detection system to run at near real-time frame rates and improve the overall detection accuracy.

In the research work published by J. Redmon et al.[9], YOLO has been established as a fast system that can process 45 frames per second in real-time and outpace the existing identification systems in time efficiency. This paper also represents the system as a single convolutional network that can instantaneously forecast multiple bounding boxes and class probabilities for those boxes and YOLO as a system trained on full images, for optimizing the detection performance. P. Boobalan et al[10] presented a paper on deep learning-based object detection system that is specifically designed for the blind community. This system incorporates the YOLO algorithm for object recognition and uses text-to-speech (TTS) technology to provide a voice-guidance technique that conveys information about the objects around people with visual impairment. The main objective of this paper was to empower visually challenged individuals to independently identify objects in a particular space without relying on external assistance.

Rekha BS et al[11] and D. Wilson et al[12] addressed the various modifications that can be done to a YOLO network to improve efficiency of object detection and also discusses the pros and cons of various YOLO based detection systems. In this work, the YOLO architecture was surveyed, the YOLO network model for object detection, pedestrian detection, obstacle detection and solder joint detection was comparatively studied.

Pranjali Deshmukh et al[13] designed a project that assists blind persons to recognise various things in their surroundings by utilising the YOLO Version 3 algorithm. This project utilise real time detection of objects method developed on the COCO dataset to detect the object in front of the person. YOLOv3, due to its real time processing speed, accuracy and ability to adapt, this project had proven adopting YOLOv3 for object identification to be a promising strategy. In order to help blind people and create an environment that may benefit individuals with visual impairments YOLO based techniques may be made more accessible and inclusive. Anjaly Siby et al proposed the concept that involves extraction of text from the image using optical character recognition and converting the text to speech by pre-recorded syllabification in their article [14]. It was an approach that helps the visually impaired to read effectually. There had few of such research work on converting text to speech in the recent decades [15, 16]. The survey paper published by Nisha P et al [15] compares various such research works involving conversion of image to text and the text to an audio output. It also explains the purpose, implementation and test results of their experiment that takes in a captured image, text localization and text to audio conversion.

Based on our study from all related works, we propose an effective model to improve accuracy and speed in detecting objects in real-time using YOLOv8 algorithm. This paper discusses the working procedure of YOLO for object detection and further, the processing method of detected object and finally, the generation of audio from detected and processed images.

III. YOLOv8

YOLO is an algorithm that works on the basis of neural networks to provide real-time object detection. Due to its speed and accuracy YOLO is gaining more popularity in recent years. YOLO algorithm uses three integrated components to detect the objects from an input image. The first component of the algorithm is the Residual blocks – In this, the input image is viewed and divided into table grid of cells of equal dimension. Each grid cell is responsible for detecting objects that appear within the image. The second component is the Bounding Box – A bounding box is a boundary that identifies an object within an image input. YOLO uses regression of a single bounding box to predict the dimension of objects like height, width and center and as well as classify the detected objects. The final component is the Intersection over union (IOU) - It accumulates the grid cell results and provides an output box which is a unique bounding box that surrounds the detected objects perfectly and locates them from the input image.

Thus, by combining all the above mentioned techniques, an object is detected from the given input image. The working model of the algorithm is shown in fig.1, where, the system receives an input from the user. The system then processes the input image by applying the above mentioned techniques for effective object detection and identifies the objects by comparing the detected objects from the input image with COCO dataset trained images.

YOLOv8 is the latest version of YOLO model, which aims to deliver improved accuracy and efficiency over previous versions. It is the latest model that is designed to be faster, more accurate, and easy to use and the YOLO versions. This makes YOLOv8 an excellent choice for a wide range of object detection and tracking as well as for instance segmentation and image classification techniques.

YOLOv8 produces higher rate of accuracy measured when the trained data set COCO and Roboflow 100 are used. As compared to other versions, YOLOv8 is available with many additional features such as.,

a) A new anchor-free detection system
b) New convolution blocks
c) Mosaic augmentation

The YOLOv8 consists of several steps that works together to detect the objects and it is discussed in proposed model section. The flowchart of YOLOv8 is shown in fig.2

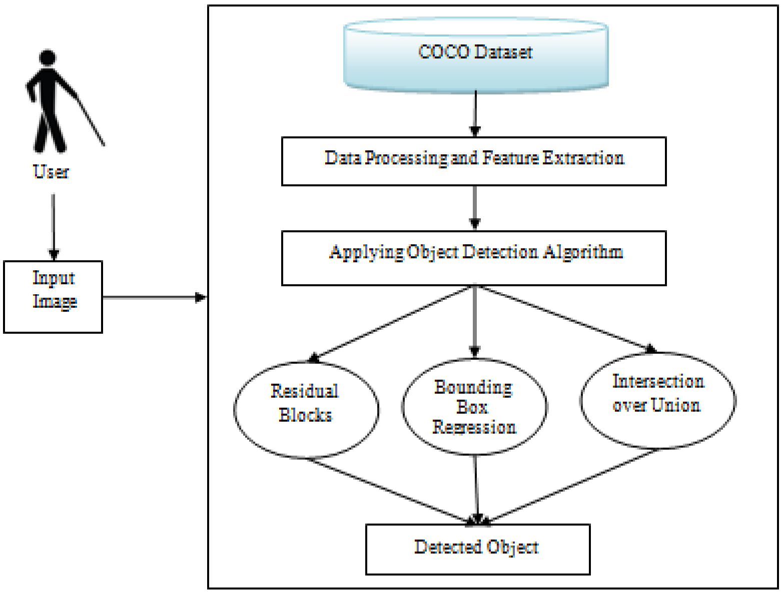

Fig.1 A model architecture of YOLO

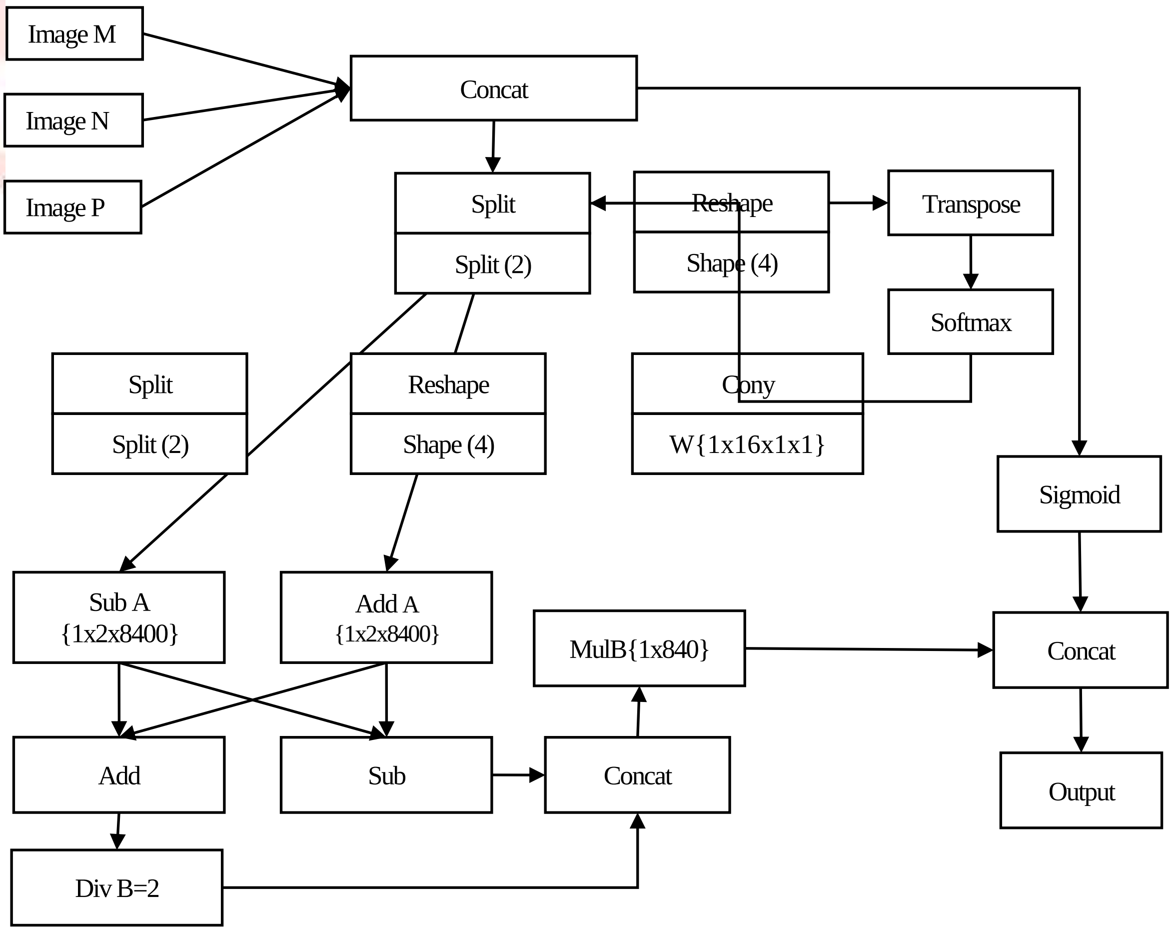

Fig.2 Flowchart of YOLOv8

IV. PROPOSED SYSTEM MODEL

The proposed system involves capturing real time images and processes them to detect various objects in them and bound the required region of interest and converts the acquired objects as an audio output. The system uses Python as a programming language to train YOLO on custom dataset in comparison with images from RF100 or COCO dataset. It uses the open source neural network, 'Darknet' as its framework responsible for object classification and bounding box prediction process. The proposed system is modelled by following steps as shown in Fig.3

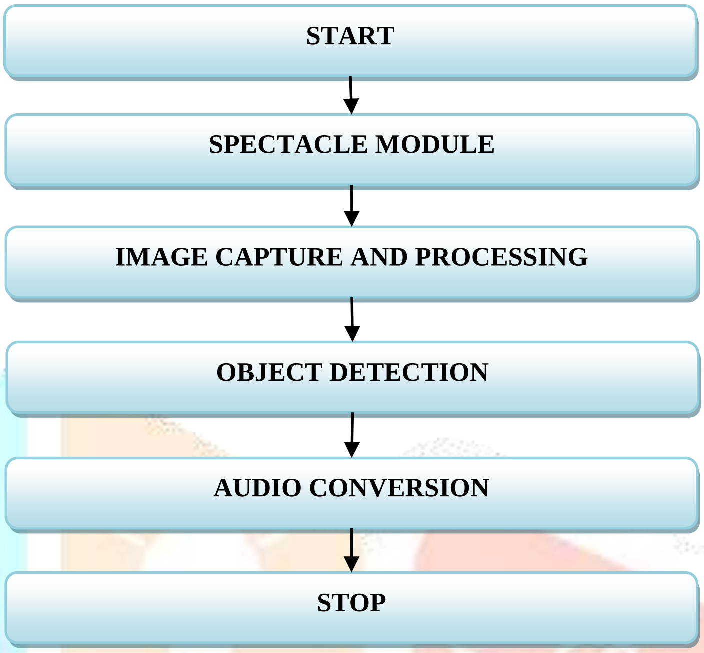

Fig.3 Workflow of Proposed Model

4.1 Spectacle Module

The spectacles used by the visually impaired people are designed in such a way, that it contains a small integrated camera embedded in it. This embedded camera can capture the surrounding environment and take the snip of the objects images on demand which may be transferred to the processing module. An earpiece attached to the modified spectacle device acts as a virtual assistant for the user through audio instructions. This modified spectacle can be the best independent device for the visually impaired user to make use of their own resource without the need of others to handle real time challenges of those people with impaired vision.

Fig.4 A Spectacle Model

4.2 Image Capture And Process

In this module, the image is captured from the integrated camera that contains objects to be detected. This module takes input image in the form of video frames. The images are pre-processed by resizing it to a fixed size and the resized frame values are normalized. The efficiency of the model is thus increased by normalization. The pre-processed image is passed through the convolution neural networks which can extract high level features from the image input. CNN utilize feature maps obtained from different layers of backbone network through information captured at various scales and resolutions.

4.3 Object Detection Module

In this module YOLOv8 algorithm is used for detecting the objects at a higher rate of accuracy. This algorithm has several detection layers. Each and every layer of the detection algorithm is responsible for detecting objects at definite scale. This algorithm uses CNN for implementing the convolution layers at different level and of different scales to detect objects within a range of sizes. This algorithm follows several steps to detect the objects from the input images captured by the camera and these processes are discussed further as follows.

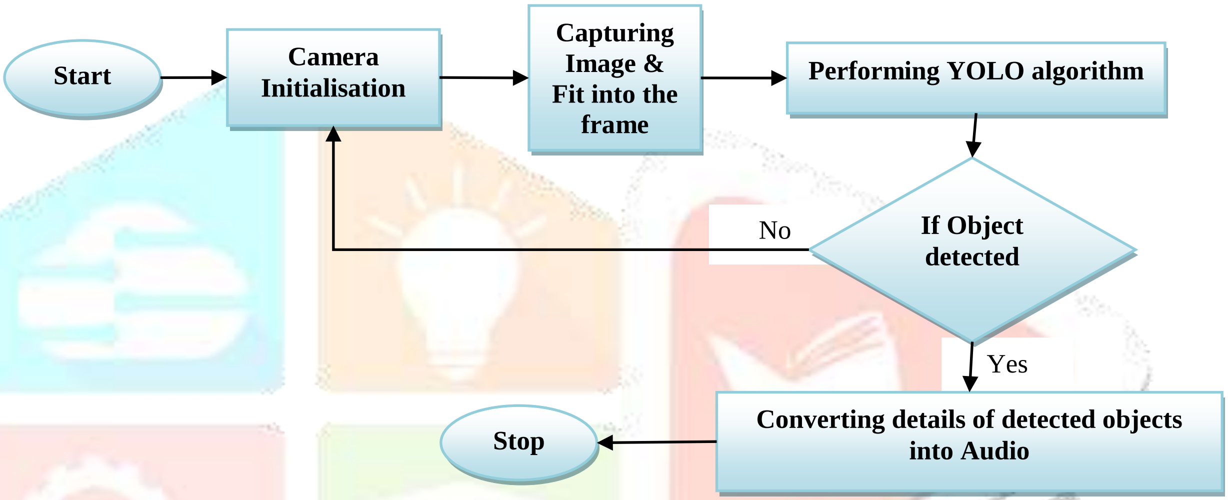

Fig.5 Proposed flow diagram

4.3.1 Anchor Free Detection

Anchor boxes are generally rectangular boxes to estimate the height and width of the objects to be found in the particular input image. YOLOv8 is an anchor-free model which predicts the center of an object directly instead of using the complicated counterpart anchor box model used in the earlier YOLO models. Thus, using YOLOv8 reduces the number of box predictions, which in turn helps to speed up the Non-Maximum Suppression (NMS), a convoluted post processing step that filter the detections after inference.

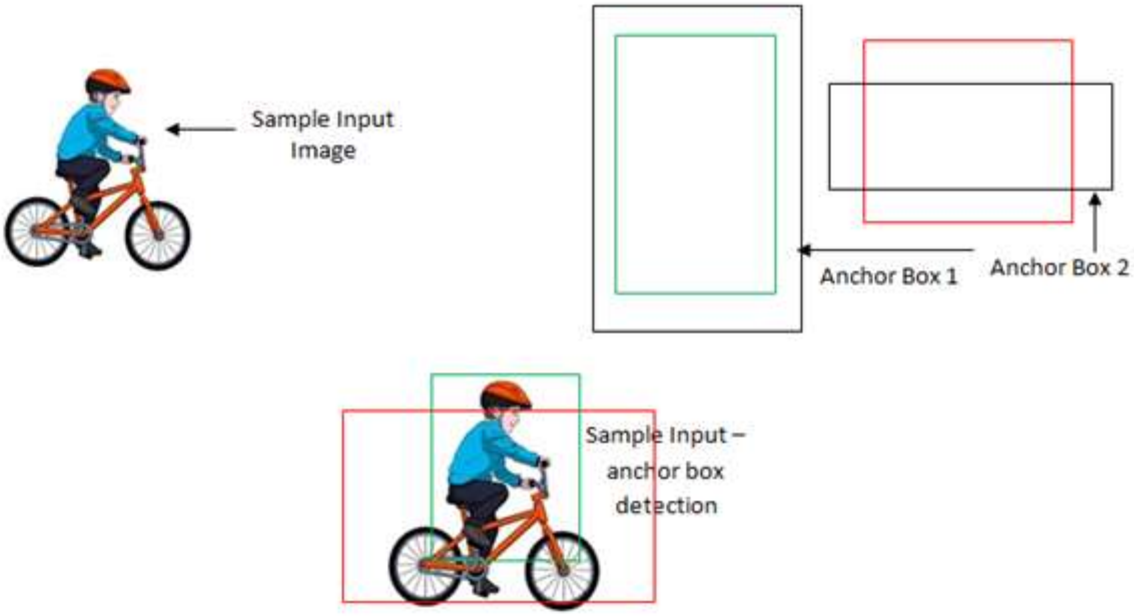

Fig.6 Anchor box detection

The above diagram shows a single grid cell with a boy in sitting in a bicycle. YOLO uses two anchor boxes: anchor box 1 is tall and thin, like a person, and anchor box 2 is shorter and wider like a bicycle.

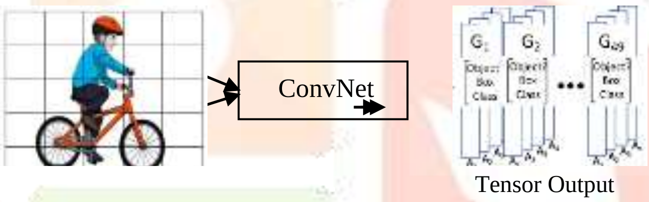

Fig.7 Output Tensor for YOLO ConvNet

4.3.2 New Convolutions

Based on defined the anchor boxes, YOLO extend the output to include a separate classification or bounding box vector for each anchor box. This allows YOLO's Convolution Network to look, in each grid cell, for objects matching the size and shape of each anchor box. By this, multiple objects in a grid cell is detected, as long as each object is associated with a different anchor box. An anchor boxes are used to improve the accuracy, by training parts of the ConvNet to become focused in detecting objects of a particular size and shape.

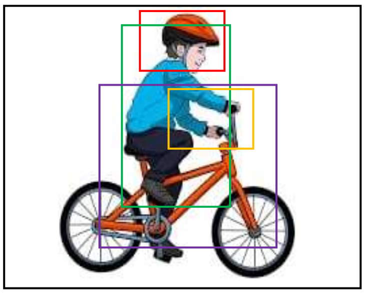

Fig.8 Predicting multiple bounding boxes

In sample training, the bounding box of the object is drawn and its classification is noted, and further it is assigned to the anchor box most similar to its bounding box. In fig.6, the boy is assigned to the first anchor box A1, and the bicycle is assigned to the second anchor box A2. Then, the target output shows the boy's classification as C1 and indicates as "person" and its bounding box is coordinated to the A1 target output vector, similarly the bicycle classification is done and the bounding box is coordinates to the target output vector A2. Thus, for each anchor box, classification and bounding box is made.

This classification and bounding box for each anchor box A1,A2,...An and for each grid cell G1,G2,...Gn is depicted in the figure 7. The output tensor produced by YOLO's ConvNet is used for final post-processing step, non-max suppression. This step helps to eliminate the redundant object detections and bounding boxes that may occur when two or more grid cells detect the same object. This can happen when an object overlaps multiple grid cells, in the figure 8 the person covers multiple grid cells, and it's not surprising that several of them detect a person exactly and estimate its bounding box.

4.3.3 Closing the Mosaic Augmentation

The Mosaic Augmentation involves mending multiple images together, forcing the model to learn objects in new locations, in partial occlusion, and against different surrounding pixels. The final output consists of a set of bounding boxes, each associated with a class label and a confidence score. These bounding boxes indicate the detected objects in the input image.

By following through the above steps, the YOLO algorithm classifies and locates multiple objects in one pass through an input image, making it fast enough to process even real-time video. It uses CNN for the processes and adds elements to the ConvNet's output that specify the identity and location of multiple objects of various shapes and sizes.

4.4 AUDIO CONVERSION

Optical Character Recognition (OCR) and text to speech (TTS) are two of the latest technologies that help in the creation of audio files from the processes text. Using these voice assistance technologies, the visually impaired people can identify the objects in their path and locate objects through the audio output effectively. OCR is a type of technology that uses visual images of characters and turns them into audio files, often in real-time. The pre-processed image when passed through the text detection function of OCR library; it extracts the bounding boxes or textual regions from the pre-processed input images. And, necessary post-processing steps are performed which helps to remove unwanted noises and also handles specific formatting requirements. Finally, the characters are recognised and stored for further use. Now, using TTS technology, the audio output is generated which states the name of the detected object and its location within the image or video frame. Thus by combining OCR and TTS technology, a voice comment is sent to the user.

This audio conversion process combined along with YOLO based object detection mechanism provides an advanced system to help visually challenged by providing real-time voice response of the obstacles and assist them navigate the places independently.

V. RESULTS

This proposed system may be great help for the visually impaired people to lead their life independently. This system uses YOLO algorithm, a convolution neural network based deep learning algorithm, for detecting the objects in front of the camera and generate an audio output after the processing the image input based on the detected object frame which uses OCR technology for visualising the characters of images and turns into audio files. Thus, the blind people may be able to find the way without any guidance. This system uses a faster, more precise and accurate technology based algorithm which helps to detect the objects accurately and also locate multiple objects within a single image frame. The fig.9 shows the visual representation of the detected object; the camera captures the image of the wall contains flower pots and figure 10 shows other similar images with objects detected in it.

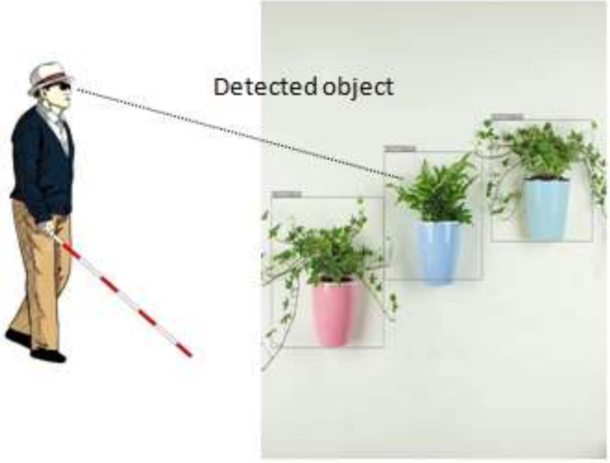

Fig.9 Resultant object detection

Fig.10 Other resultant images

VI. EVALUATION METRICES

Mean Average Precision (mAP) or Average Precision (AP) is on the most commonly used observation for evaluating and analyzing the performance of object detection and segmentation systems.

It works across several benchmarks like COCO, VOC and more. In order to analyze mAP of the system, calculation of parameters like confusion matrix, precision, recall and Intersection over Union(IoU) are estimated.

The confusion matrix is constructed based on four attributes. They are True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN). T**he prediction is a True Positives (TP) when the system predicts an object** and matches correctly. It is taken as **True Negatives (TN), if t**he model avoids an object correctly. **False Positives (FP) are attributes for t**he incorrect prediction of the model, while, the **False Negatives (FN) attributes to t**he model incorrect object avoidance.

Precision may be defined as the ratio of correctly predicted positive samples (true positives) to all samples predicted as positive (true positives + false positives).

$$\text{Precision} = \frac{\text{True Positive (TP)}}{\text{True Positives (TP)} + \text{False Positives (FP)}}$$

Recall, also known as sensitivity, is the ratio of correctly predicted positive samples (true positives) to that of all samples.

$$\text{Recall} = \frac{\text{True Positive (TP)}}{\text{Total samples}}$$

IoU, Intersection over Union helps to find the overlapping arear between two boundaries.

$$\text{IoU} = \frac{\text{Area of overlap}}{\text{Area of union}}$$

The mean average precision is average precision measured across different IoU thresholds to evaluate the performance of the model. mAP is determined as the weighted mean of precision values measured with different IoU thresholds. The introduced weights are the incremented in estimation of recall values for the successive threshold.

$$mAP = \frac{1}{N}\sum_{i=1}^{N} AP_i$$

Calculation of AP involves the following steps followed in sequence.

1. Generate prediction scores using the model.
2. Converting the scores to labels
3. Calculating TP,TN,FP,FN
4. Calculating precision and recall
5. Calculating the area of precision-recall curve
6. Measure AP

A graph plotted with the estimated precision values against the recall values as a function to evaluate the confidence score threshold is called a Precision-Recall curve. The precision-recall curve combines the tradeoff of both high precision-low recall model and high recall-low precision model and maximizes the effect giving us a better idea of the overall accuracy of the model.

mAP is considered a suitable metric for evaluating the performance of object detection applications as it can incorporate the exchange between precision and recall values considering both false positives (FP) and false negatives (FN).

The confidence score is the combination of precision and recall into a single metric that helps in comparing any two object detection models. It estimates the balance between precision and recall. Confidence scores are generally directly proportional to the precision and recall values.

$$\text{Confidence Score} = \frac{2 \text{ x } (\text{Precision} \text{ x } \text{Recall})}{\text{Precision} + \text{Recall}}$$

The table and comparison chart shows the performance of our proposed model in comparison with other classification models like CNN, Multi-layer Perceptron (MLP), k-Nearest Neighbours (k-NN), and Support Vector Machines (SVM) for face emotion recognition.

Table 1 Results Comparision

Model	**Accuracy**	**Precision**	**Recall**	**Score**
Proposed	0.88	0.87	0.90	0.88
CNN	0.85	0.84	0.87	0.85
MLP	0.79	0.77	0.81	0.79
k-NN	0.81	0.79	0.83	0.81
SVM	0.83	0.81	0.85	0.83

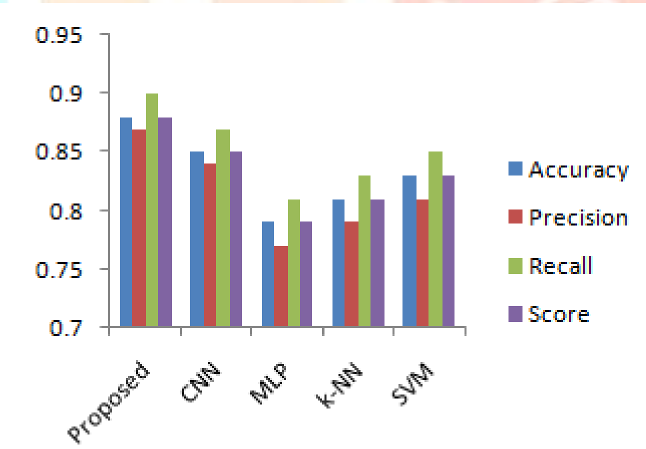

Fig 11. Result Comparision

Table 2 Evaluation Results

Metric	**Value**
mAP(IoU=0.50)	0.8 2
mAP(IoU=0.75)	0.7 2
Precision	0.8 7
Recall	0.9 0
Score	0.8 8

In our model, the estimated mean Average Precision (mAP) using an IoU threshold of 0.50, the model achieved a score of 0.82. Overall, based on these evaluation results, our proposed model shows better performance as compared to other models. While the precision and recall scores are relatively high, suggesting good accuracy and coverage, the mAP drops as the IoU threshold increases, indicating challenges in producing precise bounding boxes. These metrics provide insights into the strengths and weaknesses of the model's object detection capabilities.

VII. CONCLUSION AND FUTURE WORK

Our proposed system may acts as an artificial vision to the visually impaired people. The voice output enhances the user experience by providing a seamless and intuitive interaction, making it a valuable device for the needy people. Our innovative system offers real-time guidance, allowing users to navigate the roads confidently by receiving timely alerts about potential dangers and even, it helps them to pay a visit to crowded malls, markets, educational societies etc,. With these technical enhancements, we enact to enhance the safety and independence of visually impaired individuals, ensuring a more inclusive and secure environment for all. In future, the system can be aimed to enhance its process in zebra-crossing and traffic-signal system for blind people, helps them detect the lights of traffic signals effortlessly without any facilitation.

REFERENCES

[1] "WHO|World Health Organization", who.int. https://www.who.int/health-topics/blindness-and-vision-loss (accessed July, 2023)

[2] Rahman,F. Ritun,I.J and Farhin,N. 2018. "Assisting the visually impaired people using image processing", B.E., dissertation, Department of Computer Science and Engineering, BRAC University, Dhaka, Bangladesh.

[3] Anu Arora and Anjali Shetty, 2012. "Common Problems Faced By Visually Impaired People", International Journal of Science and Research (IJSR), Vol.3, No. 10, pp. 2002-2005.

[4] Kumar,Y.R.S. Nivethetha,T. Priyadharshini,P. and Jayachandiran,U. 2022. "Smart Glasses for Visually Impaired People with Facial recognition," 2022 International Conference on Communication, Computing and Internet of Things (IC3IoT), Chennai, India, pp.1-4, doi:10.1109/IC3IOT53935.2022.9768012.

[5] Ghosh,A. AlMahmud,S.A. Udayand,T.I.R. Farid,D.M. 2020. "Assistive Technology for Visually Impaired using Tensor Flow Object Detection in Raspberry Pi and Coral USB Accelerator," 2020 IEEE Region 10 Symposium (TENSYMP), Dhaka, Bangladesh, pp.186-189, doi:10.1109/TENSYMP50017.2020.9230630

[6] Vyas,A.S. Prajapati,H.B and Dabhi,V.K. 2019. "Survey on Face Expression Recognition using CNN," 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), Coimbatore, India, pp.102-106, doi:10.1109/ICACCS.2019.8728330.

[7] Shaoqing Ren, Kaiming He, Ross Girshick and Jian Sun, 2015. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks", Published in Advances in Neural Information Processing Systems 28.

[8] Jifeng Dai, Yi Li, Kaiming He and Jian Sun. 2016. "R-FCN: Object Detection via Region-based Fully Convolutional Networks", in: 30th Conference on Neural Information Processing Systems, Barcelona, Spain.

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

[10] Dr. Boobalan,P. Bhuvanikha,S. Sivapriya,M. and Sivakumar,R. 2023. "Object Detection with Voice Guidance to Assist Visually Impaired Using Yolov7", International Journal for Research in Applied Science and Engineering Technology (IJRASET), ISSN: 2321-9653; DOI: https://doi.org/10.22214/ijraset.2023.50182

[11] Rekha,B.S. Athiya Marium. Dr. Srinivasan,G.M. and Supreetha,A.S. 2020. “Literature Survey on Object Detection using YOLO”; International Research Journal of Engineering and Technology (IRJET), Volume: 07 Issue: 06 | June 2020.

[12] Dawn, W. Mausankar,C. and Prathibha,P.H. 2022. “Analytical study on object detection using yolo algorithm”, International Journal of Innovative Science and Research Technology(IJISRT), Volume 7, Issue 8.

[13] Prof. Pranjali Deshmukh, Ajinkya Khedkar, Shubham Kulkarni and Shriram Morkhandikar. 2023. "Object Detection for Blind People Using Yolov3", International Journal for Research in Applied Science and Engineering Technology (IJRASET), Volume 11, Issue V, Page No: 7172-7181, ISSN : 2321-9653, www.ijraset.com

[14] Anjaly Siby, Anisha P Emmanuel, Chikku Lawrance and Jain Mariya Jayan. 2020. “Text to Speech Conversion for Visually Impaired People”, International Journal of Innovative Science and Research Technology (IJISRT), Volume 5, Issue 4.

[15] Nisha,P. and Dr.Vijayakumar,J. 2021. “Survey paper: Text To Speech conversion for visually impaired peoples”, International Research Journal of Engineering and Technology (IRJET), Volume: 08 Issue: 06.

[16] Prof. Vidhyashree,C. Supriya,A.M. Supriya,H. Vedala Dinesh, and Kavya,R. 2022. “Machine Learning Based Text to Speech Converter for Visually Impaired”, International Journal for Research in Applied Science & Engineering Technology (IJRASET), Volume 10 Issue VII.