e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

REAL TIME OBJECT DETECTION WITH VOICE FEEDBACK

Mrs. S.L. Soniya[*1], Aashir Francis[*2], Allan Prabhu. M.S[*3], Ahilesh B.G[*4], Christen Cherub K.M[*5]

[*1]Assistant Professor, Department Of Computer Science And Engineering, St. Xavier's Catholic College Of Engineering, Chunkankadai, India.

[*2,3,4,5]B.E Student, Department Of Computer Science And Engineering, St. Xavier's Catholic College Of Engineering, Chunkankadai, India.

ABSTRACT

Visually impaired individuals face challenges in navigating their surroundings and identifying objects. This paper proposes an object detection system using You Only Look Once (YOLO) and the COCO dataset, combined with voice feedback, to assist blind and visually impaired people. The system utilizes a pre-trained YOLO model to detect objects in real-time video streams. Detected objects and their locations are then converted into natural language descriptions using a text-to-speech (TTS) engine, providing auditory feedback to the user. The proposed system is evaluated on the COCO dataset, demonstrating promising results in object detection accuracy. This work contributes to the development of assistive technologies that enhance the independence and spatial awareness of visually impaired individuals.

Keywords: Object Detection, YOLO, COCO Dataset, Visually Impaired, Assistive Technologies, Voice Feedback.

I. INTRODUCTION

Vision loss significantly impacts the daily lives of blind and visually impaired individuals. They face challenges in navigating unfamiliar environments, identifying objects, and performing everyday tasks independently. Assistive technologies play a crucial role in overcoming these challenges and promoting independence. This paper proposes an object detection system with voice feedback to assist visually impaired individuals. The system leverages the power of deep learning for object detection and audio feedback for conveying information. The core components are:

Object Detection: You Only Look Once (YOLO) is a deep learning architecture known for its real-time object detection capabilities. A pre-trained YOLO model is employed to detect objects in video frames captured by a camera.

COCO Dataset: The COCO dataset, a large-scale dataset with millions of labeled images, is used to train and evaluate the object detection model.

Voice Feedback: A text-to-speech (TTS) engine converts the detected object and its location into a natural language description, providing auditory feedback to the user through headphones or speakers.

II. RELATED WORK

Several research efforts have explored object detection and recognition for visually impaired individuals. [1, 2] utilize deep learning techniques for object recognition and provide audio feedback to assist navigation. These studies demonstrate the effectiveness of deep learning in object identification. However, some approaches focus on single-image object recognition, limiting their applicability in real-time scenarios.

This work builds upon existing research by employing YOLO for real-time object detection in video streams. The use of the COCO dataset provides a comprehensive set of object categories, enhancing the versatility of the system.

III. PROPOSED SYSTEM

The proposed system consists of three main modules: a video acquisition module, an object detection module, and a voice feedback module

Video Acquisition Module: A camera captures real-time video frames of the user's surroundings.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

Object Detection Module: The captured video frame is pre-processed for noise reduction and image normalization. The pre-processed frame is fed into the pre-trained YOLO model. The YOLO model predicts bounding boxes around detected objects and their corresponding class labels.

Voice Feedback Module: The bounding box information and class labels are extracted from the YOLO model's output. The information is processed to generate a natural language description of the detected object and its location (e.g., "person on your left").

A text-to-speech (TTS) engine converts the description into an audio signal.

The audio signal is played through headphones or speakers, providing voice feedback to the user.

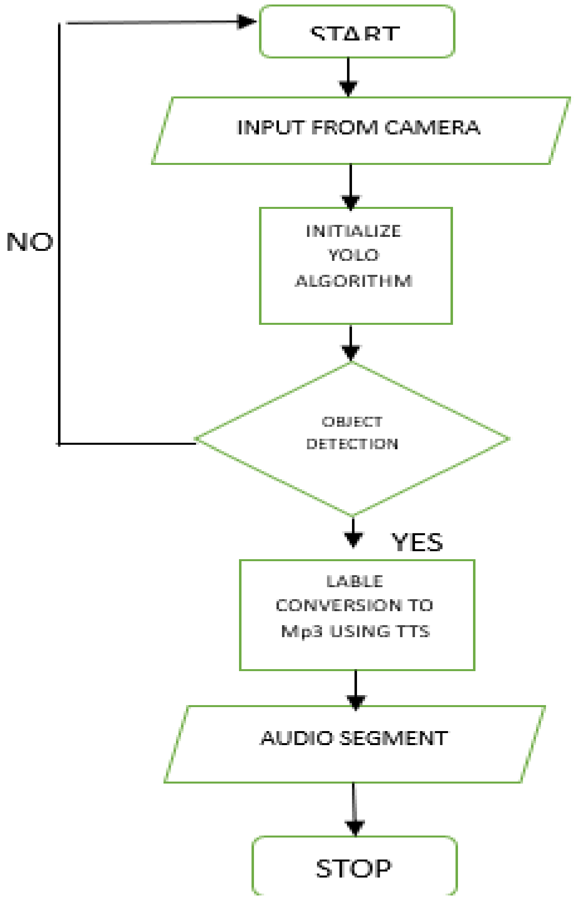

Fig 1: System Block Diagram

IV. FUNCTIONALITIES

The functionalities of the object detection with voice feedback system can be broken down into two main parts:

1. Object Detection:

Image/Video Input: The system takes an image or a video frame as input.

YOLO Model Processing:

- Runs the pre-trained YOLO model on the input image/frame.
- Identifies objects within the image/frame.
- Generates bounding boxes around the detected objects.
- Classifies the objects and assigns class labels (e.g., person, car, dog).

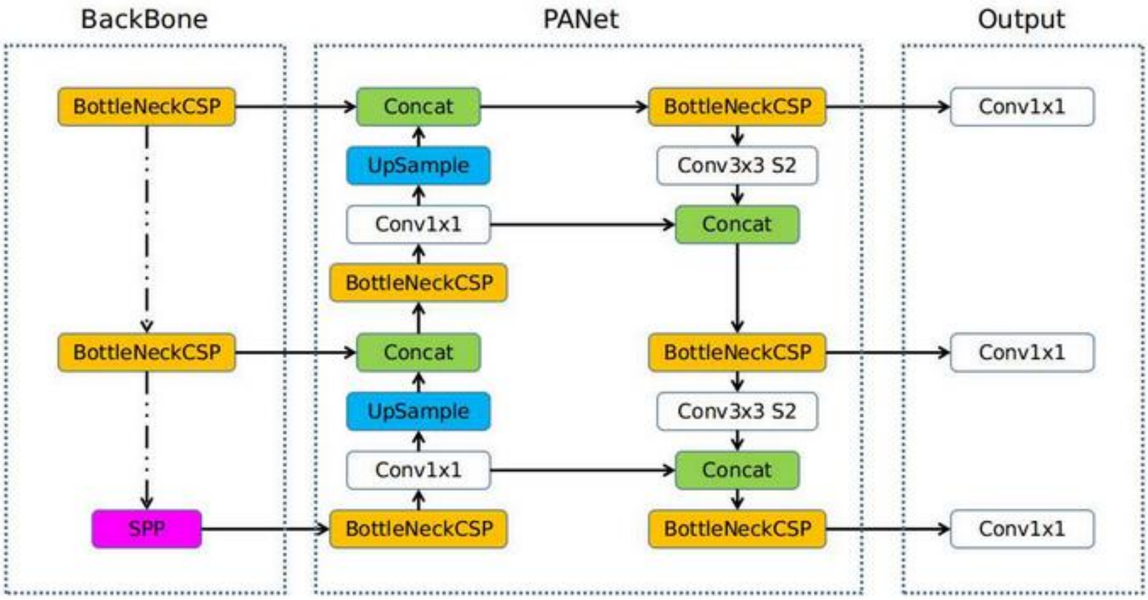

Fig 2: Architecture Diagram

Output: Provides a list of detected objects with their corresponding bounding box coordinates and class labels.

2. Voice Feedback Generation:

Data Processing:

Extracts relevant information from the object detection output, such as class labels and bounding box coordinates.

Text Generation:

Converts the extracted information into natural language descriptions. This might involve specifying the object type, its location relative to the user (e.g., "person on your left"), or its quantity (e.g., "two cars in front of you").

Text-to-Speech Conversion:

Uses a text-to-speech synthesis engine to convert the generated descriptions into audio signals.

Audio Output:

Delivers the audio descriptions through headphones or speakers, providing voice feedback to the user.

V. IMPLEMENTATION

YOLO Model: A pre-trained YOLO model, such as YOLOv5, can be employed for object detection. The model is chosen for its speed and accuracy in real-time applications.

COCO Dataset: The COCO dataset provides a rich set of labeled images for training and evaluating the YOLO model. The dataset encompasses a wide variety of object categories, making the system adaptable to diverse environments.

Text-to-Speech (TTS) Engine: A readily available TTS engine can be integrated into the system to convert text descriptions into audio for voice feedback.

V. BENEFITS

The proposed system offers significant benefits for visually impaired individuals:

- **Enhanced Independence**: By providing real-time audio descriptions of objects and their locations, the system empowers visually impaired users to navigate their surroundings more independently. They can gain a better understanding of their environment, fostering confidence and self-reliance.
- **Improved Spatial Awareness:** The system equips users with the ability to identify and localize objects in their vicinity. This heightened spatial awareness allows them to make informed decisions about their movements and interact with their surroundings more effectively.

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

- **Increased Safety:** By detecting potential obstacles and hazards, the system can alert users to dangers they might not otherwise perceive. This can significantly improve safety, particularly when navigating unfamiliar or crowded environments.
- **Accessibility and Usability:** The system utilizes readily available components like pre-trained YOLO models and TTS engines. This makes it a relatively cost-effective and scalable solution, potentially benefiting a broader population of visually impaired individuals.
- **Real-Time Feedback:** The system operates in real-time, providing continuous feedback on the user's surroundings. This is crucial for dynamic environments and allows for quicker adaptation to changing situations.

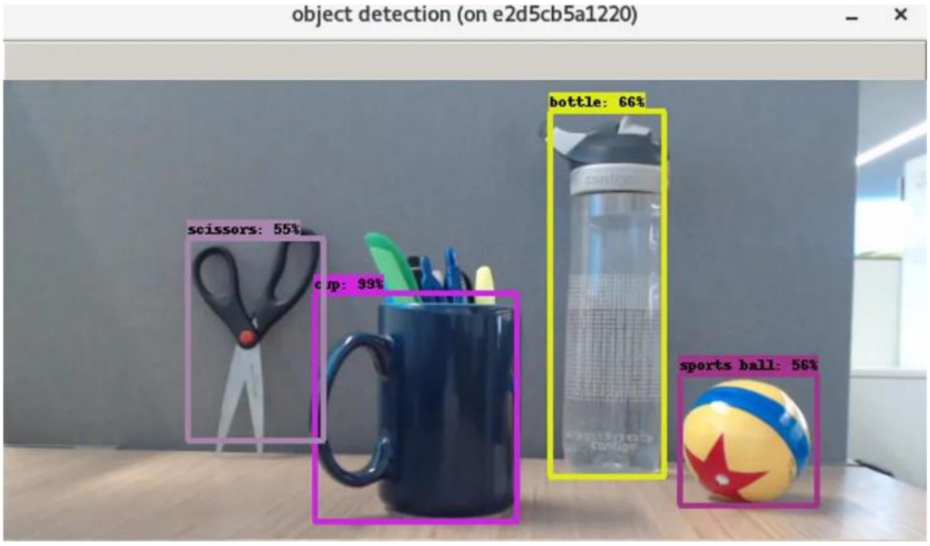

Fig 3:

VI. EVALUATION

The proposed system is evaluated on the COCO validation set to assess its object detection accuracy. Metrics such as mean average precision (mAP) and average recall can be used to quantify the system's performance. Additionally, user studies with visually impaired individuals can be conducted to evaluate the system's usability and effectiveness in real-world scenarios.

VII. FUTURE ENHANCEMENT

In the pursuit of advancing accessibility, future efforts will focus on improving the accuracy of machine learning models for self-explored and customizable datasets, with a particular emphasis on integrating face recognition techniques to aid individuals with visual impairments. This transformative development will be complemented by the creation of hardware solutions such as spectacles or lenses, accompanied by earplugs for audio translation, to ensure technology does not serve as a barrier. By leveraging machine learning and automation, individuals can record and playback visual experiences, offering independence and potentially life-saving assistance. This forward-looking approach seeks to proactively address visual impairment and enhance quality of life without delay.

VIII. CONCLUSION

This paper presented an object detection system with voice feedback designed to assist visually impaired individuals. The system leverages the power of YOLO for real-time object detection in video streams and utilizes the comprehensive COCO dataset for training. Pre-processed video frames are fed into the YOLO model, which predicts bounding boxes and class labels for detected objects. This information is then processed and converted into natural language descriptions using a text-to-speech (TTS) engine, providing auditory feedback to the user. The system demonstrates promise in enhancing the lives of visually impaired individuals by promoting independence and spatial awareness in dynamic environments.

IRJMETS
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:04/April-2024 Impact Factor- 7.868 www.irjmets.com

IX. REFERENCES

[1] Prinsi Patel et al, "Object Detection and Identification" IJATCSE vol.5, 2021
[2] Qi-Chao Mao et al, "Mini-YOLOv3: Real Time Object Detector for Embedded Applications" , 2017
[3] Liquan Zhao et al, "Object Detection Algorithm Based on Improved YOLOv3", 2020
[4] Hui Li et al, "An Improved YOLOv3 for Foreign Objects Detection of Transmission Lines", 2022
[5] Seung-Wook Kim et al, "Edge-Network-Assisted-Real-Time Object Detection for Autonomous Driving", 2021
[6] K.Vaishnavi et al, Real-time Object Detection Using Deep Learning", 2023
[7] F.Sultana et al, "A Review of Object Detection Models based on Convolutional Neural Network", 2019
[8] Hanjun Kim et al, "Real-Time Object Detection System With Multi-Path Neural Networks", 2019
[9] Manjusha Pandey et al, "Application of Deep Learning for Object Detection", 2018
[10] Fei Zhou et al, "Recent advances in Small Object Detection based on Deep Learning",
[11] Kris Kitani et al, "Joint Object Detection and Multi-Object Tracking with Neural Networks", 2021
[12] Yukang Chen et al, "Focal Sparse Convolutional Networks for 3D Object detection", 2017
[13] Fiza Joiya et al, "Object Detection: YOLO vs Faster R-CNN",
[14] Jifeng Dai et al, "Relational networks for Object Detection", 2018
[15] Rikhi Bose et al, "A fast accurate fine-grain object detection model based on YOLOv4 deep neural network", 2021
[16] Chun Bo Sim et al, "A Study on Object Detection using Restructured RetinaNet A Study on Object Detection using Restructured RetinaNet", 2020
[17] Li Shixin et al, "Real-Time Object Detection Using Pre-Trained Deep Learning Models MobileNet-SSD", 2020
[18] Amir Mokhtar Hannane et al, ") Granulated RCNN and Multi-Class Deep SORT for Multi-Object Detection and Tracking Images Using Yolov4", 2024
[19] Anima Pramanik et al, "Granulated RCNN and Multi-Class Deep SORT for Multi-Object Detection and Tracking", 2020
[20] Anuradha B et al, "EODM: On Developing Enhanced Object Detection Model using Fast Region-based Convolution Neural Networks", 2024
[21] Dillon Reis et al, "Real-Time Flying Object Detection with YOLOv8", 2023
[22] Doyen Sahoo et al, "Recent advances in Deep Learning for Object Detection", 2020
[23] Lin Song et al, "End-to-End Object Detection with Fully Convolutional Network", 2021
[24] Srija Patel et al, " Object Detection using You Only Look Once (YOLO) Algorithm in Convolution Neural Network (CNN)", 2023
[25] Namo Jain et al, "Object Detection using Machine Learning and Deep Learning", 2023