International Journal of Science and Research Archive
eISSN: 2582-8185
Cross Ref DOI: 10.30574/ijsra
Journal homepage: https://ijsra.net/

(Review Article)

Check for updates

YOLOv5 driven smart glasses for visually impaired

R. SWEATHA * and S. SATHIYA PRIYA

Electronics and Communication Engineering, Panimalar Institute of Technology, Anna university, Chennai, India.

International Journal of Science and Research Archive, 2024, 12(01), 353–374

Publication history: Received on 27 March 2024; revised on 05 May 2024; accepted on 08 May 2024

Article DOI: https://doi.org/10.30574/ijsra.2024.12.1.0804

Abstract

This project introduces an innovative AI-based interactive shopping assistance system tailored specifically for visually impaired individuals, aiming to enhance their shopping experience and promote inclusivity in retail environments. Leveraging advanced artificial intelligence algorithms, the system provides personalized assistance, navigation support, and seamless interaction for users facing visual challenges. Through intuitive interfaces and adaptive technologies, disabled individuals are empowered to navigate retail spaces with greater independence and confidence. The system utilizes a camera module and YOLO-based deep learning algorithms on a Raspberry Pi 4 for real-time object classification, converting processed information into accessible audio output. In addition to these features, the system incorporates a language customization capability using OCR and gTTS technology, enabling conversion of audio output into multiple languages based on user preference. By prioritizing the needs of visually impaired individuals, this groundbreaking system aims to enhance accessibility and promote an inclusive shopping experience for all.

Keywords: YOLOv5; OCR; gTTS; Raspberry Pi 4; Visually impaired; Deep learning algorithms.

1. Introduction

In an era of advancing technology, the quest for inclusivity and accessibility remains paramount. This project unveils a pioneering AI-based interactive shopping assistance system meticulously crafted for the benefit of visually impaired individuals.

Aimed at revolutionizing the retail landscape, this innovative system harnesses cutting-edge artificial intelligence algorithms to cater specifically to the unique needs of those facing visual challenges. By seamlessly integrating personalized assistance, navigation support, and intuitive interfaces, the system empowers visually impaired individuals to navigate retail environments with newfound independence and confidence.

Through the ingenious utilization of a camera module and YOLO-based deep learning algorithms on a Raspberry Pi4, real-time object classification is achieved, with processed information seamlessly converted into accessible audio output.

Furthermore, the system goes beyond conventional bounds by incorporating a language customization capability, facilitated by OCR and gTTS technology, enabling the conversion of audio output into multiple languages according to user preference. With an unwavering commitment to prioritizing the needs of visually impaired individuals, this groundbreaking system sets forth to redefine accessibility and foster inclusivity, heralding a new era of shopping experiences for all.

*Corresponding author: R. SWEATHA

1.1. Problem Statements

The visually impaired community faces significant challenges when navigating retail environments independently, leading to a lack of inclusivity and accessibility in shopping experiences. Traditional shopping assistance systems often fall short in catering to their specific needs, resulting in a barrier to independent shopping experiences for visually impaired individuals. To address this gap, this project proposes the development of an innovative AI-based interactive shopping assistance system tailored explicitly for the visually impaired.

The primary objective of this project is to enhance the shopping experience for visually impaired individuals and promote inclusivity in retail environments. Leveraging advanced artificial intelligence algorithms, including YOLO-based deep learning techniques, the proposed system aims to provide personalized assistance and navigation support to users facing visual challenges. By incorporating intuitive interfaces and adaptive technologies, the system seeks to empower disabled individuals to navigate retail spaces with greater independence and confidence.

Key features of the proposed system include :

- Real-time object classification using a camera module and YOLO-based deep learning algorithms deployed on a Raspberry Pi4 platform.
- Conversion of processed information into accessible audio output, enabling users to receive real-time guidance and information about their surroundings.
- Language customization capability using OCR and gTTS technology, facilitating the conversion of audio output into multiple languages based on user preference. Here I have used English and Tamil languages.
- Intuitive interfaces and adaptive technologies to ensure ease of interaction and seamless navigation for visually impaired users.
- Conversion of both detected handwritten text and printed text into voice output, providing users with audible information about the text content.

1.2. Scope of the project :

The scope of the project are listed below:

- System design and development
- Object detection and classification
- Handwritten Text detection and Conversion
- Language Customization and Accessibility
- User Interface and Interaction Design
- Testing and Validation
- Documentation and Deployment
- Ethical Considerations and Accessibility Guidelines

The scope of the project encompasses the design, development, testing, and deployment of an AI-based interactive shopping assistance system tailored for visually impaired individuals. By focusing on object detection, handwritten text recognition, language customization, and user interaction, the project aims to enhance accessibility and promote inclusivity in retail environments for all users.

2. Literature survey

2.1. Empowering differently abled individuals throughIOT: smart glasses for the visually impaired people

Authors :MAANIL LAAD, DR. VASUDHA BAHL,Information Technology, Maharaja Agrasen Institute of Technology, Delhi.

Year: June, 2023.

The paper focuses on developing smart glasses for visually impaired individuals, utilizing image processing and OCR for text recognition. The Raspberry Pi 2 acts as the central processing unit, enabling real-time image capture and processing. Results confirm the glasses' effectiveness in aiding reading tasks, highlighting the potential for improved text recognition accuracy and speed. The technology aims to bridge accessibility gaps for visually impaired individuals, emphasizing the importance of ongoing research for enhanced user experience and functionality.

2.2. Smart glasses using deep learning and stereo camera

Authors : Jung-Hwna Kim, Sun-Kyu Kim, Tea-Min Lee, Yong-Jin Lim and Joonhong Lim.

Year: February, 2020.

This paper introduces a novel approach to assist blind individuals through the development of smart glasses utilizing deep learning and stereo camera technology. The main function of these smart glasses is to detect obstacles in the user's path in real-time, providing both the location and type of obstacle through advanced image recognition algorithms. By incorporating YOLO v3 algorithm for obstacle recognition, the system can efficiently identify obstacles at different scales. The experimental results demonstrate the effectiveness of the proposed device in detecting and recognizing obstacles, thereby enhancing the safety and mobility of visually impaired individuals during navigation.

2.3. Design of smart goggle for visually impaired with audio features

Authors: Sujay C V , Suhas G R, V K Sunidi, Varuni V, Mrs.Samatha R Swamy.

Year: April ,2023.

The main content of the paper focuses on the design and implementation of a Smart Goggle for visually impaired individuals. This innovative device utilizes deep learning algorithms, specifically YOLO V3, for object detection to assist users in navigating their surroundings. The Smart Goggle provides audio feedback through voice assistance, allowing users to understand the type of obstacles they encounter. Additionally, the device includes features such as a camera module for image capture, SD card storage, and headphone output for audio feedback. Overall, the Smart Goggle aims to enhance the independence and safety of visually impaired individuals by providing real-time assistance and information about their environment.

2.4. Food detection with image processing using convolutional neural network(cnn)method

Authors: AssyifaRamdani ,AgusVirgono, CasiSetianingsih

Year: August,2020.

This paper focuses on utilizing Convolutional Neural Network (CNN) for food detection and classification in restaurant settings. The main content includes data collection of six types of food, image processing techniques for digital image conversion, and the importance of food classification for health and dietary purposes. The system design involves using CNN for food detection on a desktop application, streamlining the billing process in restaurants. The paper highlights the significance of CNN in image classification, object detection, and feature extraction, emphasizing its effectiveness in food recognition tasks. Overall, the research aims to enhance food service industry operations by automating food detection and simplifying billing processes through advanced image processing technologies.

2.5. A binary object detection pattern model to assist the visually impaired in detecting normal and camouflaged faces

Authors: S. Sajini , B. Pushpa

Year: December, 2023.

The paper introduces a Binary Object Detection Pattern Model (BODPM) designed to assist visually impaired individuals in detecting normal and camouflaged faces using computer vision techniques. The model utilizes the KERAS dataset to recognize objects with face key points and evaluate their proximity and accuracy. Information on recognized objects, including date and time, is converted into voice output for the visually impaired user. The BODPM framework consists of two key phases supported by integrated algorithms. The proposed model outperforms existing ones, showcasing remarkable performance with a peak accuracy of 98%. Overall, the study aims to enhance object recognition and proximity detection for visually impaired individuals, potentially revolutionizing assistive technologies in this field.

Table 1 Trends and Technologies discussed in literature

Project title	Methods	Pros	Cons
Empowering Differently Abled Individuals through IoT: Smart Glasses for the Visually Impaired , June 2023, IRE journals. [1]	The project utilizes the Raspberry Pi 2 as the central processing unit for the smart glasses, facilitating image acquisition and processing. The Raspberry Pi camera module is employed for real-time visual capture. Software components include image processing algorithms, the Tesseract OCR engine for text extraction, and a text-to-speech synthesizer for audio output. Additionally, Simulink software is utilized for advanced image processing operations to enhance text recognition accuracy.	1.Wearable and hands-free device that assists in reading tasks 2. By utilizing the Raspberry Pi 2 and its camera module, the project offers an affordable and accessible technology solution. 3. Integration ofIoT technology improves the quality of life for differently abled individuals	1. The need for additional image processing techniques to compensate for the limitations of the Raspberry Pi camera. 2.It does not explicitly mention multi-language support in the audio output. 3.OCR algorithms specifically focus on recognizing and interpreting characterswithin the image.
Smart Glasses using Deep Learning and Stereo Camera , February, 2020,IEEE(GCCE). [2]	The project utilizes a stereo camera, gyro sensor, vibration motors, and LED lights for hardware. Software includes ROS Kinetic Kame, Darknet-53 architecture, and YOLO v3 algorithm for deep learning. The system operates on an Intel(R) Core(TM) i7-8700K server with Ubuntu 16.04 and Windows 10, using wireless communication for data transmission.	1. The use of YOLOv3 allows for effective and efficient classification of obstacles. 2.The stereo camera system enables accurate and real-time detection of obstacles 3. Wireless communication facilitates fast data transmission to external servers for processing	1. The use of YOLOv3 in the project may result in slightly lower accuracy and speed compared to YOLOv5. 2. YOLOv5 is known for its improved speed and accuracy compared to YOLOv3, which may result in faster and more precise obstacle recognition.
Design Of Smart Goggle For Visually Impaired With Audio Features, April ,2023, IJARCCE(International Journal of Advanced Research in Computer and Communication Engineering) .[3]	The project utilizes hardware components such as ultrasonic sensors, an SD card module, a headphone module, a switch, a GPRS SIM900A module, and an Atmega328p microprocessor. On the software side, deep learning algorithms like YOLO v3 for object detection, Android OS for system operation, and Text-to-Speech (TTS) technology are used. These components work together to create a Smart Goggle system that assists visually impaired individuals by providing real-time feedback and	1. Uses the YOLO V3 algorithm for object detection, which is known for its speed and accuracy in real-time applications. 2. Integration of Voice Assistance. 3.Ease of Interaction with Surroundings.	1. Limited Language Support. 2.Dependency on Specific Hardware: The project relies on the EPSON BT-300 smart goggles. 3.Potential Delay in Speech Output: The typical delay between the smart goggles and speech

	essential information about their surroundings.		output is mentioned as 3.788 seconds .
Food detection with Image processing using Convolutional Neural Network(CNN) method ,August,2020, The 2020 IEEE international conference on artificial Intelligence, industry 4.0 and communication Technology. [4]	The hardware comprises a webcam and tripod for capturing food images. Software include image processing tools, Convolutional Neural Network (CNN) for food detection, and a desktop application for displaying processed data. This integration enables efficient food detection and automatic price estimation.	1.Achieved 100% detection accuracy for 6 types of food. 2.Built a dataset specific to the food items , which can lead to better performance. 3.Utilized a partion of 80% training data and 20% test data and low learning rate of 0.0002.	1. Limited Object Detection: The CNN method may not be as versatile as YOLOv5. 2.Training a CNN model with a high number of epochs is time consuming. 3.Require more computational resources compared to YOLOv5.
A Binary Object Detection Pattern Model to Assist the Visually Impaired in detecting Normal and Camouflaged Faces,December, 2023,Engineering, Technology & Applied Science Research (ETASR) . [5]	The hardware used in this project likely includes a camera or sensor for capturing visual data, a processing unit such as a computer or embedded system to run the object detection algorithms, and possibly a speaker or audio output device for communicating information to the visually impaired user. The project likely involves computer vision libraries or frameworks such as OpenCV for image processing and object detection, machine learning tools like Keras for training the model, and possibly text-to-speech software for converting the recognized objects into voice output for the user.	1.BODPM (Binary Object Detection Pattern Model) is a hybrid model that covers image capture, recognition, and proximity detection . 2.CamouflagedObject Detection. 3.BODPM outperformed existing models in terms of accuracy, error rate, and F-score	1. Limited Scope: BODPM is specifically designed for object detection and recognition. 2. The performance of BODPM is dependent on the dataset used for training. 3. Require a certain level of expertise in computer vision.

3. Difficulties faced by visually impaired individuals

3.1. Introduction

Visually impaired individuals encounter numerous challenges in their daily lives, particularly when navigating unfamiliar environments such as retail spaces. These challenges often stem from the reliance on visual cues and information inaccessible to individuals with visual impairments. In the context of shopping for medical products, the difficulties become even more pronounced due to the critical nature of these purchases and the potential consequences of errors or misunderstandings.

3.2. Challenges faced while buying medical products

According to World Health Organization (WHO), Out of 63 participants surveyed, 71% expressed a desire for independence in managing their medication, while 79% reported self-administering their medication despite facing difficulties. The most common challenges reported included difficulty in locating medicines (25.39%), identifying medicines and containers (17.46%), and administering liquid medications (25.39%). These difficulties resulted in issues such as inaccurate dosing (14.28%), missed doses (39.68%), and premature discontinuation of treatment (28.57%). Participants utilized various coping strategies to overcome these challenges, including using differently shaped containers, attaching medicines to their clothing, and using finger measurements for liquid medicines. Additionally, incidents of medication mishaps, such as consuming vinegar instead of a gripe mixture and mistakenly putting ear drops into the eyes, were reported

According to National Institutes of Health (NIH), In the study, 82 participants were interviewed, comprising 42 males and 40 females, demonstrating an equal distribution of genders, as depicted in Table 1.

Table 2 Gender distribution, visual impairment, level of education and braille literacy of participants

Demography	Male		Female		Total	
Gender Distribution	42	-	40	-	82	-
Level of visual impairment						
Some sight	34		30		64	78
No sight at all	8		10		18	22
Highest level of education						
No formal education	9	21.4	10	25	19	23
Primary	17	40.5	11	27.5	28	34
Secondary	10	23.8	10	25	20	24
Tertiary	6	14.3	9	22.5	15	19
Ability to read braille						
Yes	7	-	8	-	15	18.3
No	35	-	32	-	67	81.7

3.3. Existing system

Existing systems for aiding visually impaired individuals often incorporate various technologies to assist them in navigation, object recognition, and text-to-speech conversion

3.4. Mobile applications

There are several mobile applications designed to assist visually impaired individuals in navigating their surroundings. These apps often use GPS technology to provide location information and may incorporate features such as text recognition and object detection. However, they are typically limited by the processing power and capabilities of the mobile device.

3.4.1. Wearable devices

Some wearable devices, such as smartwatches or wristbands, have been developed to provide haptic feedback or audio cues to assist visually impaired users in navigation. While these devices can be useful for simple tasks like receiving notifications or providing directional guidance, they may lack the advanced capabilities needed for complex object recognition.

- **Smart Cane**: Traditional white canes for the visually impaired have been augmented with sensors and electronics to detect obstacles and provide feedback to the user through vibrations or auditory signals. However, these systems are limited in their ability to recognize complex objects or text.

3.5. Effects of environmental conditions on existing system

- **Outdoor Navigation Challenges**: Environmental factors such as varying terrain, weather conditions (e.g., rain, snow), and changes in lighting can pose significant challenges for existing systems relying on sensors or GPS technology. Obstacles may be obscured or difficult to detect, leading to potential navigation errors or safety hazards for visually impaired users.
- **Interference and Signal Degradation**: In urban environments with dense infrastructure or crowded spaces, existing systems utilizing electronic sensors or GPS signals may experience interference or signal degradation, impacting their accuracy and reliability. This interference can disrupt object detection capabilities and hinder the effectiveness of navigation assistance provided to users.

- **Limited Adaptability to Dynamic Environments**: Existing systems may struggle to adapt to dynamic or unpredictable environments, such as crowded streets, busy intersections, or construction zones. Changes in environmental conditions or the presence of temporary obstacles may not be adequately detected or accounted for, compromising the system's ability to provide accurate and timely assistance to users.
- **Accessibility in Indoor Settings**: While existing systems may excel in outdoor navigation, they may face challenges in indoor environments, such as shopping malls, airports, or public buildings. Factors like varying layouts, crowded spaces, and limited GPS coverage can hinder the system's ability to provide reliable navigation guidance and object recognition indoors, limiting its usefulness for users in these settings.
- **Integration with Environmental Sensors**: To mitigate the impact of environmental conditions on system performance, future iterations of existing systems could incorporate additional environmental sensors capable of detecting factors such as temperature, humidity, and air quality.

4. System design

The Below Block diagram shows the schematic diagram of the proposed system

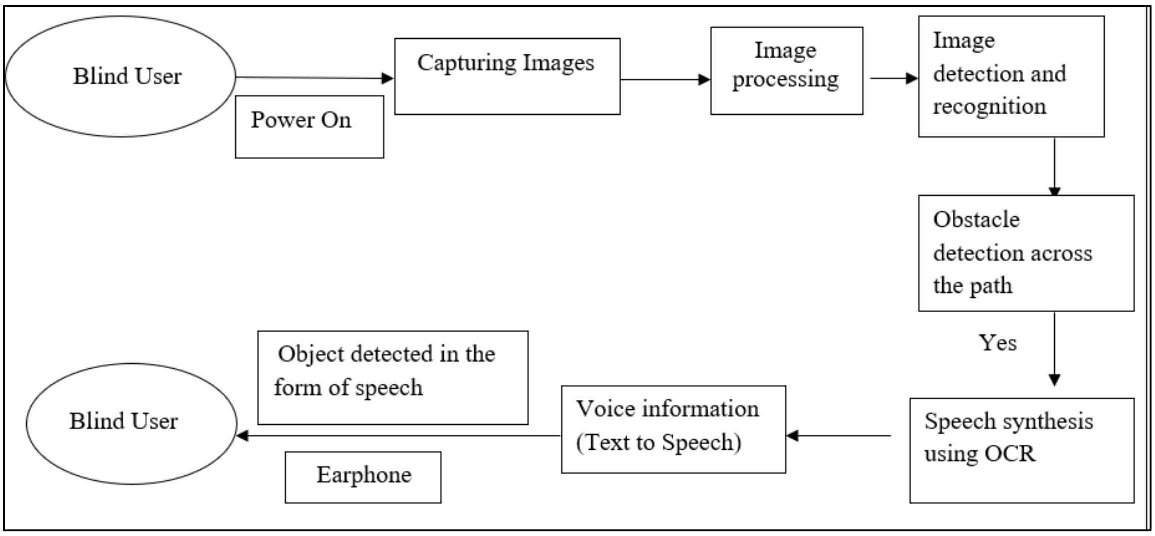

Figure 1 Schematic diagram of Proposed system

This project aimed at aiding visually impaired individuals, the input system consists of the mechanisms through which the system receives information about the user's environment and their interactions.

Through the ingenious utilization of a camera module and YOLO-based deep learning algorithms on a Raspberry Pi4, real-time object classification is achieved, with processed information seamlessly converted into accessible audio output. Furthermore, the system goes beyond conventional bounds by incorporating a language customization capability, facilitated by OCR technology, enabling the conversion of audio output into multiple languages according to user preference. With an unwavering commitment to prioritizing the needs of visually impaired individuals, this groundbreaking system sets forth to redefine accessibility and foster inclusivity, heralding a new era of shopping experiences for all.

4.1. Input system

The system begins its operation when powered on. A camera module, integral to the system, captures an image of the surrounding environment. This captured image is then sent to the Raspberry Pi, a miniaturized computer that acts as the system's brain. Within the Raspberry Pi, the image undergoes processing to enable object recognition. This recognition process analyzes the image to identify and understand the objects present within it. In some configurations, the system might also be equipped with text recognition capabilities. If this functionality is enabled, the Raspberry Pi would not only identify objects but also extract text from the captured image, further enriching the information the system can provide.

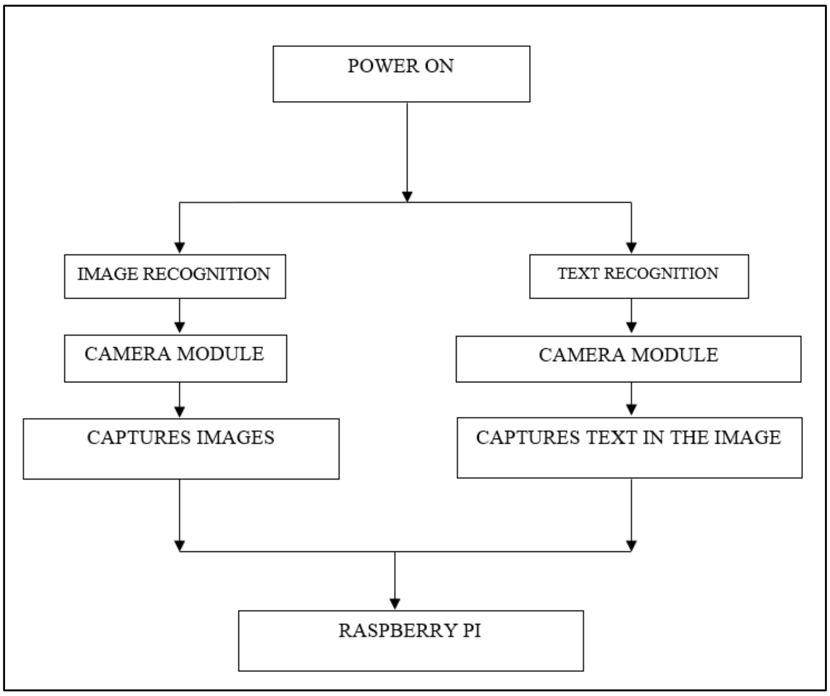

Figure 2 Schematic diagram of Input system

Object Recognition: The smart glasses utilize the YOLOv5 algorithm to detect and recognize objects, obstacles, and people in the user's surroundings .The system provides real-time feedback through the display, informing the user about the presence and location of detected objects.

Text Recognition: Using the OCR engine and gTTS technology, the smart glasses can accurately recognize printed text from images captured by the camera.

In this project, the OCR model utilized possesses the capability to detect text in both Tamil and English languages. This system is designed to identify both handwritten and printed text, offering seamless support to users. Specifically, for Tamil text detection, the OCR Tamil framework has been employed.

4.2. Output system

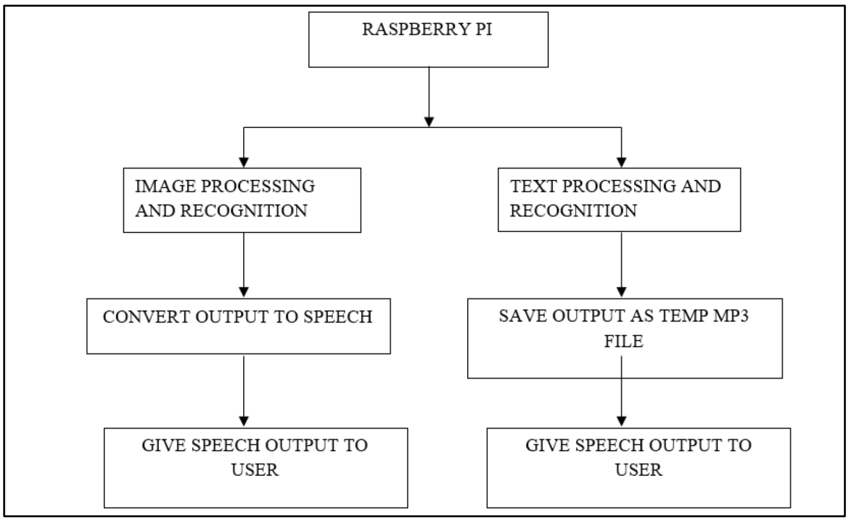

Figure 3 Schematic diagram of Output system

Object Detection with YOLOv5: At first we need to perform object detection using YOLOv5 on our desired images or video frames. This will give you information about the objects present in the input.

4.3. Text Detection with OCR

If the user wants to detect any particular text, after clicking on the tactile push button, the image will be captured, and the OCR (Optical Character Recognition) process will start to detect the text inside the image. Once done, the output will be stored as a temporary MP3 file.

- **Text-to-Speech Conversion**: After generating the text descriptions, you can use a text-to-speech (TTS) library or service to convert the text into spoken words. There are several TTS libraries , here I have used pyttsx3 in python and gTTS and OCR to support both Tamil and English text.
- **Audio Output**: Finally, play the generated audio to hear the descriptions of the detected objects. Depending on your application, you might output the audio to speakers, headphones, or integrate it into a larger system

4.4. Real time monitoring system

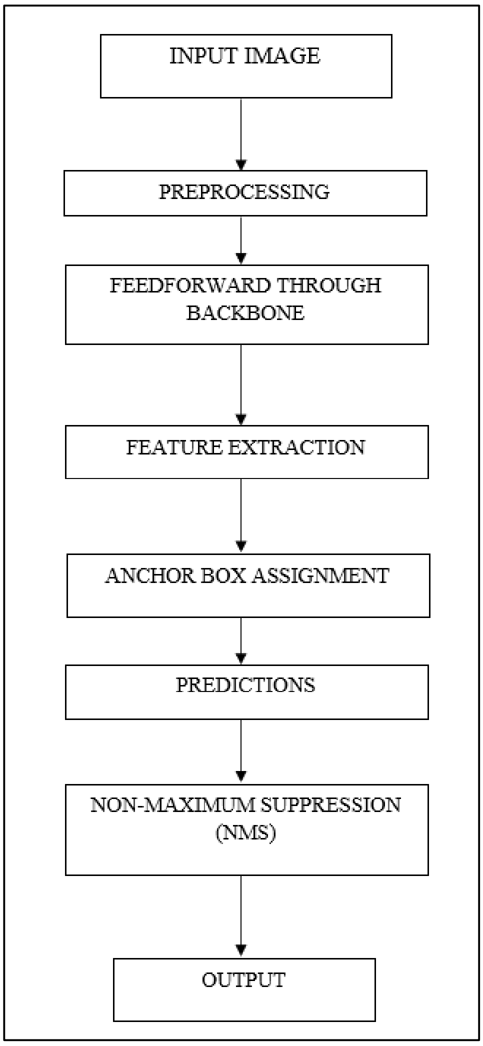

Figure 4 Schematic diagram of YOLOv5 detection process

In today's fast-paced world, real-time monitoring systems play a crucial role in various domains. These systems enable continuous observation, analysis, and response to events as they occur, providing valuable insights and enhancing decision-making processes.

One approach to implementing a real-time monitoring system is through the use of computer vision and optical character recognition (OCR) technologies. By leveraging these technologies, we can develop intelligent systems capable of analyzing visual data and extracting meaningful information in real-time.

The first component of our real-time monitoring system utilizes YOLOv5, a state-of-the-art object detection model, to detect objects of interest in images or video streams.The second component of our system focuses on text detection and recognition using OCR techniques.

The below flowchart depicts the target detection of YOLOv5 detection.

The YOLOv5 detection process involves several key steps that transform an input image into a set of detected objects with bounding boxes and class labels. Initially, an image is fed into the model, where it undergoes preprocessing to ensure it meets the requirements of the YOLOv5 architecture. This includes resizing the image to a fixed size and normalizing its pixel values.

Next, the preprocessed image is passed through the backbone network, typically a convolutional neural network such as CSPDarknet53. During this step, features are extracted from the image at different scales using techniques like the feature pyramid network (FPN).

The extracted features are then used for anchor box assignment, where anchor boxes are assigned to different feature maps based on their scales. These anchor boxes serve as reference points for the model to predict bounding boxes around detected objects.

Predictions are made for each grid cell on each feature map, comprising bounding box coordinates, confidence scores indicating the presence of objects, and class probabilities for different object categories.

To refine the detections and remove redundant bounding boxes, non-maximum suppression (NMS) is applied. This involves filtering out bounding boxes with low confidence scores and keeping only the most confident boxes when multiple overlapping detections occur.

The final step involves outputting the detection results, including the bounding box coordinates, class labels, and confidence scores for each detected object. Optionally, additional post-processing steps may be applied, such as filtering based on class probabilities or conducting further analysis on the detected objects.

Finally, the detection results are displayed on the input image or outputted in a suitable format for further processing, completing the YOLOv5 detection process.

This project employs EasyOCR library to detect and extract text from images captured by the Raspberry Pi's camera module. This allows us to analyze textual information in real-time.

By combining object detection and text recognition capabilities, our real-time monitoring system can provide comprehensive insights into the surrounding environment.

Overall, the integration of object detection and text recognition technologies into a real-time monitoring system offers numerous benefits in terms of efficiency, accuracy, and scalability. By harnessing the power of artificial intelligence and edge computing, we can create intelligent systems that enhance situational awareness, improve operational efficiency, and contribute to safer and more secure environments.

The YOLOv5 detection process involves several key steps that transform an input image into a set of detected objects with bounding boxes and class labels. Initially, an image is fed into the model, where it undergoes preprocessing to ensure it meets the requirements of the YOLOv5 architecture. This includes resizing the image to a fixed size and normalizing its pixel values.

Next, the preprocessed image is passed through the backbone network, typically a convolutional neural network such as CSPDarknet53. During this step, features are extracted from the image at different scales using techniques like the feature pyramid network (FPN).

The extracted features are then used for anchor box assignment, where anchor boxes are assigned to different feature maps based on their scales. These anchor boxes serve as reference points for the model to predict bounding boxes around detected objects.

Predictions are made for each grid cell on each feature map, comprising bounding box coordinates, confidence scores indicating the presence of objects, and class probabilities for different object categories.

To refine the detections and remove redundant bounding boxes, non-maximum suppression (NMS) is applied. This involves filtering out bounding boxes with low confidence scores and keeping only the most confident boxes when multiple overlapping detections occur.

The final step involves outputting the detection results, including the bounding box coordinates, class labels, and confidence scores for each detected object. Optionally, additional post-processing steps may be applied, such as filtering based on class probabilities or conducting further analysis on the detected objects.

Finally, the detection results are displayed on the input image or outputted in a suitable format for further processing, completing the YOLOv5 detection process.

This project employs EasyOCR library to detect and extract text from images captured by the Raspberry Pi's camera module. This allows us to analyze textual information in real-time.

By combining object detection and text recognition capabilities, our real-time monitoring system can provide comprehensive insights into the surrounding environment.

Overall, the integration of object detection and text recognition technologies into a real-time monitoring system offers numerous benefits in terms of efficiency, accuracy, and scalability. By harnessing the power of artificial intelligence and edge computing, we can create intelligent systems that enhance situational awareness, improve operational efficiency, and contribute to safer and more secure environments.

5. Hardware specification

5.1. Hardware requirements:

- Raspberry PI 4 Model B
- Camera
- Ultrasonic sensor
- USB cable
- Micro SD card
- Pushbutton
- Battery (5V, 3A)

5.1.1. Raspberry PI 4 Model B:

The Raspberry Pi 4 plays a central role as the hardware platform that hosts the AI-based interactive shopping assistance system. Its role encompasses several key functions:

- **Processing Power**: The Raspberry Pi 4 serves as the computing powerhouse for the system, capable of handling complex AI algorithms required for real-time object classification and processing. Its quad-core ARM Cortex-A72 CPU and ample RAM provide sufficient computational resources to run the deep learning algorithms efficiently.
- **Integration with Camera Module:** The Raspberry Pi 4 interfaces with a camera module, which is utilized to capture live video feed of the retail environment. This feed is then processed by the AI algorithms running on the Raspberry Pi 4 for object recognition and classification.
- **YOLO-based Deep Learning Algorithms:** The Raspberry Pi 4 executes YOLO-based deep learning algorithms responsible for real-time object classification. YOLO (You Only Look Once) is a state-of-the-art object detection system known for its speed and accuracy, making it well-suited for applications requiring rapid processing of visual data like this shopping assistance system.
- **Audio Output Generation:** Processed information from the object classification is converted into accessible audio output. The Raspberry Pi 4 likely handles the generation of this audio output, translating the detected objects into spoken words or descriptions that can be easily understood by visually impaired individuals.

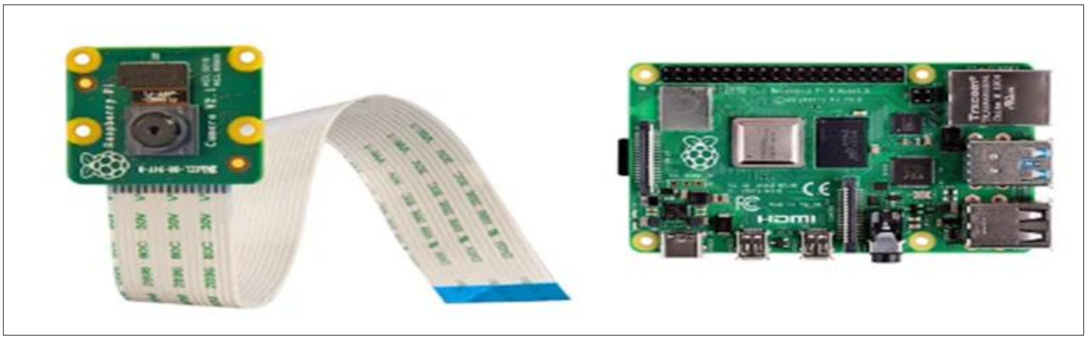

Figure 5 Raspberry pi 4 and camera module

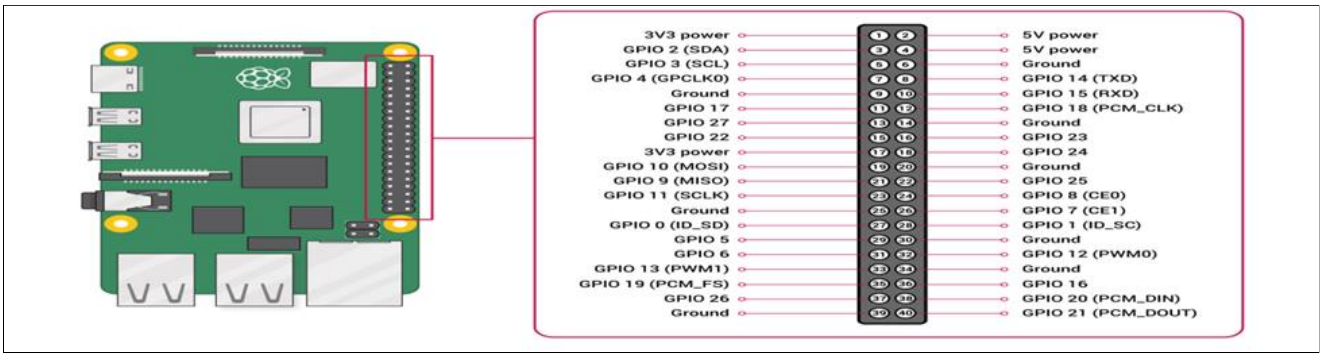

Figure 6 Raspberry pi 4 and it's GPIO pins

5.1.2. Camera

Table 3 Specifications of Raspberry PI 4 model B

Feature	**Specification**
Processor	Broadcom BCM2711 SoC
CPU	1.5 GHZ quad-core Cortex-A72 (later models: 1.8 GHz)
RAM	1GB, 2GB, 4GB, or 8GB LPDD (depending on model)
Storage	MicroSD card slot
Video output	2x micro HDMI ports (up to 4Kp60 supported)
USB	2x USB 3.0 ports, 2x USB 2.0 ports
Networking	Gigabit Ethernet port
Wireless	802.11b/g/n/ac Wi-Fi, Bluetooth 5.0
Power	5V DC via USB-C connector (minimum 3A)
GPIO	Standard 40-pin GPIO header
Operating Temperature	0-50° C
Dimensions	89.6 mm x 56.5 mm x 16.8 mm

Table 4 Specifications of Raspberry PI camera module

Feature	**Specification**
Image Sensor	Sony IMX219 PQ CMOS
Resolution	8 megapixels (3280 x 2464)
Lens	Fixed focus, field of view: ~54° diagonal
Interface	CSI-2 (MIPI Camera Serial Interface)
Connection	15-pin ribbon cable
Frame Rates	1080p @ 30fps, 720p @ 60fps, 640x480p @ 60/90fps
Video Modes	H.264, MJPEG
Still Capture Formats	JPEG, BMP
Dimensions	23.86 mm x 25 mm x 9 mm
Weight	3g
Compatibility	Works with all models of Raspberry Pi 1, 2, 3, and 4

5.1.3. Ultrasonic sensor

An ultrasonic sensor is a device that emits high-frequency sound waves and detects their reflection off nearby objects. It can be utilized in such a system for proximity detection and obstacle avoidance, which are crucial functionalities for assisting visually impaired individuals in navigating retail environments.

- **Proximity Detection**: The ultrasonic sensor can be used to detect the presence of obstacles or objects in the path of the visually impaired individual. By measuring the time it takes for sound waves to bounce back from nearby objects, the sensor can determine the distance to those objects and alert the user through audio cues or vibrations.
- **Enhancing Navigation**: In addition to object detection, the ultrasonic sensor can contribute to the overall navigation experience by providing feedback on the layout of the environment and identifying open pathways or clear spaces for the user to move through.

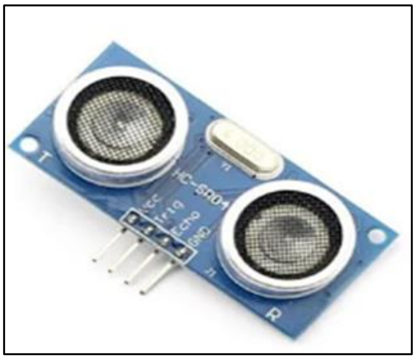

Figure 6 Ultrasonic sensor

Table 5 Specifications of Ultrasonic sensor

Feature	**Specification (HC-SR04)**
Operating Voltage	5V DC
Operating Current	< 15mA
Working Frequency	40Hz
Measuring Distance	2cm - 400cm (theoretical), 2cm - 80cm (practical)

Accuracy	Â±3mm
Measuring Angle	< 15 degrees
Trigger Input Pulse Width	10Âµs (microseconds)
Echo Output Signal	TTL level signal proportional to distance
Dimensions	45mm x 20mm x 15mm (typical)

5.1.4. MICRO SD CARD

- Storage: The microSD card acts as the primary storage device for your Raspberry Pi. It stores the Raspberry Pi operating system (OS), applications, data, and any media files you use.
- Booting: The Raspberry Pi boots its operating system from the microSD card. When you power on your Raspberry Pi, it reads the instructions and files necessary to start the OS from the microSD card.

Figure 7 Micro SD card front and back view

Table 6 Specifications of Micro SD card

Feature	**Specification**
Brand	SanDisk
Type	Micro SDHC (micro Secure Digital High Capacity)
Capacity	32GB
Speed Class	Class 4

5.1.5. Push button

A tactile push button, also known simply as a push button or tactile switch, is a type of switch that is commonly used in electronic circuits and devices. Its primary function is to provide a simple and momentary electrical contact when pressed.

Tactile push buttons are often used as user input devices in electronic systems. They allow users to trigger specific actions or functions, such as turning a device on or off, selecting options in a menu, or triggering an event in a program.

Figure 8 Four legged tactile push button

Connect one leg of the switch pair to a GPIO pin on the Raspberry Pi and the other leg to a ground (GND) pin. You'll also need to connect the two legs of the mechanical stability pair to the 3.3V or 5V pin and a GND pin for stable operation.

In your Python script, set up the GPIO pins for input and use the GPIO.input() function to detect button presses.

Table 7 Specifications of push button

Feature	Specification	Example value
Number of Legs	4	-
Contact Configuration	Single-Pole Single-Throw (SPST)	-
Actuation Type	Momentary	-
Operating Force	Typically 100 grams (may vary)	100 gf
Operating Voltage	Typically 12V DC (may vary)	12V DC
Maximum Current Rating	Typically 50 mA (may vary)	50 mA
Contact Resistance	Typically <100 milliohms (may vary)	<100 mΩ
Insulation Resistance	Typically >100 megohms (at 250V DC)	>100 MΩ
Operating Temperature	-25°C to +70°C	
Life Expectancy	Typically 100,000 to 500,000 cycles (may vary)	100,000-500,000 cycles
Termination Style	Through-hole mounting (THT)	
Dimensions	Typically 6mm x 6mm x 5mm (may vary)	6x6x5 mm
Actuator Type	Round or square	

5.1.6. Battery (5V, 3A)

The 5V, 3A battery serves as the power source for the entire project, providing the necessary electrical energy to operate the various components and systems involved.

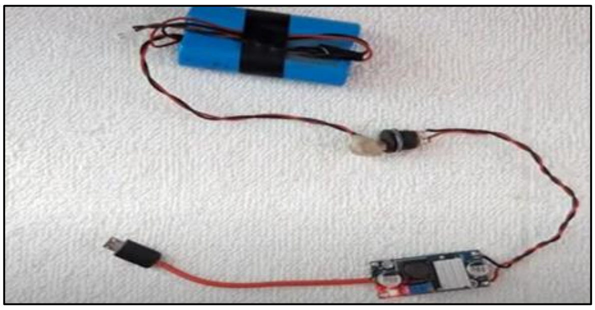

Figure 9 5V, 3A battery

After connecting all peripherals, including the USB Pendrive, the Raspberry Pi is powered on. Upon booting up, the Raspberry Pi OS guides the user through the initial setup process, including configuring Wi-Fi network settings, creating a username and password, and enabling SSH.

6. Software specification

Software required

- Python 3.8+
- Pytorch
- YOLOv5 library
- Packages for image manipulation (Open CV)
- OCR, gTTS, pyttsx3

- Pygame

6.1. PYTHON 3.8+

Python is a high-level programming language known for its simplicity and readability. Python serves as the primary programming language for writing the scripts.

6.2. PYTORCH

PyTorch is an open-source machine learning library primarily developed by Facebook's AI Research lab (FAIR). It provides tensor computation (similar to NumPy) with strong GPU acceleration and deep neural networks built on a tape-based autograd system. In this project, PyTorch is used for object detection using the YOLOv5 model.

6.3. YOLOv5 LIBRARY

YOLOv5 is a real-time object detection model developed by Ultralytics. The YOLOv5 library provides pre-trained models and utilities for running object detection inference on images, videos, and live camera feeds. In the code, the YOLOv5 library is used for object detection tasks.

6.4. Packages for image manipulation (OpenCV)

OpenCV (Open Source Computer Vision Library) is a popular open-source computer vision and image processing library. It provides a wide range of functions for tasks such as image manipulation, object detection, object tracking, and more. In this project, OpenCV is used for tasks such as capturing frames from a webcam, reading and writing images, drawing bounding boxes, and displaying images.

6.5. OCR (optical character recognition)

OCR is the process of converting images of typed, handwritten, or printed text into machine-encoded text. In the code, the OCR functionality is used to detect and recognize text from images captured by the webcam or provided as input.In this project, OCR is used to detect and recognize text from images captured by the webcam. It provides a means to extract textual information from images.

6.5.1. gTTS (Google Text-to-Speech)

gTTS is a Python library and CLI tool to interface with Google Translate's text-to-speech API. It allows converting text into spoken audio in various languages. In the code, gTTS is used to convert detected text into speech for auditory feedback.In this project, gTTS is used to convert the detected text (either in Tamil or English) into speech.

6.5.2. PYTTSX 3

Pyttsx3 is a text-to-speech conversion library in Python. It supports multiple TTS engines and provides a simple interface to convert text into spoken audio. In the code, pyttsx3 is used for speech synthesis to provide auditory feedback.In the provided code, pygame is used to handle audio playback. It initializes the audio mixer and plays back the synthesized speech generated by gTTS.

6.6. PYGAME

Pygame is a set of Python modules designed for writing video games. It provides functionality for graphics, sound, and user input. In the code, Pygame is used for audio playback to play the speech synthesized by gTTS or pyttsx3.In the provided code, pygame is used to handle audio playback. It initializes the audio mixer and plays back the synthesized speech generated by gTTS.

7. Results

7.1. Prototype model

The smart glasses system developed for visually impaired individuals successfully demonstrated its core functionalities during testing. The object detection module, powered by the YOLOv5 algorithm, effectively identified both static and moving objects in real-time captured by the camera. Objects were accurately detected with high confidence levels, providing users with crucial environmental awareness. During testing, the object detection module performed reliably in detecting various objects.

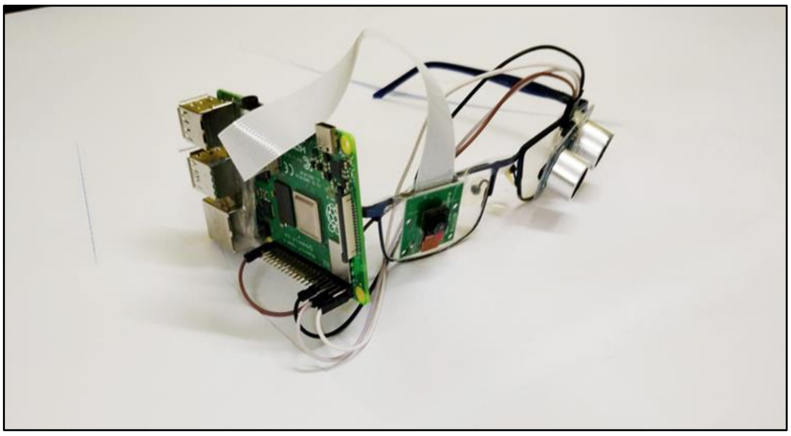

Figure 10 Prototype model of the project

Python Scripts Contains the code for object detection using YOLOv5, text recognition using EasyOCR, and audio playback using gTTS and pygame.

Raspberry Pi OS is installed on the SD card. This is achieved by flashing the Raspberry Pi OS image onto the SD card using tools like Etcher.

After connecting all peripherals, including the USB Pendrive, the Raspberry Pi is powered on. Upon booting up, the Raspberry Pi OS guides the user through the initial setup process, including configuring Wi-Fi network settings, creating a username and password, and enabling SSH.During the initial setup or later via the Raspberry Pi desktop interface, the user configures Wi-Fi network settings to enable internet connectivity.

Once the Raspberry Pi is set up and connected to the internet, the user can remotely access it via SSH from another computer. The Python scripts containing the object detection, text recognition, and audio playback code can be transferred to the Raspberry Pi and executed via the terminal.

During the operation of the Python scripts, the webcam captures video frames, which are processed for object detection and text recognition. Detected objects and recognized text are then spoken aloud using the audio playback capabilities of the Raspberry Pi.

The user can monitor the execution of Python scripts and control the Raspberry Pi remotely via SSH.

7.1.1. Output

During our training, validation, and testing phases on the Google Colab platform, we utilized Weights & Biases to monitor and visualize the training and validation processes. Throughout this training and validation, we focused on analyzing the following losses to gain a deeper understanding of the advantages offered by our proposed system.

- Box loss (box): This represents the loss associated with the bounding box coordinates. It measures how accurately the model predicts the coordinates (x, y, width, height) of the bounding boxes around objects in the image.
- Objectness loss (obj): This loss term evaluates how well the model predicts the presence of an object within a bounding box. It penalizes the model when it fails to detect objects or when it incorrectly predicts the confidence score associated with the presence of an object.
- Class loss (cls): This term represents the loss associated with predicting the class labels of the detected objects. It measures how accurately the model assigns class labels to the objects within the bounding boxes.

7.1.2. Output table

Table 8 Output table for object loss, class loss, box loss

Epoch	Box Loss	Object Loss	Class Loss
1	0.113	0.02862	0.03193
2	0.1009	0.02748	0.02876
3	0.09814	0.03047	0.02845
4	0.09048	0.03027	0.02801
5	0.08133	0.03065	0.02334
6	0.0779	0.03522	0.02058
7	0.06877	0.02911	0.02064
8	0.07353	0.03321	0.02112
9	0.07415	0.03756	0.02041
10	0.0673	0.03108	0.01978
11	0.06633	0.03197	0.02106
12	0.06024	0.03097	0.01816
13	0.06382	0.0323	0.01612
14	0.0575	0.03096	0.01948
15	0.0534	0.02783	0.0187
16	0.0579	0.0289	0.01807
17	0.0522	0.02797	0.01921
18	0.05374	0.02985	0.01628
19	0.05	0.02933	0.01625
20	0.05021	0.02901	0.01591
21	0.05151	0.02452	0.01455
22	0.05025	0.02599	0.01451
23	0.04662	0.02713	0.01236
24	0.04915	0.02476	0.01263
25	0.05327	0.02575	0.01285
26	0.04796	0.02656	0.01347
27	0.04437	0.02297	0.01184
28	0.04555	0.0231	0.01372
29	0.04874	0.02585	0.009709
30	0.04759	0.02415	0.008228
31	0.04649	0.02248	0.008292
32	0.04702	0.02486	0.00795
33	0.04905	0.02267	0.007749
34	0.04705	0.02156	0.007984

35	0.04656	0.02264	0.007616
36	0.04868	0.02149	0.007798
37	0.04439	0.02123	0.009858
38	0.0449	0.02364	0.00676
39	0.04488	0.02071	0.007322
40	0.04482	0.02212	0.005533
41	0.04272	0.0212	0.005485
42	0.04473	0.02082	0.005503
43	0.04538	0.01929	0.005698
44	0.04236	0.02218	0.005507
45	0.04106	0.02248	0.005698
46	0.04239	0.02259	0.004752
47	0.04248	0.02064	0.003645
48	0.04222	0.01906	0.007068
49	0.04196	0.02023	0.005482
50	0.0428	0.021	0.005287

7.1.3. Graphs

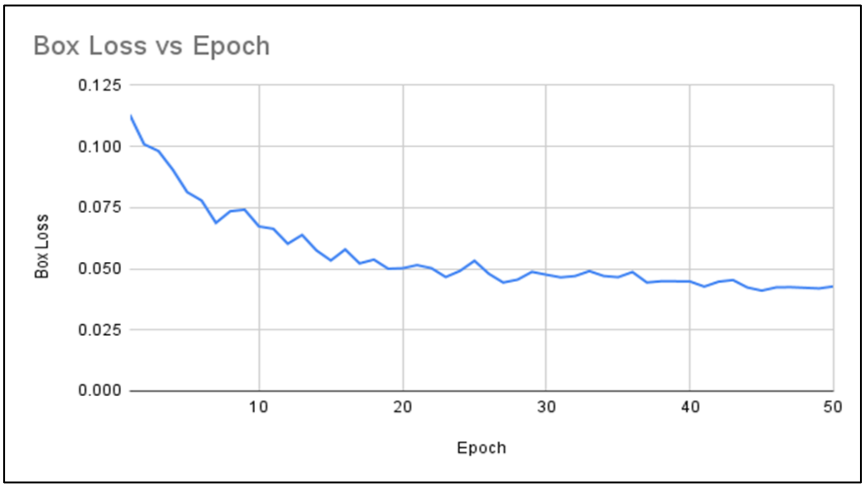

Figure 11 Graph for Box loss Vs number of epochs

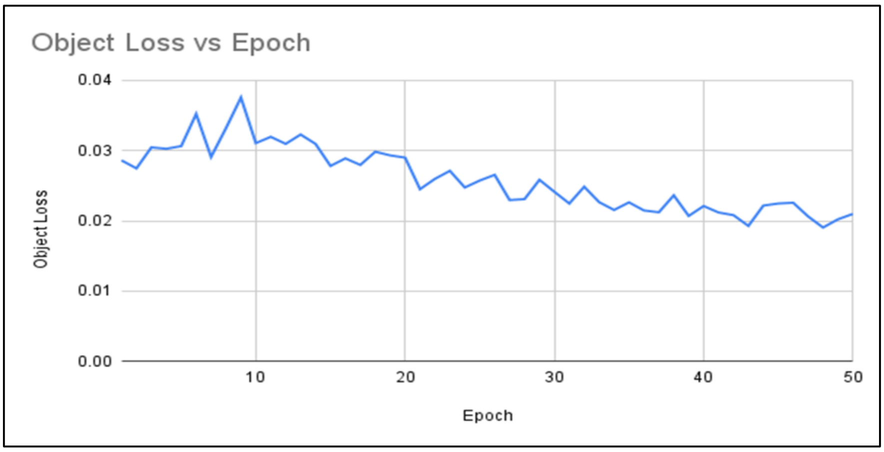

Figure 12 Graph for Object loss Vs number of epochs

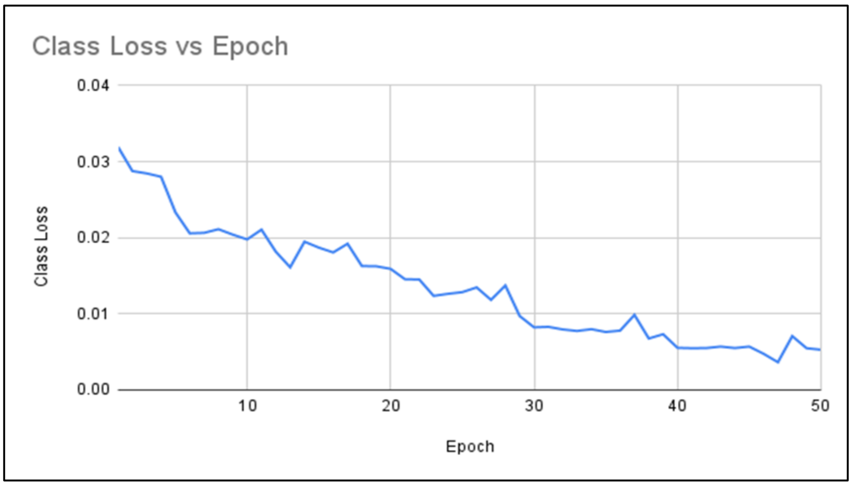

Figure 13 Graph for Class loss Vs number of epochs

8. Conclusion

The completion of this project marks a significant advancement in the development of smart glasses for visually impaired individuals. Through the integration of cutting-edge technologies such as real-time object detection using the YOLOv5 algorithm, text recognition powered by OCR, and multilingual voice command support. The results obtained from the implementation demonstrate the effectiveness of the system in accurately detecting both moving and static objects. In conclusion, the successful development and testing of the smart glasses prototype signify a significant step forward in assistive technology for the visually impaired. By harnessing the power of artificial intelligence and machine learning, we have created a tool that not only addresses practical challenges but also promotes autonomy, accessibility, and inclusivity for individuals with visual impairments. Moving forward, continued refinement and optimization of the system will be crucial in realizing its full potential and ensuring widespread adoption and impact in the community.

Future scope

The future scope of this project is vast and promising, with numerous avenues for expansion and enhancement:

- Enhanced Object Recognition: Continuously improving the AI algorithms to recognize an even wider array of objects with higher accuracy. This could include recognizing product brands, packaging details, and more subtle product features.
- Integration with Smart Devices: Integration with wearable devices such as smart glasses or haptic feedback systems to provide even more seamless navigation and interaction within retail environments.

- Augmented Reality (AR) Integration: Incorporating AR overlays to provide additional information about products, such as pricing, reviews, or nutritional information, directly within the user's field of view.
- Expansion to Other Environments: Adapting the system to work in other environments beyond retail, such as public transportation hubs, museums, or educational institutions, to assist visually impaired individuals in various aspects of daily life.
- Natural Language Processing (NLP) Integration: Incorporating NLP technology to allow for more natural and conversational interactions with the system, such as asking questions about specific products or requesting recommendations.
- Community Engagement and Education: Launching outreach programs and educational initiatives to raise awareness about the availability and benefits of the system within the visually impaired community and among retailers.

List of abbreviations

- YOLO - You Only Look Once
- OCR - Optical Character Recognition
- gTTS - Google Text-to-Speech
- CNN - Convolutional Neural Network
- BODPM - Binary Object Detection Pattern Model
- GPRS - General Packet Radio Service
- GPS - Global Positioning System
- MP3 - MPEG Audio layer 3
- NMS - Non-Maximum Suppression
- FPN - Feature Pyramid Network
- CSP - Cross Stage Partial Network
- ARM - Advanced RISC Machine
- GPIO - General Purpose Input/Output
- BMP - Bitmap
- SDHC - Secure Digital High Capacity
- CLI - Command Line Interface
- SSH - Secure Shell
- NLP - Natural Language Processing

Compliance with ethical standards

Acknowledgement

We have taken efforts to complete this project. However, it would not have been possible without the kind support and help of many individuals and our institution. We are using this opportunity to express our gratitude to everyone who supported us throughout the course of this final year project.

Our sincere thanks to Our Honourable Founder and Chairman, Late. Col. Dr. JEPPIAAR M.A., Ph.D., for educating us in his premier institution. We would like to express special gratitude and thanks to Our Beloved Secretary and Correspondent, Dr. P. CHINNADURAI, M.A., Ph.D., for providing us such a platform to complete our project.

We express warm thanks to Our Directors Mrs. C. VIJAYA RAJESWARI, Dr. C. SAKTHI KUMAR, M.E and Dr. SARANYA SREE SAKTHI KUMAR, B.E for their support and valuable guidance.

We are sincerely grateful to Our Principal, Dr. T. JAYANTHY, M.E., Ph.D., for sharing her illuminating views on a number of issues related to our project.

We are highly indebted to Our Head of the Department and internal guide, Dr. S. SATHIYAPRIYA, M.E., Ph.D.,, for her guidance and constant supervision as well as for providing necessary information regarding the project and also for her support in completing the project.

We are thankful to our Project Co-ordinator Dr. I. Poonguzhali M.E., Ph.D.,, for their aspiring guidance, invaluably constructive criticism and friendly advice during the project work.

We would also like to thank all our department staff who provided us with the facilities being required and conductive conditions for our project.

Last but not least we would like to express our gratitude towards our parents for their kind co-operation and encouragement which helped us in completion of this project.

Disclosure of conflict of interest

No conflict of interest to be disclosed.

References

[1] Empowering Differently Abled Individuals through IoT: Smart Glasses for the Visually Impaired PeopleMAANIL LAAD, DR. VASUDHA BAHL,Information Technology, Maharaja Agrasen Institute of Technology, Delhi, June, 2023.

[2] Smart glasses using deep learning and stereo camera Jung-Hwna Kim, Sun-Kyu Kim, Tea-Min Lee, Yong-Jin Lim and Joonhong Lim, February, 2020.

[3] Design Of Smart Goggle For Visually Impaired With Audio Features Sujay C V, Suhas G R, V K Sunidi, Varuni V, Mrs.Samatha R Swamy, April ,2023.

[4] Food detection with Image processing using Convolutional Neural Network(CNN) method , AssyifaRamdani , AgusVirgono, CasiSetianingsih, August,2020.

[5] A Binary Object Detection Pattern Model to Assist the Visually Impaired in detecting Normal and Camouflaged Faces , S. Sajini , B. Pushpa , December, 2023.