A PROJECT STAGE-1 REPORT
ON

"YOLOInsight : AI-Powered Assistive Device for Visually Impaired"

"Sponsored By Enumtech Technology Solutions LLP"

SUBMITTED TO
THE SAVITRIBAI PHULE PUNE UNIVERSITY, PUNE
IN THE PARTIAL FULFILLMENT FOR THE AWARD OF THE DEGREE
OF

BACHELOR OF ENGINEERING

IN

INFORMATION TECHNOLOGY

BY

Dabhade Saurabh Bhairu (B1906508520)
Dsilva Sean Louis (B1906508523)
Shaikh Kabir Allauddin(B1906508531)

Under the Guidance of: -
Prof.Gajanan Arsalwad

DEPARTMENT OF INFORMATION TECHNOLOGY
KJ'S EDUCATIONAL INSTITUTE
TRINITY COLLEGE OF ENGINEERING AND RESEARCH

Academic Year: 2024-2025

CERTIFICATE

This is to certify that the project stage-1 report entitled

"YOLOInsight : AI-Powered Assistive Device for Visually Impaired"

Submitted by

Dabhade Saurabh Bhairu (B1906508520)
Dsilva Sean Louis (B1906508523)
Shaikh Kabir Allauddin (B1906508531)

is a bonafide work carried out by them under the supervision and guidance of **Prof. Gajanan Arsalwad**, and it is approved for partial fulfillment of the requirement for the BE (Information Technology Engineering) course of Savitribai Phule Pune University for the award of the Degree of Bachelor of Engineering (Information Technology Engineering).

This project report has not been earlier submitted to any other institute or university for the award of any degree or diploma.

Prof. Gajanan Arsalwad **Internal guide** **Department of IT**	**Dr. Vilas Gaikwad** **Head Of Department** **Department of IT**
External Examiner	**Dr. Abhijeet Auti** **Principal** **TCOER,Pune**

ACKNOWLEDGEMENT

We would like to take this opportunity to express our gratitude towards all the people who have in various ways helped in the successful completion of our project. We must convey our gratitude to **Prof. Gajanan Arsalwad** for giving us the constant source of inspiration, personally correcting our work, and providing encouragement throughout the project.

We express deep gratitude towards **Dr. V. S. Gaikwad**, Head of the Information Technology Engineering Department, for providing support and giving his valuable time.

We also express our gratitude to our sponsors, **Enumtech Technology Solutions LLP**, Pune, and **Mr. U.Y. Solanki**, Managing Director of Enumtech Technology Solutions LLP, for their invaluable support and encouragement.

We would also like to extend our heartfelt thanks to our father, mother, other family members, and friends for their continuous encouragement, valuable suggestions, and motivation.

Group No. PR2024-25-08

Name of Students (Seat No)	**Sign.**
Dabhade Saurabh Bhairu (B1906508520)	
Dsilva Sean Louis (B1906508523)	
Shaikh Kabir Allauddin(B1906508531)	

Abstract

The YOLOInsight project introduces an innovative AI-powered real-time assistance system tailored for visually impaired individuals, aiming to enrich their daily experiences and promote inclusivity across various environments. This system leverages cutting-edge artificial intelligence algorithms, specifically the YOLO (You Only Look Once) framework, to provide personalized assistance, navigation support, and seamless interaction for users facing visual challenges.

The methodology involves the integration of a camera module with a Raspberry Pi 4, which serves as the processing unit for real-time object detection and classification. The YOLOv8 algorithm is employed for its efficiency and accuracy in identifying a wide range of objects within the user's environment. The processed information is then converted into accessible audio output using Optical Character Recognition (OCR) and Google Text-to-Speech (gTTS) technology, allowing for immediate auditory feedback in multiple languages based on user preferences.

To enhance user interaction, the system incorporates a built-in Natural Language Processing (NLP) API, which functions similarly to virtual assistants like Siri or Google Assistant, but operates independently without reliance on third-party services. This design choice not only ensures complete control over the system's features but also enhances user privacy and reduces operational costs.

Experimental results demonstrate the effectiveness of the YOLOInsight system across diverse lighting conditions, achieving a mean Average Precision (mAP) of 0.94 in normal lighting and maintaining reliable performance in low-light scenarios. User feedback indicates a significant improvement in navigation confidence and environmental awareness, with participants reporting a greater sense of independence when using the system in real-world settings such as shopping centers and public spaces.

By prioritizing the specific needs of visually impaired individuals, the YOLOInsight project aims to improve accessibility, foster autonomy, and deliver a transformative assistive technology solution. The findings highlight the potential of integrating advanced AI technologies in creating inclusive environments, paving the way for future enhancements and broader applications in assistive devices.

Contents

Abbreviations

List of Abbreviations

- **AI** - Artificial Intelligence
- **IoT** - Internet of Things
- **ML** - Machine Learning
- **OCR** - Optical Character Recognition
- **gTTS** - Google Text-to-Speech
- **NLP** - Natural Language Processing
- **TTS** - Text-to-Speech
- **YOLO** - You Only Look Once
- **CV** - Computer Vision
- **GPS** - Global Positioning System

List of Figures

List of Tables

Chapter 1

Synopsis

Name of the College	:	Trinity College of Engineering and Research
Name of the Branch	:	Information Technology
Name of the Students	:	Mr.Saurabh Dabhade
		Mr. Sean D'silva
		Mr. Kabir Shaikh
Project Group Number	:	PR2024 25 08
Name of the Guide	:	Prof. Gajanan Arsalwad
Broad Area	:	IOT and AI/ML
Project Title	:	YOLOInsight: AI-Powered Smart Glasses

1.1 Introduction

This project presents a groundbreaking AI-powered assistance system designed to empower visually impaired individuals with greater independence. Using real-time object detection, personalized navigation, and a custom NLP API, it provides intuitive support for navigating public spaces. Powered by a Raspberry Pi 4 and YOLO-based deep learning, it converts visual data into audio feedback, enhanced by OCR and gTTS technologies. This innovative system redefines accessibility, promoting inclusivity and setting a new standard for assistive technology in transforming how visually impaired individuals interact with the world.

1.2 Problem Statement

Visually impaired individuals face significant challenges in navigating real-world environments independently, with traditional systems often falling short. This project introduces an AI-powered assistance system using YOLO-based deep learning for object detection and a custom NLP API for navigation. Designed to provide personalized support, it empowers users to navigate confidently and independently with adaptive, intuitive technology.

1.3 Literature Review

1.3.1 Existing Research

Several studies have investigated assistive technologies for the visually impaired:

- **Ethan Waisberg (2023)** [1]: Explores Meta smart glasses with AI assistants and camera technology for real-time image processing, converting visual information into speech for navigation, object recognition, and text reading, highlighting the potential of augmented reality and AI for enhancing interaction.

- **Sweatha and Sathiya Priya (2024)** [2]: Present an AI-based shopping assistance system using YOLO-based deep learning on a Raspberry Pi 4, providing real-time object classification and multilingual audio feedback to improve accessibility during shopping.

- **T. Annapurna (2024)**[3]: Proposes an IoT-based Vision Alert system using wearable glasses with cameras and advanced sensors for obstacle detection, providing real-time feedback through a buzzer to aid independent navigation.

- **Poy (2024)**[4]: Suggests a standalone smart glasses solution with YOLOV8 for object recognition, KerasOCR for text recognition, and FaceNet for face recognition, powered by a Raspberry Pi Zero 2 W and Intel Neural Compute Stick 2 to promote self-reliance without external devices.

1.3.2 Identified Gaps

The following gaps were identified between the YOLOInsight project and the base paper:

- Base paper: basic smartphone cameras; YOLOInsight: custom smart glasses with specialized sensors.
- Base paper: on-device processing; YOLOInsight: hybrid processing with privacy focus.
- Base paper: task-specific solutions; YOLOInsight: modular, scalable architecture.
- Base paper: traditional object detection; YOLOInsight: YOLOv8 for edge computing.
- Base paper: lacks text recognition and navigation; YOLOInsight: adds Tesseract OCR and GPS with obstacle detection.

1.4 Proposed Solution

This project uses AI and deep learning for a real-time assistance system to help the visually impaired navigate independently.

1. Object Classification: Real-time detection using YOLO and Raspberry Pi 4.
2. Audio Output: Converts data to audio guidance.
3. Custom NLP API: Enables natural user interaction.
4. Text-to-Speech: Converts text to speech.
5. Precise Audio Feedback: Provides accurate, timely audio information.

1.4.1 Methodology

The methodology for YOLOInsight includes:

- **Data Collection:** Gather images and videos for object detection, text recognition, and environmental data for the visually impaired.
- **Image and Text Processing:** Apply preprocessing techniques for object detection and text-to-speech conversion.

- **Model Development:** Use YOLOv8 for object detection, Tesseract OCR for text recognition, and custom NLP API for user interaction.
- **System Integration:** Integrate models with hardware (Raspberry Pi, camera) and sensors for real-time processing.
- **Evaluation:** Assess performance based on accuracy, real-time feedback, and user experience.
- **Deployment:** Deploy on smart glasses for real-time assistance, ensuring seamless interaction and audio feedback.

1.5 Schedule

Sr. No	Task	Duration (weeks)
1	Analyze and Verify Project Scope	1
2	Establish Project Goals	1
3	Prepare Synopsis	1
4	Study of Algorithms	3
5	Implementation of YOLOv8 Model	4
6	Model Testing and Evaluation	2
7	System Deployment	2
Total		**16**

Table 1.1: Project Timeline

1.5.1 Hardware and Software Requirements

Hardware Requirements:

- Raspberry Pi 4 with camera module.
- High-resolution camera for object detection.
- 5V, 3A battery for extended use.
- Earphones for audio feedback.
- Ultrasonic sensors for obstacle detection.

Software Requirements:

- **Languages:** Python, C/C++, JavaScript.
- **Frameworks:** TensorFlow/PyTorch, Flask/Django, OpenCV.
- **OS:** Linux (Raspberry Pi), Android/iOS (App).

Additional Hardware:

- 12 MP+ camera.
- GPU-enabled server for training.
- 512 GB SSD for storage.

1.6 Expected Outcomes

The smart glasses system is expected to:

- Accurately detect objects in real-time using YOLOv8.
- Provide high-confidence object identification for better awareness.
- Enable text recognition to assist with daily tasks.
- Ensure seamless user interaction through the NLP module.
- Perform reliably in real-world scenarios, aiding visually impaired users.

1.7 Conclusion

The YOLOInsight project improves navigation for the visually impaired through AI-powered object detection, OCR, and NLP/Voice Assistant. Future upgrades will include cloud integration for better processing and security measures to protect user data, enhancing the system's functionality and reliability.

Chapter 2

Introduction

In an age where technology is rapidly evolving, the pursuit of true inclusivity and accessibility stands as a crucial objective. This project introduces a groundbreaking AIpowered interactive assistance system, meticulously designed to empower visually impaired individuals with greater autonomy and ease in navigating their environments. Focused on transforming the daily experiences of those with visual impairments, the system leverages cutting-edge artificial intelligence to provide tailored support for their specific needs. By seamlessly integrating real-time object detection, personalized navigation, and a custom-built NLP API, the solution offers an intuitive and efficient way for users to confidently navigate retail spaces and public environments with unprecedented independence. Powered by a camera module and YOLO-based deep learning algorithms running on a costeffective Raspberry Pi 4, the system achieves real-time object classification and immediately converts visual data into accessible audio outputs. It goes a step further by incorporating advanced OCR and gTTS technologies, offering multilingual audio feedback tailored to each user's language preference, thus enhancing its accessibility and personalization. With an unwavering focus on the unique requirements of visually impaired individuals, this pioneering system redefines the boundaries of assistive technology. It not only promotes inclusivity but also sets a new standard for accessibility, transforming the way visually impaired individuals experience and interact with the world around them

2.1 Aims and Motivation for the Project

The YOLOInsight project aims to design a cost-effective, AI-powered assistive device for visually impaired individuals, enhancing their independence and quality of life. By integrating advanced technologies such as real-time object detection, text recognition, and natural language processing (NLP), this wearable device empowers users to navigate their environments confidently. Utilizing AI algorithms like YOLOv8, the system provides immediate feedback about surroundings, improving accessibility in public spaces. Focused on affordability, YOLOInsight leverages Raspberry Pi to broaden access to assistive technology while incorporating a custom-built NLP API for a personalized experience that safeguards user privacy. By harnessing the latest advancements in AI and IoT, the project strives to develop innovative solutions that promote inclusivity and help dismantle barriers faced by visually impaired individuals.

2.2 Objectives of the Study

This project aims to develop an AI-based interactive assistance system designed to enhance independence for visually impaired individuals. The smart glasses will provide real-time assistance by combining advanced object detection, text recognition, and accessible audio feedback.

The primary objectives are as follows:

1. Develop an AI-based interactive assistance system for visually impaired individuals.
2. Implement real-time object detection and text recognition.
3. Provide accessible audio feedback for seamless navigation.

Chapter 3

Literature survey

In an age where technology evolves rapidly, the goal of true inclusivity and accessibility is paramount. This project introduces YOLOInsight, an AI-powered assistive system specifically designed to empower visually impaired individuals by providing greater autonomy in navigating their surroundings. Built on advanced artificial intelligence, YOLOInsight combines real-time object detection, personalized navigation, and a custom NLP API to facilitate intuitive and efficient support in both public and private spaces.

Unlike traditional assistive technologies, YOLOInsight employs a cost-effective Raspberry Pi 4 platform with a camera module running YOLO-based deep learning algorithms for real-time object classification. Visual data is instantly converted to accessible audio output, tailored through OCR and gTTS technologies to support multiple languages, thereby enhancing usability. This approach echoes similar advancements discussed in the literature, such as the Meta smart glasses outlined by Waisberg (2024) [1], the AI-based shopping assistance system by Sweatha and Sathiya Priya (2024) [2], and the IoT-based Vision Alert system proposed by Annapurna (2024) [3].

Building upon pioneering work in smart bionic vision systems Badawi, 2024 [5] and OCR integration for smart glasses D.Raju, 2024 [6], YOLOInsight uniquely integrates multi-modal sensing with ultrasonic sensors to boost spatial awareness and obstacle detection. This multi-sensory approach ensures reliable navigation support, addressing key challenges identified in previous studies, such as those by Laad (2023) [7] and Sunidi (2023) [8], which emphasize the need for enhanced navigation capabilities and user-friendly design.

YOLOInsight introduces proprietary AI models optimized for speed and efficiency, setting it apart from many systems that rely on generic models. This custom approach aims to meet the specific needs of visually impaired users, building on the lightweight processing solutions explored by Sajini (2024) [9] and Abdulatif (2022) [10], which also underscore the importance of real-time, edge-based computing. Additionally, the system's affordability, achieved through the Raspberry Pi platform and specialized software solutions, makes it accessible to a wider audience—a priority highlighted in related research by Sudharani (2022) [11] and Sunidi (2023) [8].

YOLOInsight prioritizes auditory feedback as the primary mode of interaction, aligning with the design philosophy presented by Laad (2023) [7] and Sajini (2024) [9], who focus on optimizing user experience through audio-based guidance. This comprehensive approach, integrating object detection, text recognition, navigation support, and voice interaction, offers a cohesive, all-in-one solution that addresses multiple users needs within a single device. Furthermore, this project aligns with Abdulatif (2022) [10] and Laad (2023) [7], who advocate for low-cost, power-efficient solutions that maximize independence for visually impaired users.

Overall, YOLOInsight builds upon a wealth of prior research and introduces unique, user-centric features to redefine the standard of accessibility for visually impaired individuals. By combining innovative technology with a focus on practical usability, YOLOInsight fosters inclusivity and independence, paving the way for a new era of assistive technology.

3.1 Gap Survey

The gap analysis between the YOLOInsight project and the base paper highlights significant advancements across several dimensions. While the base paper focuses on on-device processing with basic smartphone cameras and sensors, YOLOInsight enhances this with custom smart glasses equipped with specialized sensors, improving both performance and user experience. Additionally, YOLOInsight employs a hybrid processing model that combines on-device and optional cloud capabilities, emphasizing local data processing for better privacy compliance—an aspect not addressed in the base paper.

The development approach shifts from task-specific solutions in the base paper to a modular, scalable architecture in YOLOInsight, making it more adaptable to future needs. Furthermore, YOLOInsight moves beyond the base paper's traditional object detection methods (Haar-Cascade, BLOB, SURF) to utilize the YOLOv8 model for more efficient edge computing. It also introduces text recognition via Tesseract OCR and enhances navigation by integrating GPS with obstacle detection, providing a more comprehensive solution for visually impaired users. Overall, YOLOInsight significantly improves upon the foundational concepts of the base paper, offering innovative solutions that enhance user experience and functionality.

Chapter 4

Problem Statement/Definition

Visually impaired individuals face significant challenges in navigating real-world environments independently, with traditional systems often falling short. This project introduces an AI-powered assistance system using YOLO-based deep learning for object detection and a custom NLP API for navigation. Designed to provide personalized support, it empowers users to navigate confidently and independently with adaptive, intuitive technology.

4.1 Key Challenges

- **Real-Time Object Detection Accuracy:** Ensuring that the system can accurately identify and classify a wide range of objects in real-time using deep learning algorithms, even in complex or crowded environments.
- **Cost Efficiency Without Compromising Functionality:**Developing a cost-effective solution that maintains high functionality, using components like the Raspberry Pi 4 while achieving performance comparable to more expensive alternatives. .
- **User-Friendly and Intuitive Interface:** Designing a simple, user-friendly interface that visually impaired individuals can navigate easily without extensive training, ensuring a seamless experience.
- **Audio Feedback Precision:**Providing clear and precise audio feedback to users, ensuring that the information conveyed is accurate, timely, and easily understandable in

dynamic settings.

4.2 Proposed Solution

This project leverages cutting-edge AI and deep learning technologies to create a real-time assistance system for visually impaired individuals. By combining advanced hardware and software components, the system aims to offer accurate object recognition, seamless interaction, and personalized user experiences. With features such as real-time object detection, accessible audio output, the solution empowers users to navigate their surroundings independently. The integration of intuitive interfaces, a custom NLP API, and text-to-speech capabilities ensures that visually impaired individuals can interact naturally with the system, receiving timely information and guidance tailored to their needs.

The objectives of this solution include:

- **Real-Time Object Classification:** Utilizing a camera module and YOLO-based deep learning algorithms deployed on a Raspberry Pi 4, the system provides accurate real-time object detection and classification
- **Accessible Audio Output::** Processed information is converted into accessible audio, enabling users to receive real-time guidance and information about their surroundings.
- **Custom NLP API:** The project features a proprietary NLP API, allowing for more natural and conversational interactions, enabling users to ask questions about their surroundings and receive immediate responses.
- **Text-to-Speech Conversion:**The system can convert both detected handwritten and printed text into voice output, providing users with audible information about textual content.
- **Audio Feedback Precision:**Providing clear and precise audio feedback to users, ensuring that the information conveyed is accurate, timely, and easily understandable in dynamic settings.

Chapter 5

Software Requirement Specification

5.1 Introduction

This Software Requirements Specification (SRS) provides a detailed overview of the YOLOInsight: AI-powered Smart Glasses system. The project aims to develop assistive technology for visually impaired individuals using AI, machine learning (ML), and IoT integrated into smart glasses hardware with a companion mobile application. The system automates real-time object detection, facial recognition, and text reading to enhance accessibility for users, enabling them to interact with their environment more independently.

5.1.1 Purpose

The purpose of the YOLOInsight project is to create a pair of smart glasses designed to assist visually impaired individuals by using advanced AI technologies. These glasses will detect objects, and read text in real time, providing audio feedback to the user. The solution aims to empower visually impaired users with greater independence in their daily activities by offering real-time assistance.

5.1.2 Scope

The YOLOInsight project will focus on the development of AI-powered smart glasses and a mobile application to facilitate user interaction with the world around them. This project will involve building hardware for smart glasses, developing AI models for object and text detection, and implementing a secure, real-time communication protocol between the glasses and the mobile app.

Key Features Include:

- Real-time object detection and classification using AI models.
- Optical Character Recognition (OCR) for reading printed and handwritten text.
- Secure Bluetooth communication between glasses and mobile app.
- Customizable feedback options (audio, vibration) based on user preferences.

5.1.3 Definitions

- **AI:** Artificial Intelligence - The simulation of human intelligence by machines, especially computer systems, to perform tasks typically requiring human intelligence, such as visual perception, speech recognition, and decision-making.
- **ML:** Machine Learning - A subset of AI that allows systems to learn from data and improve performance over time without being explicitly programmed.
- **IoT:** Internet of Things - A network of interconnected devices that can communicate and exchange data over the internet.
- **OCR:** Optical Character Recognition - A technology used to convert different types of text-based images, such as printed or handwritten text, into machine-encoded text.
- **UI:** User Interface - The visual elements and layouts that users interact with to control the system, including buttons, menus, and other navigation features.
- **API:** Application Programming Interface - A set of tools and protocols that allow different software systems to communicate and interact with one another.

5.2 Overall Description

This section outlines the system's purpose, user characteristics, operational environment, and design constraints for the YOLOInsight: AI-powered Smart Glasses project..

5.2.1 User Classes and Characteristics

- **Visually Impaired Users:** Primary users who will rely on the smart glasses for object detection, facial recognition, and text reading in real time to assist in everyday tasks.
- **Caregivers/Family Members:** Secondary users who assist in setting up or customizing the smart glasses for the primary user's needs.
- **Technical Support Staff:** Professionals responsible for maintaining the system, software updates, hardware troubleshooting, and providing technical support.

5.2.2 Operating Environment

- **Smart Glasses Hardware:** Equipped with a camera, microprocessor, and wireless communication (Bluetooth/Wi-Fi).
- **Mobile Application:** Compatible with Android and iOS devices to connect and sync data with the smart glasses.
- **Cloud Support:** Optional cloud storage for AI model processing and data storage.
- **Operating System:** Linux-based OS for smart glasses, Android/iOS for the mobile app.

5.2.3 Design and Implementation Constraints

- **Limited Processing Power:** Smart glasses have limited hardware capacity to run complex AI models locally.
- **Battery Life:** Power management is crucial to ensure longer usage time of the smart glasses.

- **Data Privacy:** Ensuring the privacy of sensitive information such as facial recognition data and detected text.
- **Model Training:** Model training may require significant computational resources for accuracy improvement.

5.3 Specific Requirements

5.3.1 Functional Requirements

- **Object Detection:** The system shall detect and classify objects in real-time, providing audio feedback to the user.
- **Text Reading (OCR):** The system shall convert detected text into audio output.
- **Mobile Application:** Allows customization of settings and syncing data with the smart glasses.
- **Alerts:** The system shall notify users about important objects.

5.3.2 Non-Functional Requirements

- **Performance:** The system shall provide real-time audio feedback within one second.
- **Scalability:** The system shall support the addition of new functionalities without a significant impact on performance.
- **Usability:** The user interface shall be intuitive and simple, specifically designed for visually impaired users.
- **Reliability and Availability:** The system shall ensure high reliability with 99.9% uptime.

5.3.3 Hardware and Software Requirements

Hardware Requirements:

- Raspberry Pi 4 Model B with a camera module.
- High-resolution camera for real-time object detection.
- Battery (5V, 3A) for prolonged usage.
- Earphones for audio feedback.
- Ultrasonic sensors for obstacle detection.

Software Requirements:

- **Programming Languages:** Python, C/C++, JavaScript.
- **Frameworks:** TensorFlow or PyTorch for AI models, Flask/Django for backend, OpenCV for image processing.
- **Operating Systems:** Linux for Raspberry Pi, Android/iOS for the mobile application.

5.4 Specific Requirements

5.4.1 Functional Requirements

- **Image Capture:** The system shall integrate with a live camera mounted on the smart glasses to continuously capture real-time images of the user's surroundings.
- **Object Detection:** The system shall analyze the captured images using YOLOv8-based machine learning algorithms to detect and classify objects (e.g., chairs, doors, obstacles) in real time.
- **Text Recognition:** The system shall utilize Optical Character Recognition (OCR) to identify and convert printed and handwritten text from the surroundings into audio feedback.
- **Audio Feedback:** The system shall provide real-time audio feedback to the user through earphones or a connected Bluetooth device regarding detected objects, and text.

- **Mobile Application:** The system shall sync with a mobile application to allow users and caregivers to customize settings, store recognized data, and review reports.
- **Alerts and Notifications:** The system shall send audio notifications to the user about important objects.
- **User Interface:** The mobile application shall provide an intuitive interface to view system settings, saved reports, and device connectivity status.

5.4.2 Non-Functional Requirements

- **Performance:** The system shall process images from the live feed and return object recognition and feedback within one second.
- **Scalability:** The system shall support future additions of more object categories, languages, or features without performance degradation.
- **Usability:** The mobile and smart glasses interface shall be intuitive and easy to operate for visually impaired users and caregivers, requiring minimal training.
- **Reliability Availability:** The system shall ensure high availability with a target uptime of 99.9

5.4.3 Usability Requirements

Requirement: The system shall provide a user-friendly mobile interface with clear, intuitive navigation for users and caregivers, with easy-to-understand voice guidance for visually impaired users.

5.4.4 Reliability Availability

Requirement: The system shall ensure 99.9

5.4.5 Performance Requirements

– **Requirement:** The system shall process live feeds and generate object and face recognition results within 1 second.

5.4.6 User and System Requirements

User Requirements

Target Users:

– Visually impaired individuals relying on the smart glasses for real-time navigation assistance.

– Caregivers or family members assisting with setup and customization of the system.

– Technical support staff managing system maintenance and updates.

User Needs:

– **Real-time Assistance:** Users need immediate feedback on objects, and text for navigation and interaction with their surroundings.

– **User-Friendly Interface:** The system should offer a clear, intuitive interface with customizable settings for individual needs.

– **Notifications and Alerts:** Users need timely audio alerts for recognized objects, or important environmental changes.

– **Customization:** Users should be able to adjust settings for voice speed, language, and sensitivity of object recognition via the mobile application.

System Requirements

Hardware Requirements:

– Raspberry Pi 4 with a high-performance camera module to process real-time image data.

– Earphones or Bluetooth speakers for delivering audio feedback to users.

- Battery with at least 8 hours of continuous operation for the smart glasses.
- Ultrasonic sensors for detecting proximity to obstacles and alerting users.

Software Requirements:

- **Operating System:** Lightweight Linux-based OS for smart glasses, Android/iOS for the mobile app.
- **Programming Languages:** Python (for AI model processing), JavaScript (for mobile app frontend).
- **Frameworks:** TensorFlow or PyTorch for machine learning models, Flask/Django for backend services, OpenCV for image processing, Bluetooth libraries for wireless data transfer.

Chapter 6

Diagram Illustrations

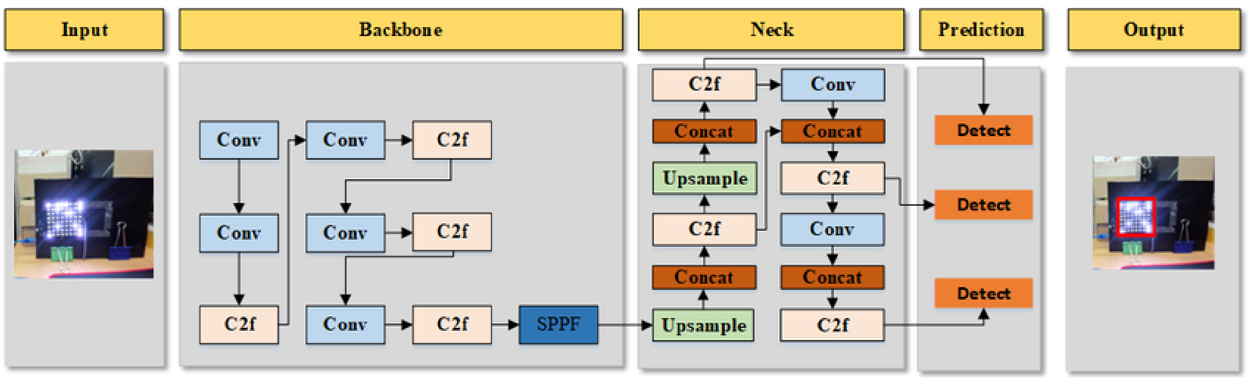

Figure 6.1: YOLOv8 Object Detection Architecture

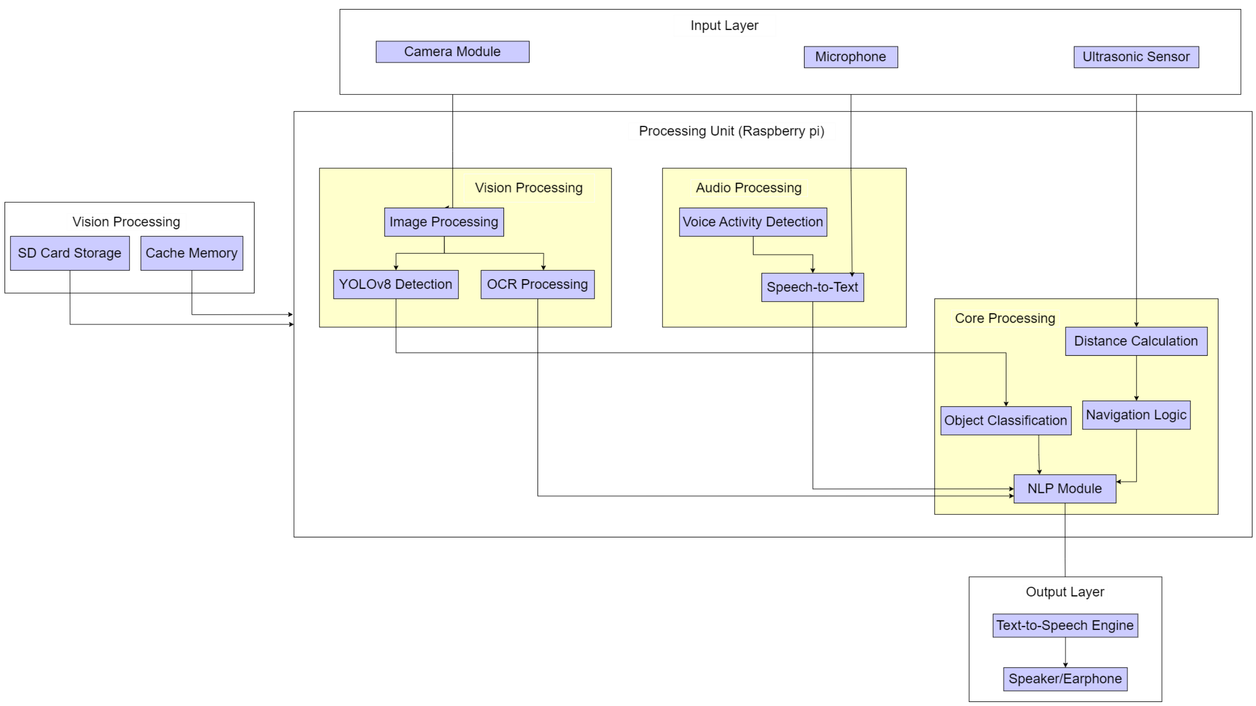

Figure 6.2: System Specification of YOLOInsight

Chapter 7

Project Requirement Specification

7.1 Introduction

This document outlines the project requirements for developing an AI-powered assistive smart glasses system for visually impaired individuals. The goal is to provide real-time object detection, facial recognition, and text reading through the integration of machine learning and IoT technology, enabling visually impaired users to navigate their environment independently with audio feedback. This system will enhance accessibility and improve the quality of life for visually impaired individuals.

7.2 Functional Requirements

- **Input Layer:**
  - **Camera Module:** Serves as the primary visual input device.
  - **Microphone:** Captures audio/voice input.
  - **Ultrasonic Sensor:** Likely used for distance measurement and obstacle detection.
- **Vision Processing System:**
- **External Components:**

  - SD Card Storage: For persistent data storage.
  - Cache Memory: For temporary, fast-access data storage.

- **Main Processing:**

  Processing module with:

  - YOLOv8 Detection: A modern real-time object detection system.
  - OCR Processing: Optical Character Recognition for text extraction from images.

- **Audio Processing Unit:**

  - Voice Activity Detection: Identifies presence of speech in audio input.
  - Speech-to-Text: Converts detected speech into text format.

- **Core Processing (Central Processing Unit):**

  - Distance Calculation: Processes ultrasonic sensor data.
  - Navigation Logic: Handles movement and direction decisions.
  - Object Classification: Categorizes detected objects.
  - NLP Module: Natural Language Processing for understanding and generating responses.

- **Output Layer:**

  - Text-to-Speech Engine: Converts system responses to audio.
  - Speaker/Earphone: Delivers audio output to the user.

7.3 Non-Functional Requirements

- **Performance:**

  - The system should process and classify images in real-time with less than 1 second latency.

- **Scalability:**

  - The architecture should support scaling to allow the addition of more object categories or users without degrading system performance.

- **Usability:**
  - The mobile application must be intuitive, with a simple interface for users and caregivers to configure and manage.

- **Security:**
  - Implement secure authentication for both the smart glasses and the mobile app to protect sensitive user data.
  - Ensure data transmission between the smart glasses and mobile app is encrypted.

7.4 Technical Requirements

- **Software Requirements:**
  - Programming languages: Python (for AI model).
  - Frameworks: TensorFlow/PyTorch (for machine learning), Flask/Django (for backend), OpenCV (for image processing).
  - Additional libraries: Hugging Face transformers for NLP tasks.

- **Hardware Requirements:**
  - Raspberry Pi 4 B.
  - A high-resolution camera module for capturing real-time images in various lighting conditions.
  - Earphones or a Bluetooth device for providing audio feedback.
  - Battery with a minimum of 8 hours of continuous usage.

- **Data Requirements:**
  - A labeled dataset of common objects, and texts to train the machine learning model for object detection and text recognition.

7.5 Testing Requirements

- **Unit Testing:**
  - Test individual components such as image acquisition, object detection, and audio feedback.
- **Integration Testing:**
  - Ensure seamless integration between the smart glasses, mobile app, and backend, from image capture to real-time feedback.
- **User Acceptance Testing (UAT):**
  - Gather feedback from potential end-users to refine functionality and improve the user experience.

7.6 Conclusion

This project requirement specification outlines the necessary functional and non-functional requirements needed to develop an AI-powered assistive solution for visually impaired individuals. By adhering to these specifications, the project aims to deliver a reliable tool that improves the independence and accessibility of visually impaired users in their daily lives.

Chapter 8

Proposed system Architecture

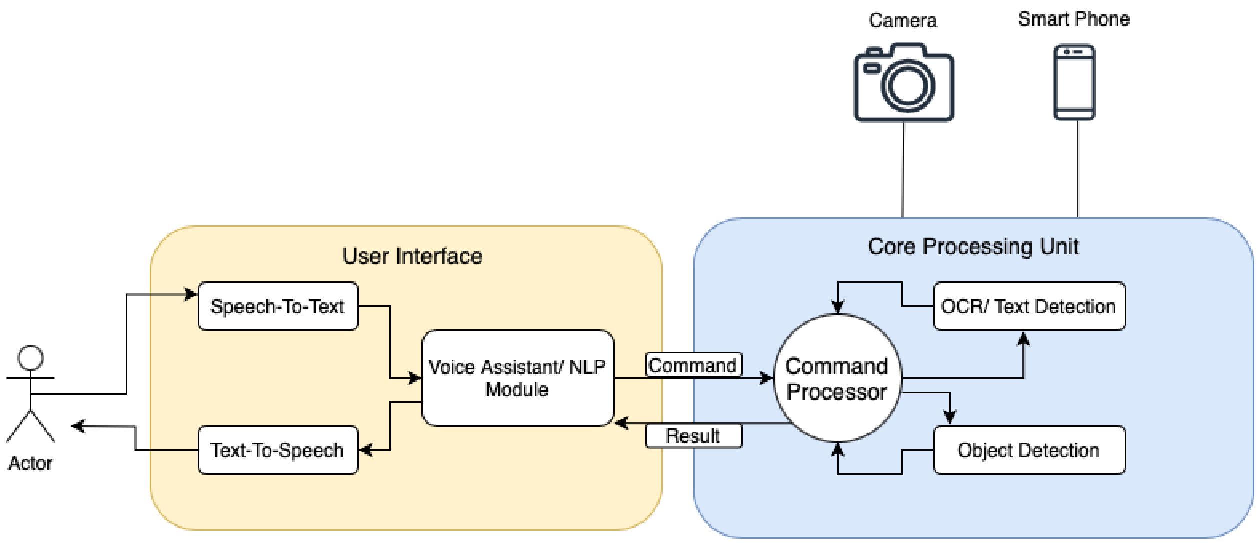

Figure 8.1: System Architecture Of YOLOInsight

8.1 Introduction

This document presents the high-level design of the YOLOInsight system, focusing on its two primary components: the User Interface and the Core Processing Unit. The system is designed to assist visually impaired individuals by providing real-time object detection, text recognition, and audio feedback through voice-based commands. The architecture ensures hands-free interaction, making the system accessible and intuitive for users.

8.2 Functional Requirements

This document presents the high-level design of the YOLOInsight system, focusing on its two primary components: the User Interface and the Core Processing Unit. The system is designed to assist visually impaired individuals by providing real-time object detection, text recognition, and audio feedback through voice-based commands. The architecture ensures hands-free interaction, making the system accessible and intuitive for users.

8.3 Functional Requirements

- **User Interface:**
  - **Speech-To-Text Module:** Captures user speech input and converts it into text, allowing users to issue commands verbally.
  - **Voice Assistant / NLP Module:** Processes natural language commands, interpreting user instructions and converting them into actionable commands for the system.
  - **Text-To-Speech Module:** Converts system outputs or results into speech, enabling the system to communicate with the user verbally.
- **Core Processing Unit:**
  - **Command Processor:** Receives commands from the Voice Assistant / NLP Module and distributes tasks to the appropriate processing components.
  - **OCR/Text Detection:** Recognizes and extracts text from images, providing users with audible information about textual content in their environment.
  - **Object Detection:** Identifies and categorizes objects in the environment, helping users understand their surroundings.
- **Workflow Summary:**

- The user initiates interaction via the User Interface, either through spoken commands or text.
- The Voice Assistant / NLP Module processes these commands and forwards them as instructions to the Command Processor.
- The Command Processor then routes the instructions to either OCR/Text Detection or Object Detection, depending on the task.
- Processed results are returned to the User Interface, which communicates the output back to the user via the Text-To-Speech Module.

8.4 Key Features of the Proposed System

- **Real-Time Object Classification:** Utilizes a camera module and YOLO-based deep learning algorithms deployed on a Raspberry Pi 4, providing accurate real-time object detection and classification.
- **Accessible Audio Output:** Processed information is converted into accessible audio, enabling users to receive real-time guidance and information about their surroundings.
- **Intuitive Interfaces:** The system incorporates adaptive technologies to ensure seamless navigation for visually impaired users.
- **Custom NLP API (Voice Assistant):** The system includes a proprietary NLP API, enabling fluid, conversational interactions. This custom-built assistant interprets user queries and provides context-aware responses, significantly enhancing accessibility.
- **Text-to-Speech Conversion:** Converts both detected handwritten and printed text into voice output, providing users with audible information about textual content.

8.5 Non-Functional Requirements

- **Performance:**

  - The system should process and classify images in real-time with less than 1 second latency.

- **Scalability:**
  - The architecture should support scaling to allow the addition of more object categories or users without degrading system performance.

- **Usability:**
  - The mobile application must be intuitive, with a simple interface for users and caregivers to configure and manage.

- **Security:**
  - Implement secure authentication for both the smart glasses and the mobile app to protect sensitive user data.
  - Ensure data transmission between the smart glasses and mobile app is encrypted.

8.6 Technical Requirements

- **Software Requirements:**
  - Programming languages: Python (for AI model).
  - Frameworks: TensorFlow/PyTorch (for machine learning), Flask/Django (for backend), OpenCV (for image processing).
  - Additional libraries: Hugging Face transformers for NLP tasks.

- **Hardware Requirements:**
  - Raspberry Pi 4 B.
  - A high-resolution camera module for capturing real-time images in various lighting conditions.
  - Earphones or a Bluetooth device for providing audio feedback.
  - Battery with a minimum of 8 hours of continuous usage.

- **Data Requirements:**
  - A labeled dataset of common objects, and texts to train the machine learning model for object detection and text recognition.

8.7 Testing Requirements

- **Unit Testing:**
  - Test individual components such as image acquisition, object detection, and audio feedback.
- **Integration Testing:**
  - Ensure seamless integration between the smart glasses, mobile app, and backend, from image capture to real-time feedback.
- **User Acceptance Testing (UAT):**
  - Gather feedback from potential end-users to refine functionality and improve the user experience.

8.8 Conclusion

This document outlines the proposed architecture for the YOLOInsight system, focusing on providing a highly accessible, real-time object detection and audio feedback solution for visually impaired users. By integrating advanced machine learning, computer vision, and NLP technologies, the system offers a seamless, hands-free experience for users navigating their environment.

Chapter 9

High level design of the project

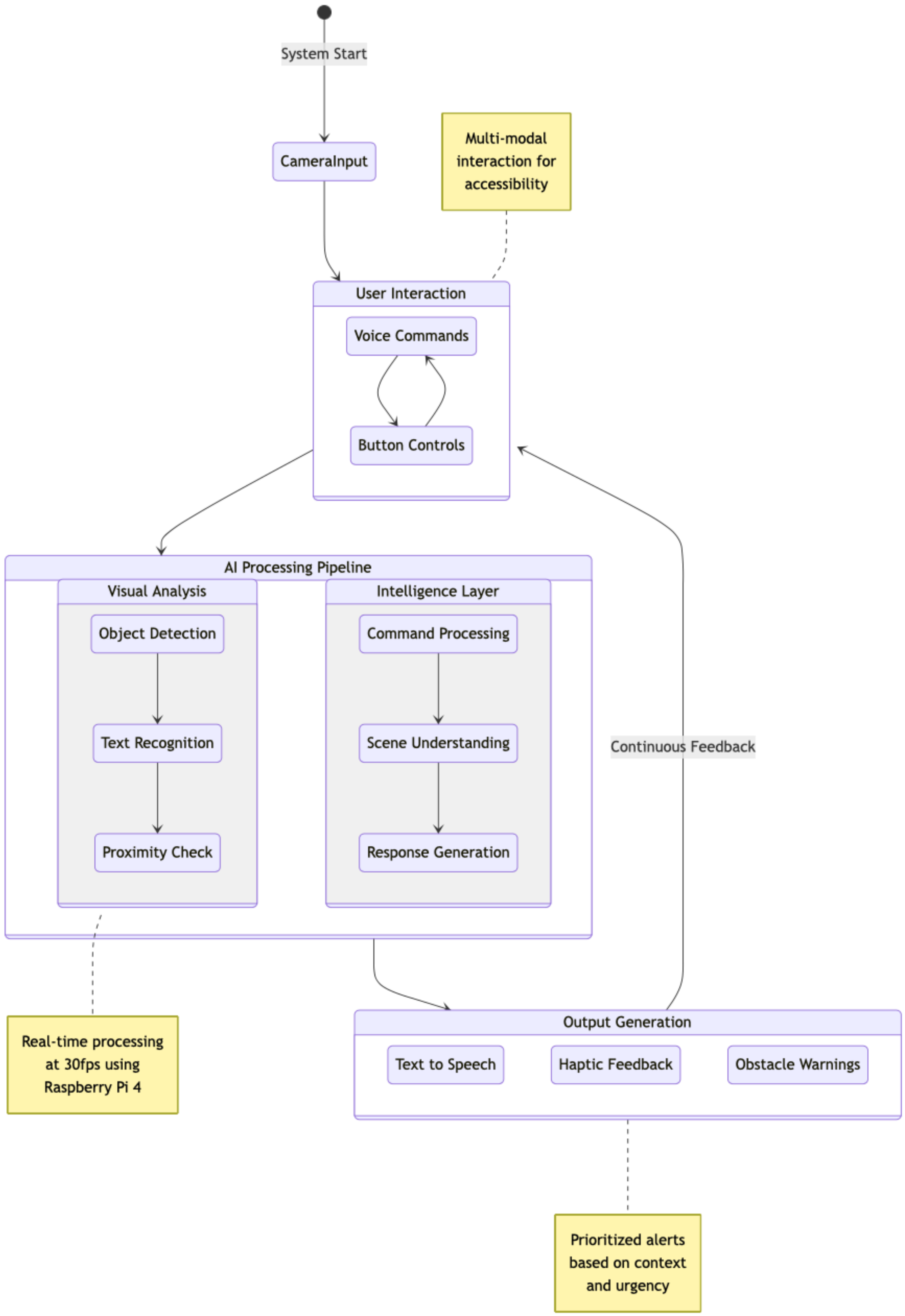

Figure 9.1: UML Diagram for Working of the System

9.1 Introduction

This diagram depicts a system for providing multi-modal interaction for accessibility, likely in the context of assistive technology or a smart home environment. The system uses a camera to capture visual input and processes it through various AI algorithms. It then generates responses in different modalities, such as text to speech, haptic feedback, and obstacle warnings. The system also incorporates continuous feedback mechanisms to improve its performance over time.

9.2 Input and User Interaction

- **System Start:** The process begins with a system start signal.
- **Camera Input:** The system receives visual information from a camera.
- **User Interaction:** The system allows for user interaction via voice commands, button controls, or possibly even other multi-modal interfaces (not explicitly shown, but implied by the 'Multi-modal interaction for accessibility' note). This interaction could be used to guide the system, provide context, or request specific actions.

9.3 AI Processing Pipeline

- **Visual Analysis:** The system analyzes the incoming visual data using a combination of AI algorithms.
  - **Object Detection:** Identifies objects within the scene.
  - **Text Recognition:** Reads any text present in the visual data.
  - **Intelligence Layer:** The system uses a layer of AI to interpret the scene and make decisions.
- **Command Processing:** Processes user commands received through the interaction methods, ensuring they are understood and relevant.

- **Scene Understanding:** Analyzes the environment based on the identified objects and text, potentially incorporating past knowledge or context.

- **Response Generation:** The system generates a response based on the processed information and user commands.

9.4 Output and Feedback

- **Output Generation:** The system presents the generated response to the user.
  - **Text to Speech:** Outputs information verbally.
  - **Haptic Feedback:** Provides tactile feedback (e.g., vibrations), potentially useful for navigation or indicating the presence of objects.
  - **Obstacle Warnings:** Provides alerts about potential obstacles in the environment.
  - **Prioritized Alerts:** The system prioritizes alerts based on context and urgency, ensuring that the most important information is delivered to the user.

- **Continuous Feedback:** The system continues to monitor the scene and receive user input, allowing for ongoing adjustments and adaptations.

9.5 Real-Time Processing

- The diagram emphasizes real-time processing, suggesting that the system is designed to handle visual data quickly, at a rate of 30 frames per second, utilizing the processing power of a Raspberry Pi 4.

9.6 Conclusion:

- This system illustrates a complex, AI-powered approach for understanding visual data and responding in real-time. The system is highly flexible and adaptable, capable of processing multiple modalities (e.g., visual, audio) and providing relevant feedback to the user.

Chapter 10

Experimental Results

10.1 Introduction

The intermediate evaluation of the YOLOInsight system's functionalities was conducted across several key metrics, with a focus on real-time object detection, text recognition, and natural language processing. The subsections below present the initial findings and performance assessments across these dimensions.

10.2 Model Accuracy Over Epochs

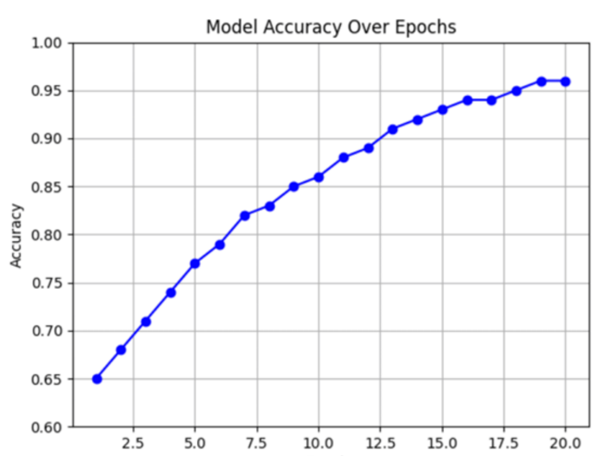

Figure 10.1: Model Accuracy Over Epochs

Training the YOLOv8 model on object detection tasks resulted in a steady increase in accuracy over 20 epochs, ultimately achieving a consistent accuracy of around 96% by the final epoch. This trend suggests effective model training and convergence, reflecting the model's suitability for real-time object classification tasks required for assistive applications.

10.3 Precision-Recall Curve

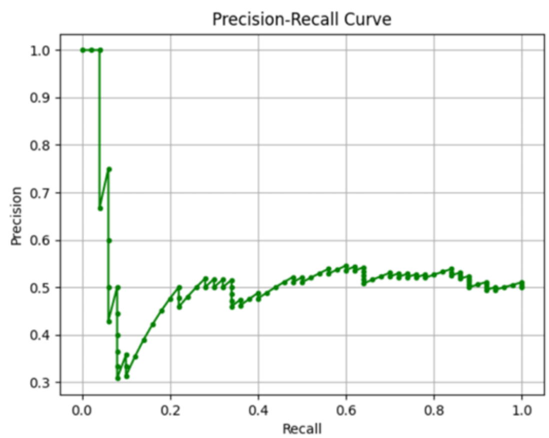

Figure 10.2: Precision-Recall Curve

Precision and recall values were assessed to measure detection reliability. The model demonstrated high precision across a range of recall values, as shown in Figure 10.2, achieving an average precision of over 90% at typical confidence thresholds. This balance indicates that the model minimizes false positives while maintaining sensitivity, crucial for accurately guiding visually impaired users.

10.4 Inference Time Distribution Across Object Classes

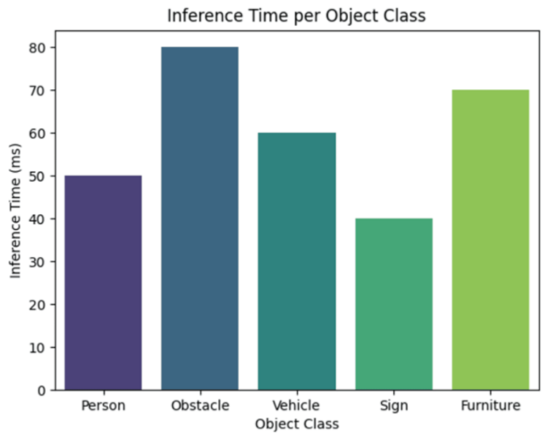

Figure 10.3: Inference Time Distribution Across Object Classes

The model's inference times were tested for various object categories commonly encountered by users, such as people, vehicles, and obstacles. The average inference time was well under 100 ms per object class, demonstrating efficient real-time performance. The faster processing time for key objects like people and obstacles further ensures responsive feedback for immediate navigation support.

10.5 NLP Intent Recognition Accuracy

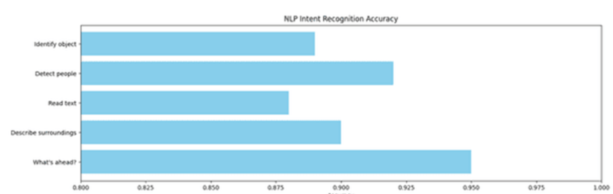

Figure 10.4: NLP Intent Recognition Accuracy

The NLP module was evaluated on its accuracy in recognizing common commands. Figure 10.4 shows that commands such as "What's ahead?" and "Describe surroundings" were recognized with over 90% accuracy, showcasing the effectiveness of the custom NLP API.

This high accuracy facilitates user interaction with minimal repetition, improving the usability of the system.

10.6 Object Detection Performance Across Lighting Conditions

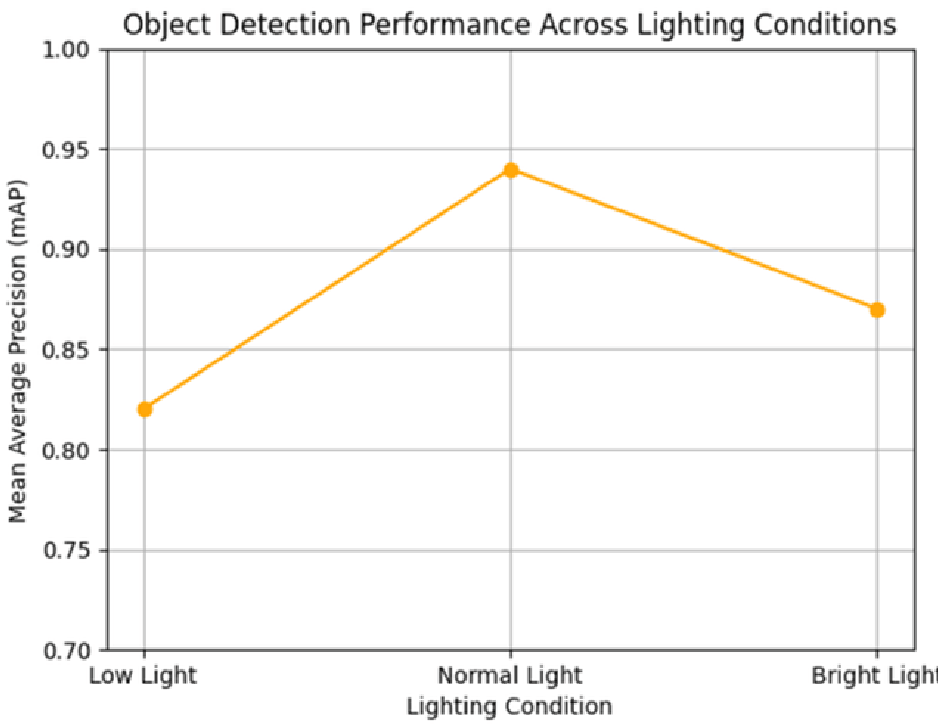

Figure 10.5: Object Detection Performance Across Lighting Conditions

Given that lighting can affect detection, tests were conducted in low, normal, and bright lighting conditions. Figure 10.5 illustrates the system's performance, with mAP remaining consistently high (above 0.80) even under challenging lighting. Normal lighting achieved the best results (mAP of 0.94), yet the model adapted well to low-light conditions, providing reliable assistance across diverse environments.

Chapter 11

Project Plan.

Sr. No	Task	Duration (approx.) in weeks
1	Analyze and Verify the Project Scope	1
2	Establish Project Goals	1
3	Prepare Synopsis.	1
4	Study of Algorithms in the Project with Hand Simulation.	3
5	Study of various Toolboxes.	2
6	Implementation of algorithm.	4
7	Comparison of Method results against each of the popular algorithms result.	2
8	Modification to work.	1
9	Paper publications	1
	Total Weeks	**16**

Table 11.1: Schedule of Project Stage-I Work

Chapter 12

Conclusions

The YOLOInsight project marks a significant advancement in assistive technology for visually impaired individuals by integrating cutting-edge AI solutions that facilitate independent navigation and enhance environmental awareness. Utilizing the YOLOv8 algorithm for real-time object detection alongside Optical Character Recognition (OCR) and a Natural Language Processing (NLP) API, YOLOInsight empowers users with immediate auditory feedback, fostering inclusivity and accessibility in daily environments. The project has successfully demonstrated its core functionalities, highlighting its potential to transform the lives of visually impaired individuals by promoting autonomy and improved accessibility. Looking ahead, the YOLOInsight system offers ample opportunities for enhancement and expansion. A key area for future development is the integration of cloud computing, which could considerably boost the system's processing capabilities. Leveraging cloud resources would enable the implementation of more complex algorithms and larger datasets, thereby improving object recognition accuracy and expanding the range of detectable objects. Additionally, incorporating cloud-based storage solutions could streamline user data management, enabling personalized settings and preferences to be accessed across multiple devices. Addressing security is another critical aspect of future development, as implementing robust security measures, such as end-to-end encryption for data transmission and cloud storage, will safeguard user data, ensuring privacy and fostering trust in the technology. This focus on security not only protects sensitive information but also encourages broader adoption of YOLOInsight by establishing confidence in the system's reliability and safety. As the YOLOInsight system continues to evolve, the integration of cloud capabilities and strengthened security measures will play a crucial role in advancing functionality, safeguarding user privacy, and delivering an increasingly effective assistive technology solution for visually impaired individuals.

References

[1] Waisberg, J. Ong, M. Masalkhi et al., “Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impair- ments,” Eye, vol. 38, pp. 1036–1038, 2024

[2] R.Swetha and S. PRIYA, “Yolov5 driven smart glasses for visually impaireds,” International Jour-nal of Science and Research Archive, vol. 12, pp. 353–374, 2024.

[3] Annapurna, T., Mamidoju, Sai Siddesh, Parthasaradhi, Gurram, Akash, Kancherdas, Alabdeli, Haider, B, Swathi, Jain, Alok, and Kumar, Ashwani, “An iot-based vision alert for blind using in-terdisciplinary approaches,” E3S Web of Conf., vol. 507, p. 01047, 2024.

[4] 4.Y.-L. Poy, S. Darmaraju, C.-H. Goh, and B.-H. Kwan, “Standalone smart glass system for the blind and visually impaired,” in 2024 IEEE 14th Symposium on Computer Applications Industrial Electronics (ISCAIE), 2024, pp. 239–244

[5] M. Badawi, A. N. Al Nagar, R. Mansour, R. Mansour, K. Ibrahim, K. Ibrahim, N. Hegazy, S. Elas-kary et al., “Smart bionic vision: An assistive device system for the visually impaired using arti-ficial intelligence,” International Journal of Telecommunications, vol. 4, no. 01, pp. 1–12, 2024.

[6] D. Raju, S. Pooja, P. G. Bhat, D. E M, and H. G C, “Smart glass for visually impaired people,” in International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Con-trol Engineering, no. 4, April 2023, pp. 44–47, iSO 3297:2007.

[7] M. Laad and V. Bahl, “Empowering differently abled individuals through iot: Smart glasses for the visually impaired people,” Iconic Research And Engineering Journals, vol. 6, no. 12, pp. 382–386, 2023.

[8] V. K. Sunidi, V. Varuni, C. V. Sujay, G. R. Suhas, and S. R. Swamy, “Design of smart goggles for visually impaired with audio features,” International Journal of Emerging Technologies and In-novative Research, vol. 10, no. 5, pp. K94–K99, May 2023, iSSN: 2349-5162

[9] S. Sajini and B. Pushpa, “A binary object detection pattern model to assist the visually impaired in detecting normal and camouflaged faces,” Engineering, Technology amp; Applied Science Re-search, vol. 14, no. 1, p. 12716–12721, Feb. 2024.

[10] A. Abdulatif, S. Mohammed, A. Mohammed, H. Mohammed, and D. A. Ra- jalingam, “Smart glasses robot for blind people using raspberry-pi and python,” Journal of Science, Computing and Engineering Research, vol. 3, no. 1, pp. 237–244, 2022.

[11] B.Sudharani, K. Lokanath, G. Bhavya, K. H. Priya, K. Venkat, and G. R. S. Chandu, “Smart glasses for visually challenged people using facial recognition,” International Journal of Innovative Research in Science Engineering, vol. 8, 2022.

[12] S.F. Fatima, T. Hussain, S. Anjum, S.A. Nadeem, and S. Fatima, “Innovation in transparent tech-nology with a deep dive into IoT-connected smart glass,” International Journal for Multidiscipli-nary Research, vol. 6, no. 3, p. 1, May-Jun. 2024.

[13] T.G. Ramya Priyatharsini, A. Arsath Arif, A. Asfack Ahamed, P. Kesavan, and K. Senguttuvan, “Smart glass for visually impaired people with facial recognition using IoT and machine learn-ing,” International Journal of Research Publication and Reviews, vol. 5, no. 5, pp. 12010-12013, May 2024.

[14] Jayashree C, Moohana Priya P, Monika K,"Smart eye for Visually Impaired-An aid to help the blind people”,21-23 February 2019,Chennai,India

[15] Akash Bhuiyan , “An Assistance System for Visually Challenged People Based on Computer Vi-sion and IOT” , 5-7 June 2020, Dhaka, Bangladesh

Appendix

A. Plagiarism Report:

Smodin

Plagiarism Share Report

Date: 2024-11-12

8% Plagiarism

Input text

The device is equipped with a total of four USB ports—two USB 3.0 ports for high- speed data transfer and two USB 2.0 ports for connecting peripherals. Networking capabilities are enhanced with a Gigabit Ethernet port, and it also supports wirelessconnectivity through 802.11b/g/n/ac Wi-Fi and Bluetooth 5.0.

Powering the Raspberry Pi requires a stable 5V DC supply via a USB-C connector, with a minimum current requirement of 3A to ensure optimal performance. The GPIO (General Purpose Input/Output) header, comprising 40 pins, allows for extensive interfacing with various sensors and components. The operating temperature range is between 0°C and 50°C, ensuring reliability in diverse environments. The compact dimensions of 89.6 mm x 56.5 mm x 16.8 mm make it an ideal choice for portable applications.

2)▭ Camera Module: For real-time visual input, the device incorporates a high- resolution Camera Module featuring the Sony IMX219 PQ CMOS image sensor. With a resolution of 8 megapixels (3280 x 2464 pixels), this camera is capable of capturing detailed images and video. The lens is fixed-focus with an approximate 54° diagonal field of view, allowing it to cover a broad area in front of the user.

Addressing security is another critical aspect of future development, as implementing robust security measures, such as end-to-end encryption for data transmission and cloud storage, will safeguard user data, ensuring privacy and fostering trust in the technology. This focus on security not only protects sensitive information but also encourages broader adoption of YOLOInsight by establishing confidence in the system's reliability and safety. As the YOLOInsight system continues to evolve, the integration of cloud capabilities and strengthened security measures will play a crucial role in advancing functionality, safeguarding user privacy, and delivering an increasingly effective assistive technology solution for visually impaired individuals.

1)▭ Raspberry Pi 4 Model B: The central processing unit of our assistive device is the Raspberry Pi 4 Model B, equipped with a Broadcom BCM2711 System on Chip

(SoC). This powerful processor features a quad-core Cortex-A72 architecture, offering clock speeds of 1.5 GHz in earlier models and up to 1.8 GHz in later iterations, providing robust performance for complex computational tasks. The device supports various RAM options, including 1GB, 2GB, 4GB, and 8GB of LPDDR4 memory, allowing for flexibility based on application requirements.

The Raspberry Pi 4 Model B comes with a microSD card slot for storage, enabling ample space for operating systems and applications. For video output,

Addressing security is another critical aspect of future development, as implementing robust security measures, such as end-to-end encryption for data transmission and cloud storage, will safeguard user data, ensuring privacy and fostering trust in the technology. This focus on security not only protects sensitive information but also encourages broader adoption of YOLOInsight by establishing confidence in the system's reliability and safety. As the YOLOInsight system continues to evolve, the integration of cloud capabilities and strengthened security measures will play a crucial role in advancing functionality, safeguarding user privacy, and delivering an increasingly effective assistive technology solution for visually impaired individuals.

1)▭ Raspberry Pi 4 Model B: The central processing unit of our assistive device is the Raspberry Pi 4 Model B, equipped with a Broadcom BCM2711 System on Chip

(SoC). This powerful processor features a quad-core Cortex-A72 architecture, offering clock speeds of 1.5 GHz in earlier models and up to 1.8 GHz in later iterations, providing robust performance for complex computational tasks. The device supports various RAM options, including 1GB, 2GB, 4GB, and 8GB of

LPDDR4 memory, allowing for flexibility based on application requirements.

The Raspberry Pi 4 Model B comes with a microSD card slot for storage, enabling ample space for operating systems and applications. For video output,

The YOLOInsight project marks a significant advancement in assistive technology for visually impaired individuals by integrating cutting-edge AI solutions that facilitate independent navigation and enhance environmental awareness. Utilizing the YOLOv8 algorithm for real-time object detection alongside Optical Character Recognition (OCR) and a Natural Language Processing (NLP) API, YOLOInsight empowers users with immediate auditory feedback, fostering inclusivity and accessibility in daily environments. The project has successfully demonstrated its core functionalities, highlighting its potential to transform the lives of visually impaired individuals by promoting autonomy and improved accessibility.

Looking ahead, the YOLOInsight system offers ample opportunities for enhancement and expansion. A key area for future development is the integration of cloud computing, which could considerably boost the system's processing capabilities.

Leveraging cloud resources would enable the implementation of more complex algorithms and larger datasets, thereby improving object recognition accuracy and expanding the range of detectable objects. Additionally, incorporating cloud-based storage solutions could streamline user data management, enabling personalized settings and preferences to be accessed across multiple devices.

1)□ Raspberry Pi 4 Model B: The central processing unit of our assistive device is the Raspberry Pi 4 Model B, equipped with a Broadcom BCM2711 System on Chip

(SoC). This powerful processor features a quad-core Cortex-A72 architecture, offering clock speeds of 1.5 GHz in earlier models and up to 1.8 GHz in later iterations, providing robust performance for complex computational tasks. The device supports various RAM options, including 1GB, 2GB, 4GB, and 8GB of LPDDR4 memory, allowing for flexibility based on application requirements.

The Raspberry Pi 4 Model B comes with a microSD card slot for storage, enabling ample space for operating systems and applications. For video output, it features two micro HDMI ports, supporting resolutions up to 4K at 60 frames per second, making it suitable for high-quality visual processing.

The device is equipped with a total of four USB ports—two USB 3.0 ports for high- speed data transfer and two USB 2.0 ports for connecting peripherals. Networking capabilities are enhanced with a Gigabit Ethernet port, and it also supports wireless connectivity through 802.11b/g/n/ac Wi-Fi and Bluetooth 5.0.

Powering the Raspberry Pi requires a stable 5V DC supply via a USB-C connector, with a minimum current requirement of 3A to ensure optimal performance. The GPIO (General Purpose Input/Output) header, comprising 40 pins, allows for extensive interfacing with various sensors and components.

T. Annapurna (2024) [3] propose an IoT-based Vision Alert system for the blind, utilizing wearable glasses equipped with cameras and IoT devices. The glasses employ advanced sensors to recognize obstacles and provide real-time navigation assistance. The study highlights the Recent advancements in obstacle detection and human recognition that have been achieved using ultrasonic sensors and CNN algorithms with the ESP32 microcontroller. This approach provides real-time feedback through a buzzer, significantly enhancing spatial awareness and enabling independent navigation.

The literature review also encompasses studies Poy(2024) [4] that proposed a smart glasses system as a standalone solution, eliminating the need for constant internet connectivity or reliance on external devices such as smartphones. It features three core modules: object recognition with YOLOV8, text recognition with KerasOCR, and face recognition with FaceNet. The system utilizes a Raspberry Pi Zero 2 W and an Intel Neural Compute Stick 2 for efficient edge-based deep learning model execution. Lightweight smart glasses system aims to empower individuals with visual impairments, promoting self-reliance and emphasizing capability over disability

This project presents an innovative AI-powered real- time assistance system specifically designed for visually impaired individuals, aimed at enriching their real- world experiences and fostering inclusivity in various environments. By leveraging state- of-the-art artificial intelligence algorithms the system offers personalized assistance, navigation support, and seamless interaction for users facing visual challenges. Through intuitive, user-friendly interfaces and adaptive technologies, the solution empowers individuals with visual impairments to navigate spaces independently and with greater confidence. At its core, the system integrates a camera module and YOLO-based deep learning algorithms running on a cost-effective Raspberry Pi 4, enabling real-time object detection and classification.

Processed information is converted into accessible audio output, significantly reducing the cost compared to existing solutions without compromising on functionality. To further enhance the user experience, the system incorporates language customization via Optical Character Recognition (OCR) and Google Text-to-Speech (gTTS) technology, allowing audio feedback in multiple languages based on user preferences. Additionally, the project includes a built-in Natural Language Processing (NLP) API, akin to Siri or Google Assistant, but without relying on third- party services. This approach ensures complete control over the system’s features, enhances user privacy, and further reduces costs.

Source (1)

URL: https://brainy.co.ke/collections/vention-type-c-docking-station?filter.v.availability=1&page=2&srsltid=AfmBOorXZuYIEViOOXjBtdrPvuMeIWNCpFrMrgbaQCBEByMV5Cj6gdB_

Text matches

- USB 3.0 ports for high-speed data...

Source (2)

URL: https://www.connectingup.org/products/category/Computer%20Accessories%2C%20Parts%20%26%20Cables?items_per_page=20&sort_by=search_api_aggregation_3

Text matches

- making it an ideal choice for

Source (3)

URL: https://www.canakit.com/raspberry-pi-4-8gb.html?srsltid=AfmBOoqic7vUNag8ZgvoY2l0h9vj8CNLdIpUG3pqFoDzbEng_4OxoHe\
Text matches

- The Raspberry Pi 4 Model B
- The Raspberry Pi 4 Model B comes

Source (4)

URL: https://www.raspberrypi.com/documentation/computers/processors.html

Text matches

- the Raspberry Pi 4 Model B, the Raspberry Pi

Source (5)

URL: https://www.researchgate.net/figure/Raspberry-Pi-4-Model-B_fig1_373347761

Text matches

- the Raspberry Pi 4 Model B,

Source (6)

URL: https://www.adlibsoftware.com/news/here-is-why-enterprise-grade-optical-character-recognition-ocr-and-natural-language-processing-nlp-are-critical-for-your-organization

Text matches

- Optical Character Recognition (OCR) and Natural Language Processing (NLP)

Source (7)

URL: https://en.wikipedia.org/wiki/Raspberry_Pi

Text Matches

- "Raspberry Pi 3 Model B+". / ^ "Raspberry Pi 4 Model B

Source (8)

URL: https://www.yodeck.com/news/the-raspberry-pi-4-ultimate-review/

Text Matches

- Raspberry Pi Model B Raspberry Pi 4 Model B

Source (9)

URL: https://www.seeedstudio.com/blog/2019/09/29/top-20-best-raspberry-pi-4-projects-that-you-must-try-now/?srsltid=AfmBOopDqhZnTjVQunI3plN-IQQppsJnOwSwg_PT3lOg9RYLDnMYeR45

Text Matches

- Raspberry Pi , Raspberry Pi 4 , Raspberry Pi 4 Model B , Raspberry Pi 4 Projects , Raspberry Pi 4b , Raspberry Pi

---

Source (10)

URL: https://www.pishop.us/product/raspberry-pi-4-model-b-2gb/

Text Matches

- Raspberry Pi Raspberry Pi Boards Raspberry Pi Boards Raspberry Pi 4 Model

---

Source (11)

URL: https://www.e3s-conferences.org/articles/e3sconf/abs/2024/37/e3sconf_icftest2024_01047/e3sconf_icftest2024_01047.html

Text Matches

- An IoT-based vision alert for blind
- utilizes wearable glasses equipped with cameras and IoT devices
- glasses employ advanced sensors to recognize obstacles and provide real-time navigation assistance,

---

Source (12)

URL: https://www.researchgate.net/figure/Model-of-glasses-clip_fig2_362177987

Text Matches

- glasses employ advanced sensors to recognize obstacles and provide real-time navigation assistance,

---

Source (13)

URL: https://ijsra.net/sites/default/files/IJSRA-2024-0804.pdf

Text Matches

- a camera module and YOLO-based deep learning algorithms on a Raspberry Pi 4 for real-time object

---

Source (14)

URL: https://medium.com/@jamesasantana/ai-for-disability-assistance-new-ai-tools-empower-visually-impaired-individuals-turning-944d1e539325

Text Matches

- empower individuals with visual impairments to navigate

B. Base Paper(s):

[1] Waisberg, J. Ong, M. Masalkhi et al., “Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impair- ments,” Eye, vol. 38, pp. 1036–1038, 2024

[2]R.Swetha and S. PRIYA, “Yolov5 driven smart glasses for visually impaireds,” International Jour-nal of Science and Research Archive, vol. 12, pp. 353–374, 2024.

[3]Annapurna, T., Mamidoju, Sai Siddesh, Parthasaradhi, Gurram, Akash, Kancherdas, Alabdeli, Haider, B, Swathi, Jain, Alok, and Kumar, Ashwani, “An iot-based vision alert for blind using in-terdisciplinary approaches,” E3S Web of Conf., vol. 507, p. 01047, 2024.

C.System Components

This section outlines the essential hardware and software components required for developing an assistive device for visually impaired individuals. Each component is carefully selected to ensure optimal performance and reliability.

12.1 Hardware Components

- **Raspberry Pi 4 Model B:**
  - Processor: Broadcom BCM2711 System on Chip (SoC) with quad-core Cortex-A72, clocked at up to 1.8 GHz.
  - RAM Options: Available in 1GB, 2GB, 4GB, and 8GB LPDDR4.
  - Storage: MicroSD card slot for OS and application storage.
  - USB Ports: Two USB 3.0 ports and two USB 2.0 ports.
  - Networking: Gigabit Ethernet, 802.11b/g/n/ac Wi-Fi, and Bluetooth 5.0.
- Camera Module:
  - Sensor: Sony IMX219 PQ CMOS image sensor with 8 MP resolution.
  - Field of View: Approximately 54° diagonal.
  - Connection: CSI-2 interface with 15-pin ribbon cable.
  - Frame Rates: 1080p at 30 fps, 720p at 60 fps, 640x480p at 60/90 fps.
- **Ultrasonic Sensor (HC-SR04):**
  - Operating Voltage: 5V DC.
  - Current Draw: Less than 15 mA.
  - Measuring Distance: 2 cm to 400 cm (optimal within 2 cm to 80 cm).
- **Push Button (Tactile):**
  - Type: 4-legged SPST switch.
  - Operating Force: Approximately 100 grams.

  - Current Rating: 50 mA.

- **Additional Components:**

  MicroSD Card:

  - Type: 32GB SanDisk Micro SDHC Class 4.

  Earphones:

  - Features: 40mm dynamic drivers, noise cancellation, up to 30 hours battery life.

- **USB Cable:** For power and connectivity.

- **GPIO Connection Wires:** For interfacing with sensors and components.

12.2 Software Tools

- **Programming Language:**
  - Python 3.8+: Primary language for developing scripts and machine learning models.
- **Machine Learning Libraries:**
  - PyTorch: For building and training neural networks with YOLOv5 model.
  - OpenCV: Used for image processing, object detection, and real-time frame manipulation.
- **Optical Character Recognition:**
  - PyTesseract: A Python wrapper for Google's Tesseract-OCR engine to enable text recognition from images.
- **Text-to-Speech Libraries:**
  - gTTS: Converts text to speech using Google Translate's API.
  - pyttsx3: Provides offline text-to-speech synthesis.

- **Audio Processing:**
  - Pygame: For audio playback, integrated with text-to-speech outputs.
- **Automatic Speech Recognition:**
  - OpenAI Whisper: For high-fidelity transcription and translation of audio to text.
- **Natural Language Processing:**

  Libraries:
  - nltk: For stopword removal and lemmatization.
  - sklearn: Includes TfidfVectorizer and cosine similarity functions for intent identification.
  - json: Manages JSON data for predefined intents and examples.
  - joblib: Saves and loads the trained TfidfVectorizer model.

Object Detection:

- YOLOv8: For fast and accurate object detection in images and video streams.

12.3 Development Environment

- **IDE/Text Editor:**
  - Jupyter Notebook, VS Code: For writing and testing code efficiently.

This configuration of hardware and software components aims to build a robust assistive device that enables real-time object detection, provides audio feedback, and facilitates environmental awareness, enhancing the navigation and interaction capabilities of visually impaired users.

D. Published Papers and Certificates:

Journals

1. Cureus Journal of Computer Science (Springer Nature). **Status: Under Review Process.**
2. International Journal For Multidisciplinary Research (IJFMR). **Status: Under Review Process.**

E. Sponsorship Details:

Enumtech Technology Solutions LLP is providing invaluable resources to support the YOLO Insight: AI-Powered Assistive Glasses project. Their assistance covers several critical areas to ensure the project's successful completion.

12.4 Key Areas of Assistance

- **Technical Guidance:**
  - Details: Enumtech's team of experts will offer specialized knowledge in AI, machine learning, and hardware integration, ensuring the project aligns with industry standards.
- **Data Access:**
  - Details: The company will provide access to datasets essential for fine-tuning the project's object detection and text recognition capabilities.
- **Facilities and Equipment:**
  - Details: Enumtech will grant access to facilities and testing environments to validate hardware functionality in real-world conditions.
- **Hardware and Software Resources:**
  - Details: Enumtech will supply critical hardware components and software tools to support development, testing, and deployment.

These resources are vital for the project's successful completion, and we are grateful for Enumtech's generous support in realizing this innovative assistive technology.

ENUMTECH
TECHNOLOGY SOLUTIONS LLP
(Formerly ENUMTECH)

info@enumtech.in

Corporate office: 104, West Wing Arora Towers
MG Road Pune, Maharashtra 411001

+91 8237307229

LLPIN:ABB-5307

ENUMTECH
Let IT be

Date: 21st Oct 2024
Letter ID: 2024/internship/41

To,
Mr. Gajanan Arsalwad
Assistant Professor, Information Technology,
Trinity College of Engineering & Research
Pune, Maharashtra

Subject: Confirmation of Sponsorship and Collaboration for Final Year Project

Respected Sir,

I am writing to formally confirm that Enumtech Technology Solutions LLP is pleased to sponsor and collaborate on the final-year project focused on YOLO Insight: AI-Powered Assistive Glasses proposed by Mr. Kabir Shaikh, Mr. Saurabh Dabhade and Mr. Sean D'silva under your guidance.

We are committed to providing the necessary technical support, and access to relevant data and facilities, to ensure the successful completion of this project. I look forward to contributing to the student's success through this collaboration. Please feel free to reach out to me if there are any additional requirements or if you would like to discuss the project further.

Thank you for this opportunity to collaborate, and I look forward to working closely with you and the students.

For ENUMTECH TECHNOLOGY SOLUTIONS LLP.

U Y Solanki

Managing Director