Cureus Journal of Computer Science

Open Access Original Article

YOLOInsight: Artificial Intelligence-Powered Assistive Device for Visually Impaired Using Internet of Things and Real-Time Object Detection

Received 11/18/2024
Review began 11/18/2024
Review ended 12/25/2024
Published 12/26/2024

DOI: https://doi.org/10.7759/s44389-024-02353-2

Gajanan Arsalwad [1] , Saurabh Dabhade [1] , Kabir Shaikh [1] , Sean D'silva [1]

1. Information Technology, KJ's Educational Institute Trinity College of Engineering and Research, Pune, IND

Corresponding authors: Gajanan Arsalwad, gajananarsalwad.tcoer@kjei.edu.in, Saurabh Dabhade, saurabhdabhade2002@gmail.com, Kabir Shaikh, 1234kabirshaikh@gmail.com, Sean D'silva, dsilvasean1@gmail.com

Abstract

This project presents an innovative artificial intelligence (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 AI algorithms, the system offers personalized assistance, navigation support, and seamless interaction for users with 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 functionality. To further enhance the user experience, the system incorporates language customization via optical character recognition and Google Text-to-speech technology, allowing audio feedback in multiple languages based on user preferences. Additionally, the project includes a built-in natural language processing application programming interface, 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. By prioritizing the specific needs of visually impaired individuals, this innovative system aims to improve accessibility, foster independence, and offer a more inclusive experience in environments like shopping centers and public spaces. Overall, the system achieved 96% object detection accuracy, 91-95% intent recognition accuracy, and an average response latency of 650 ms, demonstrating its feasibility as a low-cost assistive technology for visually impaired users.

Categories: Embedded Systems, IoT Applications, Deep Learning
Keywords: assistive technology, visually impaired, object detection, yolov8, raspberry pi, natural language processing (nlp), optical character recognition (ocr), real-time navigation, audio feedback, machine learning

Introduction

In an age where technology evolves rapidly, the goal of true inclusivity and accessibility is paramount. This project introduces YOLOInsight, an artificial intelligence (AI)-powered assistive system specifically designed to empower visually impaired individuals by providing greater autonomy in navigating their surroundings. Built on advanced AI, YOLOInsight combines real-time object detection, personalized navigation, and a custom natural language processing (NLP) application programming interface (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 are instantly converted to accessible audio output, tailored through optical character recognition (OCR) and Google Text-to-Speech (gTTS) technologies to support multiple languages, thereby enhancing usability. This project develops a real-time AI system to assist visually impaired individuals. It uses a camera module and YOLOv8 algorithms on a Raspberry Pi 4 Model B for object detection and classification. The information is converted into audio feedback using OCR and gTTS, offering multilingual support. A built-in NLP API (voice assistant) enhances privacy and affordability, reflecting advancements in similar technologies, such as the Meta smart glasses outlined by Waisberg et al. [1], the AI-based shopping assistance system by Sweatha and Sathiya Priya [2], and the internet of things (IoT)-based vision alert system proposed by Annapurna et al. [3].

The smart glass system for the blind and visually impaired (BVI) features object, text, and face recognition using YOLOv8, KerasOCR, and FaceNet. Powered by a Raspberry Pi Zero 2 W and Intel Neural Compute Stick 2, it operates offline with speech recognition via Vosk-API, offering audio feedback. This lightweight, standalone device promotes independence for the BVI community [4].

How to cite this article
Arsalwad G, Dabhade S, Shaikh K, et al. (December 26, 2024) YOLOInsight: Artificial Intelligence-Powered Assistive Device for Visually Impaired Using Internet of Things and Real-Time Object Detection. Cureus J Comput Sci 1 : es44389-024-02353-2. DOI https://doi.org/10.7759/s44389-024-02353-2

Building upon pioneering work in smart bionic vision systems [5] and OCR integration for smart glasses [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 and Bahl [7] and Sujay et al. [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 and Pushpa [9] and Abdulatif et al. [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 et al. [11] and Sujay et al. [8].

YOLOInsight prioritizes auditory feedback as the primary mode of interaction, aligning with the design philosophy presented by Laad and Bahl [7] and Sajini and Pushpa [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 user needs within a single device. Furthermore, this project aligns with Abdulatif et al. [10] and Laad and Bahl [7], who advocate for low-cost, power-efficient solutions that maximize independence for visually impaired users.

The project's innovations extend beyond object detection, incorporating insights from research such as food detection with image processing using convolutional neural networks (CNN) [12], which highlights the effectiveness of CNN in image classification tasks. Similarly, studies on OCR-based solutions (e.g., Abhinav et al. [13]) and text-to-speech conversion (Tesseract-OCR) inspire YOLOInsight's design for text recognition and conversion. The processed text is relayed to users through auditory feedback, ensuring seamless interaction and usability.

Further building on foundational research in smart glasses, such as smart glasses using deep learning and stereo camera by Kim et al. [14], YOLOInsight incorporates ultrasonic sensors for enhanced spatial awareness and obstacle detection. These smart glasses showcased the utility of YOLOv3 algorithms for obstacle recognition and provided real-time location and type classification of obstacles, highlighting the benefits of multi-modal sensing 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.

Materials And Methods

Hardware components

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. 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 5 V DC supply via a USB-C connector, with a minimum current requirement of 3 A 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 × 56.5 mm × 16.8 mm make it an ideal choice for portable applications.

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 × 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. The camera connects via the CSI-2 (MIPI

Camera Serial Interface) using a 15-pin ribbon cable, ensuring a seamless integration with the Raspberry Pi. It supports various frame rates, including 1080p at 30 frames per second, 720p at 60 frames per second, and 640 × 480p at both 60 and 90 frames per second. Video modes include H.264 and MJPEG, providing flexibility for different applications, while still capture formats include JPEG and BMP. The compact dimensions (23.86 mm × 25 mm × 9 mm) and lightweight design (only 3g) allow for easy integration into portable devices. Importantly, this camera module is compatible with all Raspberry Pi models, including versions 1, 2, 3, and 4.

Ultrasonic Sensor (HC-SR04)

To enhance environmental sensing capabilities, the device utilizes the HC-SR04 Ultrasonic Sensor. Operating at a voltage of 5 V DC and drawing less than 15 mA of current, this sensor employs a working frequency of 40 Hz to measure distances with high accuracy. The theoretical measuring distance ranges from 2 cm to 400 cm; however, practical applications typically yield reliable readings within 2 cm to 80 cm. This sensor provides critical spatial awareness for visually impaired individuals, allowing the device to detect obstacles and navigate safely.

Push Button (Tactile)

The assistive device includes a tactile push button configured as a four-legged, single-pole single-throw switch. This button is designed to withstand an operating force of approximately 100 g, ensuring ease of use. Typically operating at 12 V DC and capable of handling a maximum current rating of 50 mA, it boasts a contact resistance of less than 100 mΩ, ensuring reliable activation. With an insulation resistance greater than 100 MΩ (at 250 V DC) and an operational temperature range of -25°C to +70°C, this button is designed for durability and resilience. It has a life expectancy of 1,00,000 to 5,00,000 cycles, making it suitable for frequent use in assistive technologies. Its compact size (typically 6 mm × 6 mm × 5 mm) allows for easy integration into various interfaces, and it is mounted using through-hole technology.

Power Supply/Battery

The device requires a stable power supply of 5 V, with a current capacity of 3 A, to ensure all components operate efficiently. While the specific type of power supply is not detailed, it is crucial to provide a consistent power source for optimal performance.

Additional Components

MicroSD card: A 32GB SanDisk Micro SDHC Class 4 card is utilized for storage, ensuring sufficient space for the operating system and application files.

Earphones: The device includes earphones featuring 40 mm dynamic drivers, offering up to 30 h of battery life with noise-canceling capabilities. Adaptive noise cancelation technology allows for ambient sound control, enhancing the user's experience in various environments.

USB cable: A USB cable is included for power and connectivity, allowing the Raspberry Pi to interface with other devices and peripherals.

GPIO connection wires: These wires facilitate connections between the Raspberry Pi and various components, enabling the integration of sensors, buttons, and other devices into the assistive system.

This hardware configuration is meticulously designed to create a portable, efficient, and reliable assistive device for visually impaired individuals. It enables real-time object detection, provides audio feedback, and senses the surrounding environment, ultimately enhancing the user's navigation and interaction capabilities.

Software components

Python 3.8+

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

PyTorch

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

OpenCV

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

OCR

OCR is the process of converting images of typed, handwritten, or printed text into machine-encoded text. In this project, OCR functionality detects and recognizes text from images captured by the webcam, allowing for text extraction from images. PyTesseract, a Python wrapper for Google's Tesseract-OCR engine, enables OCR capabilities, supporting multi-language recognition and easy integration with libraries like Pillow (PIL) and OpenCV.

gTTS

gTTS is a Python library that interfaces with Google Translate's text-to-speech API, enabling text conversion to spoken audio in various languages. In this project, gTTS converts detected text (in languages like Tamil or English) into speech for auditory feedback.

pyttsx3

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 this project, pyttsx3 is used for speech synthesis to offer auditory feedback. Pygame is used alongside pyttsx3 to handle audio playback by initializing the audio mixer and playing the synthesized speech generated by gTTS.

Pygame

Pygame is a set of Python modules designed for writing video games, providing functionality for graphics, sound, and user input. In this project, Pygame is used for audio playback, playing the speech synthesized by gTTS or pyttsx3.

OpenAI Whisper

OpenAI Whisper is an advanced automatic speech recognition system that transcribes and translates audio into text. Known for its accuracy, Whisper handles various accents, background noise, and multiple languages, making it ideal for transcribing spoken language with high fidelity.

NLP

The NLP module is the core module used for interaction with smart glasses. It is used to process what the user is trying to say and map their commands to appropriate functionalities. It achieves this through the use of various libraries for preprocessing the text, vectorizing it, calculating similarity, and finally determining the corresponding function.

Natural Language Toolkit: Natural Language Toolkit is employed for various natural language processing tasks. Specifically, it helps remove stop words using the stopwords module and performs text lemmatization with WordNetLemmatizer.

Scikit-learn: Scikit-learn is used to interpret the user command and correspond it to appropriate functionalities such as object detection, text recognition, etc. Scikit-learn provides a comprehensive set of tools to efficiently handle the intent classification process.

i. TfidfVectorizer: Transforms textual data into numerical form based on term frequency-inverse document frequency (TF-IDF), enabling more accurate similarity measurements.

ii. Cosine similarity: Calculates similarity between processed user input and predefined examples, helping to determine the closest matching intent.

JSON: Handles reading and updating a JSON file that contains predefined intents and examples.

Joblib: Saves and loads the trained TfidfVectorizer model to enhance performance and avoid re-training the model in every run.

NLP and Intent Identification Process

Data preprocessing: Text examples undergo preprocessing steps, including tokenization, lemmatization, and stopword removal, which cleans and reduces text to essential terms.

Text vectorization: Processed examples are converted into vectors using TF-IDF, enabling numerical comparison.

Similarity measurement: The user query, once processed, is converted to a vector. Cosine similarity is calculated to find the closest matching intent from predefined examples. If a match exceeds a set similarity threshold, the corresponding action is triggered.

YOLOv8 Library

YOLOv8 is the latest version in the popular YOLO (You Only Look Once) model series, known for fast and accurate object detection in images and videos. With each update, YOLO models become more efficient, and YOLOv8 continues this trend, boasting improved speed and accuracy. This model is ideal for real-time applications, such as identifying people, vehicles, or objects in video streams. It is particularly useful in scenarios where split-second decisions are critical, like self-driving cars and security systems, setting new benchmarks for speed and precision in object detection.

System architecture

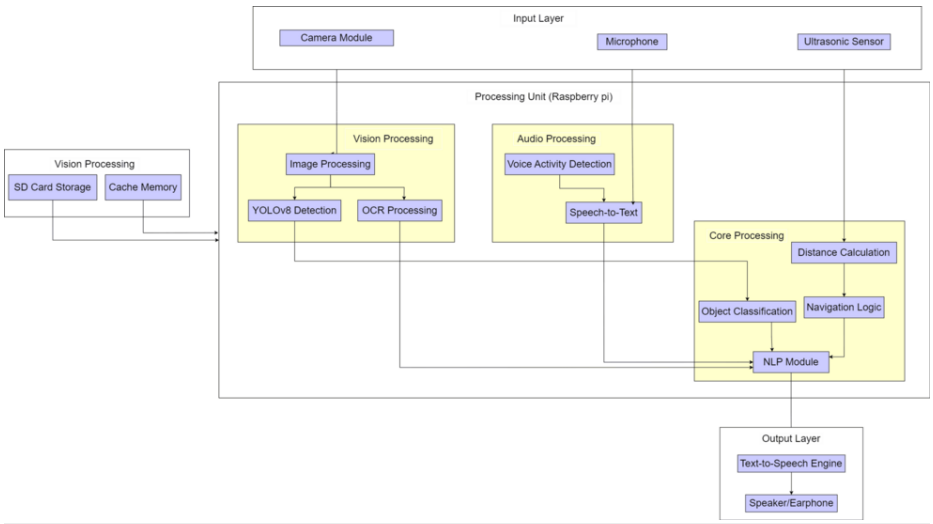

FIGURE 1: System architecture of YOLOInsight

Figure *1* presents a system architecture diagram of a comprehensive assistive device built on a Raspberry Pi platform that integrates multiple input sources (camera, microphone, and ultrasonic sensor) to understand and interact with the environment. The system processes information through three main channels: vision processing (which handles image processing, YOLOv8 object detection, and OCR), audio processing (managing voice detection and speech-to-text conversion), and core processing (handling distance calculations, object classification, and navigation logic). All these inputs are processed through an NLP module that helps in understanding and contextualizing the information. The system uses both SD card storage and cache memory for data management, and ultimately outputs information through a text-to-speech engine connected to a speaker or earphone. This architecture appears designed to create a smart assistive device that can understand its surroundings through multiple sensors, process this information intelligently, and communicate effectively with the user through audio feedback.

Proposed architecture

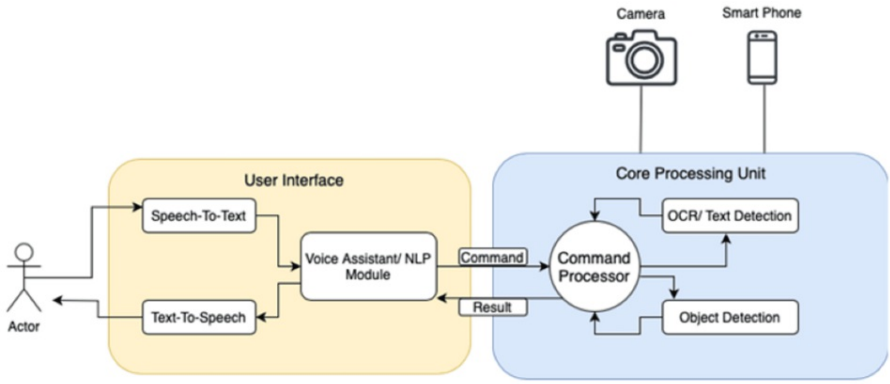

FIGURE 2: Proposed architecture of YOLOInsight

Figure *2* presents a high-level design of the YOLOInsight system, featuring two primary components:

A. User Interface:

The User Interface is responsible for interacting with the user and relaying commands to the Core Processing Unit. It includes the following modules:

i. Speech-To-Text Module: Captures user speech input and converts it into text, allowing users to issue commands verbally.

ii. Voice Assistant/NLP Module: Processes natural language commands, interpreting user instructions and converting them into actionable commands for the system.

iii. Text-To-Speech Module: Converts system outputs or results into speech, allowing the system to verbally communicate with the user.

This part of the architecture enables hands-free, accessible interaction, which is essential for a project focused on assistive technology.

B. Core Processing Unit

The Core Processing Unit handles the main functionalities of YOLOInsight. It processes commands received from the User Interface and performs complex computations to detect and interpret visual data. It consists of the following key features:

i. Command Processor: Acts as the central coordinator, receiving commands from the Voice Assistant/NLP Module and distributing tasks to the respective processing components.

ii. OCR/Text Detection: Responsible for recognizing and extracting text from images. This feature can assist visually impaired users by reading out textual information from their surroundings.

iii. Object Detection: Identifies and categorizes objects in the environment, helping users understand what is around them.

In this setup, the Core Processing Unit leverages machine learning and computer vision capabilities to process visual data, interpret it, and return relevant results to the user through the User Interface.

Workflow summary

i. The user initiates interaction via the User Interface, either through spoken commands or text.

ii. The Voice Assistant/NLP Module processes these commands and forwards them as instructions to the Command Processor.

iii. The Command Processor then routes the instructions to either OCR/Text Detection or Object Detection, depending on the task.

iv. Processed results are returned to the User Interface, which communicates the output back to the user via the Text-To-Speech Module.

This architecture ensures an intuitive and efficient experience, combining NLP with real-time image analysis to deliver critical information to the user in an accessible format.

Used approach in YOLOInsight

Real-time object detection: Utilizes the YOLOv8 algorithm for detecting objects in real-time.
OCR: Converts text from the environment into speech, allowing users to read signs, labels, and other textual information.
NLP: Provides auditory feedback and interprets user commands, enhancing interaction with the system.
Adaptability to different lighting conditions: The system is designed to perform well in both normal and low-light environments.
User-centric design: Focuses on the needs of visually impaired users to ensure accessibility and ease of use.

Pros

High accuracy: The YOLOv8 algorithm achieves a high mean average precision (mAP), ensuring reliable object detection.
Immediate feedback: Users receive real-time auditory feedback, which enhances their situational awareness and navigation capabilities.
Versatility: The system adapts well to various lighting conditions, making it effective in diverse environments.
Enhanced independence: By facilitating navigation and environmental awareness, the technology empowers visually impaired individuals to navigate independently.
Integration of multiple technologies: Combining object detection, OCR, and NLP creates a comprehensive assistive experience.

Cons

Dependence on technology: Users may become reliant on the system, which could be problematic if the technology fails or is unavailable.
Privacy and security concerns: The use of cloud computing for data processing raises concerns about user data privacy and security.
Complexity of use: Some users may find the technology complex or challenging to use, requiring training and adaptation.
Cost: The development and implementation of advanced AI systems can be expensive, potentially limiting accessibility for some users.
Environmental limitations: While the system adapts to various lighting conditions, extreme environments (e.g., heavy rain, fog) may still pose challenges for effective operation.

Cost estimation

Table *1* provides details of the estimated costs for each component required for YOLOInsight system.

Component	Specification	Estimated Cost (INR)
Raspberry Pi 4 Model B	4GB RAM, quad-core Cortex-A72 processor	₹ 4,150
Camera Module	Sony IMX219 PQ CMOS sensor, 8MP, 1080p video resolution	₹ 2,075
Ultrasonic Sensor (HC-SR04)	Obstacle detection, range: 2 cm to 400 cm	₹ 415
MicroSD Card	32GB SanDisk MicroSDHC, Class 10	₹ 830
Earphones	Dynamic drivers, adaptive noise cancellation	₹ 1,660
Push Button (Tactile)	Single-Pole Single-Throw, durable	₹ 165
Power Supply	5 V, 3 A USB-C adapter	₹ 830
GPIO Connection Wires	For connecting sensors and components	₹ 415
Additional Components	Resistors, LEDs, heat sinks, casing, etc.	₹ 830
Total		₹ 11,370

TABLE 1: Cost estimation

Results

Table 2 provides a summary of the intermediate evaluation results for the YOLOInsight system's functionalities, which was conducted across several key metrics, with a focus on real-time object detection, text recognition, and NLP. The subsections below present the initial findings and performance assessments across these dimensions.

Attribute	Technical Description	Result
Model Accuracy (Final Epoch)	YOLOv8's ability to classify objects accurately	96%
Average Precision (AP@0.5)	Precision at IoU threshold 0.5 across object classes	92%
Average Recall (AR@0.5)	Average recall at IoU threshold 0.5 across all object classes	89%
Inference Time per Object Class	YOLOv8 inference time in milliseconds for key object classes	50 ms (Person), 60 ms (Vehicle), 70 ms (Obstacles)
NLP Intent Recognition Accuracy	Accuracy in identifying intent for commands such as "What's ahead ?"	91-95%
NLP Response Latency	Average latency for processing and responding to NLP-based command	650 ms
OCR Text Recognition Accuracy	Optical Character Recognition (OCR) accuracy for printed and handwritten text	85%
Audio Feedback Latency	Delay from detection to spoken feedback output	800 ms

TABLE 2: Summary of intermediate results

IoU: Intersection over Union

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 (see Figure 3). This trend suggests effective model training and convergence, reflecting the model's suitability for real-time object classification tasks required for assistive applications.

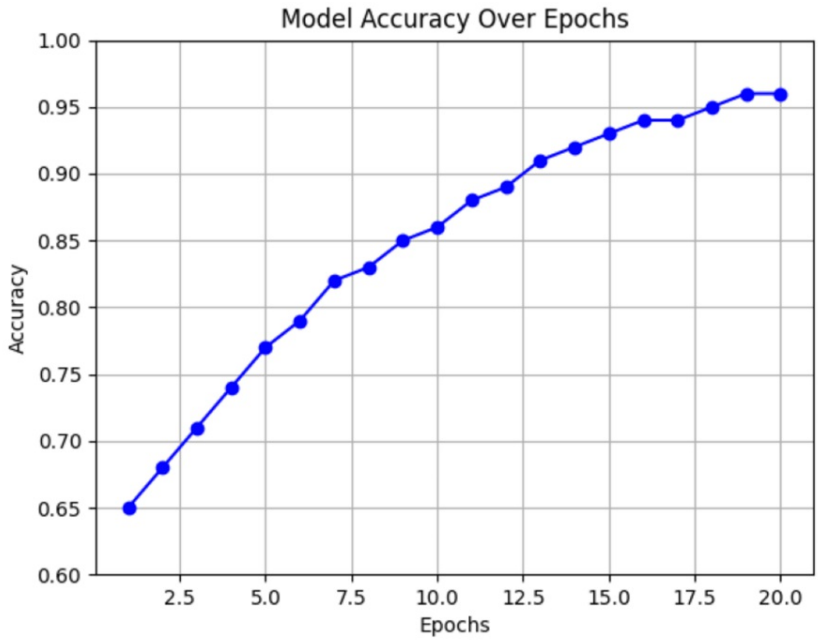

FIGURE 3: Model accuracy over epochs

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 *4*, 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.

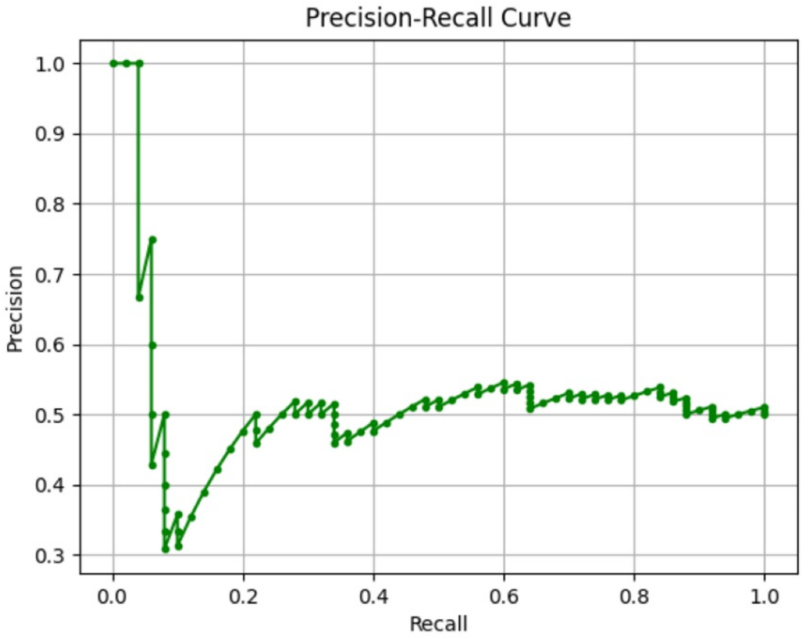

FIGURE 4: Precision-recall curve

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 (Figure 5). 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.

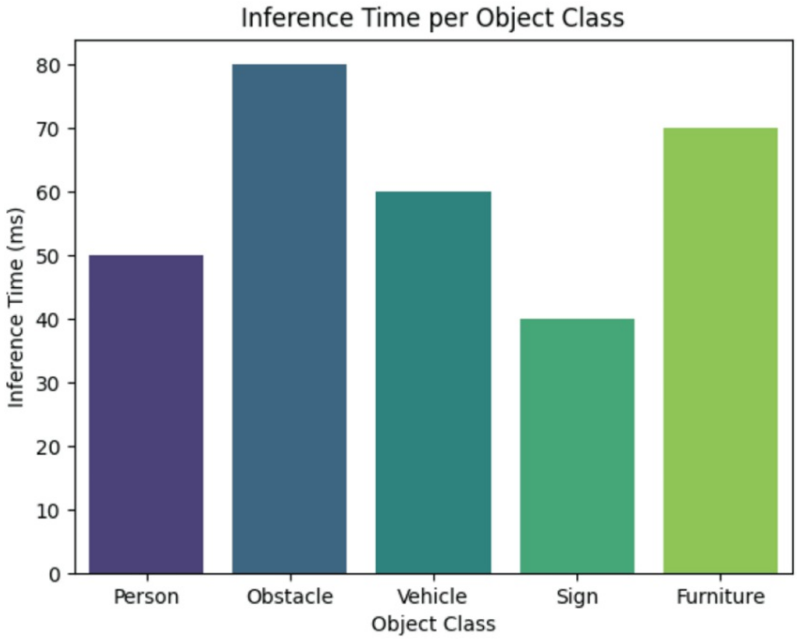

FIGURE 5: Inference time per object class

NLP intent recognition accuracy

The NLP module was evaluated on its accuracy in recognizing common commands. Figure 6 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.

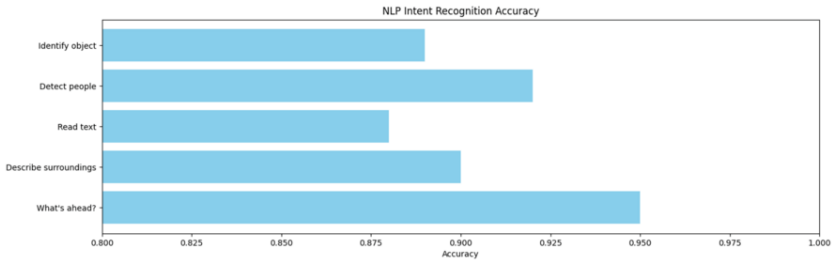

FIGURE 6: NLP intent recognition accuracy

Object detection performance across lighting conditions

Given that lighting can affect detection, tests were conducted in low, normal, and bright lighting conditions. Figure 7 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.

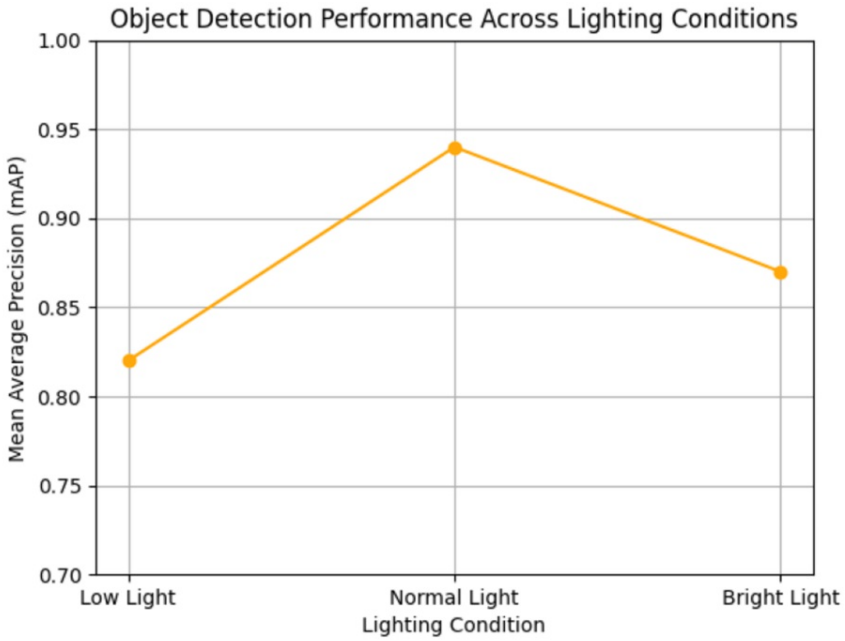

FIGURE 7: Object detection performance across lighting conditions

Comparative approach of different vision assistance systems

In the realm of assistive technologies for visually impaired individuals, technological innovation is crucial for enhancing independence and accessibility. This comparative approach, given in Table *3*, critically evaluates emerging solutions, examining their capabilities across object detection, processing speed, sensor integration, and user experience. By systematically analyzing existing technologies against the YOLOInsight system, the research reveals the progressive advancements in AI-driven assistive devices, highlighting the critical importance of interdisciplinary approaches in creating more effective, user-centric navigation and interaction solutions for the visually impaired community.

Parameter	YOLOInsight	Meta Smart Glasses	AI-Based Shopping Assistance	IoT-Based Vision Alert System
Core Technology	YOLOv8, TesseractOCR, NLP	Meta AI	YOLOv5 object detection	IoT object recognition
Hardware	Raspberry Pi 4 B	Raspberry Pi Zero 2 W	Raspberry Pi 4	ESP32 + Camera
Accuracy	96%	90%	85%	80%
Speed (ms/object)	50-70	100-120	80-90	120-150
Language Support	Multilingual (Custom NLP)	Limited	Basic	Single
Offline Capability	Partial	Full	Limited	Minimal
Sensors	Ultrasonic + Camera	Camera only	Camera + Basic sensors	Ultrasonic
Privacy	Complete control	Third-party dependent	Limited controls	Basic measures
Cost	High affordability	Moderate	Moderate	Low
Accessibility	High	Moderate	Moderate	Low

TABLE 3: Comparison of vision assistance systems

Meta Smart Glasses, Waisberg et al. [1]

AI-Based Shopping Assistance, Sweatha and Sathiya Priya [2]

IoT-Based Vision Alert System, Annapurna et al. [3]

Discussion

Assistive technology for visually impaired individuals has advanced significantly, with recent innovations increasingly centered around smart glasses. These devices are designed to improve independence and enhance quality of life. Several notable projects serve as a foundation in this field. For example, a study by Sweatha and Sathiya Priya from Anna University leverages YOLOv5-driven smart glasses to integrate object detection and OCR, enabling visually impaired users to navigate their environments more autonomously. Similarly, the “Blind Guider” application from the Sri Lanka Institute of Information Technology employs an Android-based platform, combining image processing and GPS to help users avoid obstacles and move safely.

The "Al Amal Glasses," created by Badawi and colleagues, stand out by combining AI and IoT for text reading and obstacle detection, prioritizing affordability to make the device more accessible. Facial recognition has also been explored, offering enhanced social interactions by helping users identify familiar faces. Research led by Professor Syeda Faqera Fatima and her team emphasizes IoT-enabled real-time object recognition, which aims to improve spatial awareness and ease of movement, while a project at the Narayanamma Institute of Technology and Science addresses limitations in existing technologies to improve user experience. In another effort to balance affordability with functionality, Maanil Laad and Dr Vasudha Bahl introduce smart glasses using a Raspberry Pi for text recognition, while T.G. Ramya Priyatharsini’s team focuses on facial recognition and navigation, enhancing social and environmental interaction for visually impaired users.

Our project, YOLOInsight, builds upon these foundational efforts and introduces several unique features that address gaps and limitations identified in existing research.

Integrated NLP API/voice assistant

Unlike other solutions that focus on object detection and OCR, YOLOInsight incorporates a custom-built NLP API/voice assistant. This API enables natural, conversational interactions that offer users a more intuitive, context-aware experience. By prioritizing seamless communication, YOLOInsight moves beyond basic command-based interactions, delivering a more sophisticated and user-friendly interface.

Multi-modal sensing

In addition to a camera-based object detection system, YOLOInsight integrates ultrasonic sensors to enhance spatial awareness and obstacle detection. This multi-sensory approach provides users with more reliable navigation support, significantly improving functionality in complex environments.

AI model

Many projects rely on standard computer vision and OCR models, which, while effective, may not fully meet specific performance needs for visually impaired users. YOLOInsight aims to develop its own optimized AI models, focusing on efficiency and speed. This custom approach seeks to deliver superior performance, especially in real-time scenarios.

Cost-effectiveness

YOLOInsight aims to maintain affordability through the use of a Raspberry Pi platform and custom software solutions. This approach keeps production costs low, making the device accessible to a broader user base and positioning it as a practical alternative to more costly prototypes.

Emphasis on auditory feedback

Designed with visually impaired users in mind, YOLOInsight prioritizes auditory feedback as the primary mode of interaction. Audio output provides clear and timely information about detected objects, navigation cues, and text recognition, offering a truly hands-free experience.

Comprehensive system integration

While previous projects focus on specific aspects of assistive technology, YOLOInsight takes a holistic approach by integrating object detection, text recognition, navigation support, and voice-based control into a single device. This all-in-one system offers a cohesive experience, effectively addressing multiple needs within a single, user-friendly platform.

Future scope

The YOLOInsight system holds significant potential for further advancements, paving the way for a more inclusive and innovative future in assistive technology. One promising area is the integration of augmented reality, which can enhance the user experience by overlaying spatial cues, such as directional arrows or object descriptions, on real-time images. This feature would be particularly useful for navigating complex environments, such as crowded public spaces or shopping malls, where additional visual context can greatly improve situational awareness.

The inclusion of cloud-based processing could significantly enhance the system's computational capabilities. By leveraging the cloud, YOLOInsight could incorporate more advanced algorithms for object detection, manage personalized user preferences, and facilitate dynamic updates to object databases. This approach would also enable seamless access to user data across devices, fostering a collaborative ecosystem where improvements benefit all users. However, with cloud integration, robust security measures must be prioritized. Implementing end-to-end encryption, secure protocols, and local processing options can ensure user privacy and build trust in the system.

Expanding the NLP module offers another exciting avenue for development. Future iterations could support more complex, context-aware interactions, allowing users to issue layered commands such as “Describe the closest object and its distance.” Multilingual support could also be broadened to accommodate global users, while features like emotion recognition could make the system more empathetic and user-friendly.

Field testing in diverse real-world environments is another critical area for growth. Trials in urban spaces, rural settings, and indoor locations such as malls and airports will provide valuable insights into the system's adaptability to different scenarios. These tests will help refine the device’s functionality, ensuring reliable performance regardless of environmental conditions.

User-driven customization can further enhance YOLOInsight’s utility. By allowing users to tailor features such as alert types, frequency, and mode selection, the system can cater to individual preferences. Additionally, optimizing power management through energy-efficient hardware and low-power modes can extend battery life, ensuring prolonged use without sacrificing performance.

Another exciting possibility lies in developing wearable versions of YOLOInsight, such as lightweight smart glasses or headsets. These wearables would offer hands-free usability while maintaining the robust features of the current system. Integration with smartwatches or other IoT devices could add functionalities like health monitoring, making the device even more versatile.

Collaborating with smart infrastructure initiatives can further elevate YOLOInsight’s capabilities. Features such as beacon technology for precise location guidance and integration with public transportation systems could significantly improve navigation for visually impaired users. These advancements would allow seamless interaction with smart cities, fostering independence in both urban and rural settings.

Finally, YOLOInsight can play a transformative role in education and workplaces. It can assist visually

impaired individuals in navigating educational institutions, reading printed materials, or performing tasks in office environments. By bridging the gap between accessibility and productivity, YOLOInsight can empower individuals to participate fully in academic and professional spaces.

In conclusion, the future scope of YOLOInsight encompasses a wide range of enhancements that aim to improve scalability, usability, and integration with emerging technologies. By focusing on these areas, YOLOInsight can evolve into a comprehensive, all-in-one platform that transforms assistive technology and fosters a more inclusive world.

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 OCR and 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. This would allow users to enjoy tailored experiences seamlessly, regardless of the device in use. 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.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Sean D'silva, Saurabh Dabhade, Gajanan Arsalwad

Drafting of the manuscript: Sean D'silva, Saurabh Dabhade, Kabir Shaikh

Critical review of the manuscript for important intellectual content: Sean D'silva, Saurabh Dabhade, Gajanan Arsalwad

Supervision: Gajanan Arsalwad

Acquisition, analysis, or interpretation of data: Kabir Shaikh

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

References

1. Waisberg E, Ong J, Masalkhi M, Zaman N, Sarker P, Lee AG, Tavakkoli A: Meta smart glasses—large language models and the future for assistive glasses for individuals with vision impairments. Eye. 2024, 38:1036-1038. 10.1038/s41433-023-02842-z
2. Sweatha R, Sathiya Priya S: YOLOv5 driven smart glasses for visually impaired . International Journal of Science and Research Archive. 2024, 12:353-374. 10.30574/ijsra.2024.12.1.0804
3. Annapurna T, Mamidoju SS, Parthasaradhi G, et al.: An IoT-based vision alert for blind using

interdisciplinary approaches. E3S Web of Conferences. 2024, 507:01047. 10.1051/e3sconf/202450701047

4. Poy Y-L, Darmaraju S, Goh C-H, Kwan B-H: Standalone smart glass system for the blind and visually impaired. 2024 IEEE 14th Symposium on Computer Applications & Industrial Electronics (ISCAIE). 2024, 239-244. 10.1109/ISCAIE61308.2024.10576410
5. Badawi M, Al Nagar E, Mansour R, Ibrahim Kh, Hegazy N, Elaskary S: Smart bionic vision: An assistive device system for the visually impaired using artificial intelligence. International Journal of Telecommunications. 2024, 4:1-12. 10.21608/ijt.2024.342832
6. Raju D, Pooja S, Bhat PG, Darshan EM, Harisha GC: Smart glass for visually impaired people . International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering. 2023, 4:44-47.
7. Laad M, Bahl V: Empowering differently abled individuals through IoT: Smart glasses for the visually impaired people. Iconic Research and Engineering Journals. 2023, 6:382-386.
8. Sujay CV, Suhas GR, Sunidi VK, Varuni V, Swamy SR: Design of smart goggle for visually impaired with audio features. International Journal of Advanced Research in Computer and Communication Engineering. 2023, 12: 2319-5940.
9. Sajini S, Pushpa B: A binary object detection pattern model to assist the visually impaired in detecting normal and camouflaged faces. Engineering, Technology & Applied Science Research. 2024, 14:12716-12721. 10.48084/etasr.6631
10. Abdulatif A, Mohammed S, Mohammed A, Mohammed H, Rajalingam A: Smart glasses robot for blind people using raspberry-pi and python. Journal of Science, Computing and Engineering Research. 2022, 3:237-244.
11. Sudharani B, Lokanath K, Bhavya G, Hema Priya K, Venkat K, Revanth Sri Chandu G: Smart glasses for visually challenged people using facial recognition. International Journal of Innovative Research in Science Engineering. 2022, 8:37-45.
12. Ramdani A, Virgono A, Setianingsih C: Food detection with image processing using convolutional neural network (CNN) method. 2020 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT). 2020, 91-96. 10.1109/IAICT50021.2020.9172024
13. Abhinav GM, Chandrachood Bharadwaj P, Bhat SR, Gopala Krishna UJ, Srinath R, Bharadwaj RS: Smart glasses to assist the blind. International Research Journal of Modernization in Engineering Technology and Science. 2022, 4:3985-3992.
14. Kim J-H, Kim S-K, Lee T-M, Lim Y-J, Lim J: Smart glasses using deep learning and stereo camera . 2019 IEEE 8th Global Conference on Consumer Electronics (GCCE). 2019, 294-295. 10.1109/GCCE46687.2019.9015357