VisionAid: Real-Time Object Detection and Recognition for Visually Impaired Individuals Using AI

Step 1: Prototype Selection

Abstract

Vision is a crucial ability that enables individuals to perceive the world around them. Many visually impaired people face significant challenges in identifying and navigating their surroundings, making it difficult to move around and socialize in new environments. This project aims to assist visually impaired individuals by providing navigational support and auditory feedback, allowing them to identify specific objects or pathways more efficiently and reduce their reliance on others. The primary goal is to empower visually impaired individuals, thereby benefiting all stakeholders. This study explores the concept of interpreting visual objects through auditory cues, leveraging the interplay between auditory and optical senses to locate items in space.

1. Problem Statement

Visually impaired individuals encounter significant obstacles in their daily lives due to their inability to perceive their surroundings visually. These challenges include difficulty in identifying objects, recognizing pathways, and navigating unfamiliar environments, which often leads to increased dependency on others for basic mobility and interaction. Traditional aids, such as canes and guide dogs, provide some level of assistance but do not offer a comprehensive solution for real-time object recognition and navigation.
Our solutions helps to address such problems.

2. Market/Customer/Business Need Assessment

Globally, over 285 million individuals are visually impaired, including 39 million who are completely blind. This significant demographic faces daily challenges that impact their independence and confidence in performing routine activities. The primary need for this population is to gain autonomy and enhance their quality of life through effective assistive technologies.

This project aims to address this critical need by developing an affordable and effective solution for object recognition and navigation. The system is particularly beneficial for small businesses or startups focused on creating assistive technology for visually impaired users. By offering a cost-effective tool that integrates real-time object detection and auditory feedback, this solution can significantly improve the daily lives of visually impaired individuals.

In India, the need for such technology is especially pronounced. According to the National Programme for Control of Blindness (NPCB), approximately 8 million people in India are blind, with many more experiencing low vision. This substantial demographic is in dire need of tools and technologies that can enhance their independence and daily functionality.

Small to medium-sized businesses (SMBs), including local tech companies, healthcare providers, and assistive technology firms, represent a significant market for this AI solution. These businesses are well-positioned to adopt and distribute this technology, making it accessible to a broader audience of visually impaired individuals. By leveraging this solution, SMBs can offer innovative products that meet a crucial need, fostering greater independence and improving the quality of life for their visually impaired customers.

3. Target Specifications and Characterization

The primary target include:

- Visually Impaired Individuals: People who are blind or have low vision, seeking tools to enhance their independence and navigate their environments more effectively.
- Organizations Supporting Visually Impaired People: Non-profits, educational institutions, and advocacy groups dedicated to improving the quality of life for visually impaired individuals.
- Startups and Small Businesses in the Assistive Technology Sector: Companies developing innovative solutions aimed at assisting visually impaired users.

Customer Characteristics:

- Need for Independence and Safety: Visually impaired individuals prioritize products that enable them to perform daily activities independently and safely.
- Dependence on Auditory Information: Given the lack of visual input, these customers rely heavily on auditory cues to understand and interact with their surroundings.
- Requirement for Real-Time and Accurate Object Detection: To be effective, the technology must provide immediate and precise identification of objects in the environment, ensuring users can trust and rely on the information provided.

4. External Search

Research on similar technologies, existing patents, and scientific literature was conducted using online databases like IEEE Xplore, Google Scholar, and relevant industry reports.

5. Benchmarking Alternate Products

OrCam MyEye: A wearable device that utilizes Optical Character Recognition (OCR) and deep learning to read text, recognize faces, and identify objects. This device offers hands-free operation and instant feedback, enhancing the user's ability to interact with their environment.

Aira: A service that connects visually impaired users with remote agents through live video feeds. These agents provide real-time visual information, helping users navigate and interpret their surroundings. Aira combines human assistance with advanced technology to offer personalized support.

Seeing AI: A Microsoft-developed app that uses artificial intelligence to narrate the world around the user. It recognizes people, objects, and text, providing descriptive audio feedback to help visually impaired individuals understand and interact with their environment more effectively.

6. Applicable Patents

· US10275917B2: "Systems and methods for object recognition"

7. Applicable Regulations

· Rights of Persons with Disabilities Act (RPWD), 2016: Ensuring our product meets accessibility standards set by the Indian government.

· GDPR (General Data Protection Regulation): If handling user data in the EU.

· Central Drugs Standard Control Organization (CDSCO): If marketed as a medical device in India.

· FCC (Federal Communications Commission) regulations for electronic devices.

8. Applicable Constraints

· Limited computational power for real-time processing on portable devices.

· Budget constraints for small businesses or startups.

· Requirement for user-friendly interfaces that cater to visually impaired users.

9. Business Opportunities:

· Growing Market Demand: As the global population of visually impaired individuals increases, so does the need for effective assistive technologies. VisionAid addresses this demand by offering a cutting-edge solution that combines real-time object detection with auditory feedback, providing a unique and valuable tool for users.

· Untapped Market Segments: There is substantial potential within underserved markets, including emerging economies where access to advanced assistive technology is limited. By focusing on affordability and accessibility, VisionAid can tap into these growing markets and expand its reach.

· Partnership Potential: Collaborating with non-profit organizations, educational institutions, and healthcare providers offers significant opportunities for growth. These partnerships can facilitate widespread adoption and integration of VisionAid into existing support systems and educational programs.

· Innovation and Differentiation: With advancements in AI and machine learning, VisionAid stands out by leveraging YOLO for real-time object detection and integrating it with text-to-speech technology. This innovative approach provides a competitive edge and meets the evolving needs of visually impaired users more effectively than existing solutions.

· Scalable Revenue Models: The project's diverse monetization strategies—including direct sales, subscriptions, B2B partnerships, and a freemium model—allow for flexible and scalable revenue streams. This versatility enhances the potential for financial growth and market penetration.

10. Concept Generation

The concept generation phase focuses on developing innovative solutions to address the needs of visually impaired individuals through real-time object recognition. This process involves several key steps:

· Identifying the Need: The primary challenge is to provide visually impaired individuals with a reliable means of real-time object recognition. This technology is essential for enhancing their independence and safety by enabling them to identify objects and navigate their environment with greater ease.

· Brainstorming Solutions: Leveraging existing technologies such as YOLO (You Only Look Once) for object detection and text-to-speech systems, we explored various ways to address the identified need. YOLO offers powerful real-time object detection capabilities, while text-to-speech technology provides auditory feedback, making it possible to convey information about detected objects to users in a meaningful way.

· Evaluating Feasibility: We assessed the practicality of integrating YOLO and text-to-speech technologies into a cohesive and user-friendly device. This evaluation considered factors such as hardware requirements, software compatibility, user interface design, and overall system performance. The goal was to ensure that the final solution would be both effective and accessible for visually impaired users, providing them with a seamless and intuitive experience.

11. Concept Development

Development of a Wearable Device or Mobile Application: Design and build a system that utilizes YOLO (You Only Look Once) for real-time object detection. The choice between a wearable device and a mobile application will depend on factors such as user preference, portability, and integration capabilities. The system will leverage YOLO's advanced object detection algorithms to accurately identify objects in the user's environment.

Integration of Text-to-Speech (TTS) Capabilities: Incorporate TTS technology to convert detected objects into auditory feedback. This feature will provide users with real-time, spoken descriptions of the objects identified by YOLO, enhancing their ability to navigate and interact with their surroundings. Ensuring clear, natural, and easily understandable speech output will be critical for user satisfaction.

Design for Lightweight and Portability: Ensure that the device or application is designed to be lightweight and portable, offering maximum convenience and ease of use. For a wearable device, this involves minimizing weight and optimizing comfort. For a mobile application, it includes ensuring compatibility with a range of devices and providing an intuitive user interface.

12. Final Product Prototype/ Product Details

The final product will be a wearable device, such as smart glasses, designed to assist visually impaired individuals through advanced real-time object detection and auditory feedback. The development process involves the following key components:

· Wearable Device: The solution will be embodied in a pair of smart glasses equipped with a camera for capturing the user's environment. This wearable format ensures that the technology is unobtrusive and seamlessly integrates into the user's daily life.

· Image Acquisition: The camera will be connected to a Raspberry Pi 4 Model B, which will handle the acquisition and initial processing of images. The Raspberry Pi 4 provides the necessary computational power for real-time image processing while maintaining a compact and lightweight form factor.

· Data Transfer: Acquired images will be transferred to Google Drive using rclone, a command-line program that facilitates secure and efficient data transfer between local storage and cloud services. This enables the system to manage and store image data effectively.

· Object Detection: The YOLO (You Only Look Once) algorithm will process the video feed to detect and identify objects within the captured environment. YOLO's real-time object detection capabilities will ensure that users receive timely and accurate information about their surroundings.

· Auditory Feedback: The gTTS (Google Text-to-Speech) module will be used to convert detected object names into spoken words. This auditory feedback will provide users with real-time, spoken descriptions of objects, enhancing their ability to understand and interact with their environment.

How It Works:

· Real-Time Environment Capture: The smart glasses are equipped with a camera that continuously captures the user's environment in real-time. This ensures that the system provides up-to-date information about the surroundings.

· Data Processing on Raspberry Pi: The live video feed from the camera is transmitted to a Raspberry Pi 4 Model B. This compact and powerful computing unit processes the video feed locally, handling the computational tasks required for object detection.

· Image Transfer to Cloud Storage: Using rclone, the images captured by the camera are securely transferred to Google Drive. This cloud storage solution facilitates easy management and access to the image data for further processing.

· Object Detection with YOLO: The YOLO (You Only Look Once) algorithm analyzes the video feed in real-time to detect and identify objects within the environment. YOLO's advanced object detection capabilities ensure accurate and timely identification.

· Text-to-Speech Conversion: Detected objects are converted into spoken words using the gTTS (Google Text-to-Speech) module. This conversion allows the system to provide auditory feedback about the identified objects.

· Auditory Feedback: The user receives real-time auditory feedback through earphones or built-in speakers integrated into the smart glasses. This audio output conveys the names and descriptions of detected objects, aiding in navigation and interaction.

Raspberry Pi 4 Model B

The Raspberry Pi 4 Model B is a compact, high-performance single-board computer ideal for real-time applications. Key features include:

· Powerful Processor: Equipped with a quad-core ARM Cortex-A72 CPU at 1.5 GHz, it provides the processing power needed for tasks such as real-time image processing.

· Memory and Storage: Available in 2GB, 4GB, and 8GB RAM variants, with microSD card support for storage, it accommodates various application needs.

· Connectivity: Includes two USB 3.0 ports, two USB 2.0 ports, Gigabit Ethernet, built-in Wi-Fi, and Bluetooth, ensuring versatile connectivity options.

· Camera Interface: Features a dedicated CSI interface for direct connection to a camera module, enabling efficient image acquisition.

· Efficient Cooling: Supports heat management solutions to maintain performance during intensive tasks.

YOLO Object Detection

YOLO (You Only Look Once) is a highly efficient algorithm designed for real-time object detection using Convolutional Neural Networks (CNNs). YOLO distinguishes itself from other object detection methods through its unified approach to object detection, combining both localization and classification tasks in a single neural network.

Key Features of YOLO:

· Unified Detection Process: Unlike Region-Based Convolutional Neural Networks (R-CNN) and its variations, which involve multiple stages and are often slower and more complex to optimize, YOLO uses a single neural network to predict bounding boxes and class probabilities for objects in one pass. This significantly accelerates the detection process and makes YOLO suitable for real-time applications.

· Network Architecture: YOLO employs a neural network architecture known as Darknet-53. This architecture, as depicted in Fig. 3, is a fully convolutional network consisting of 53 convolutional layers. Darknet-53 utilizes convolutional layers exclusively, allowing it to efficiently process and analyze images.

· Image Down-Sampling: YOLO processes images by down-sampling them through a factor known as stride. For instance, with a stride of 16, an input image of size 512×512 pixels will be down-sampled to an output size of 32×32 pixels. This down-sampling reduces the spatial dimensions of the image while retaining essential features for object detection.

· Bounding Box and Probability Prediction: YOLO divides the image into a grid and predicts bounding boxes and class probabilities for each grid cell. This approach allows the network to detect multiple objects within each region of the image, improving the accuracy and efficiency of object detection.

The YOLO algorithm employs several key techniques to achieve efficient and accurate object detection. These methods include:

· Residual Blocks: YOLO utilizes residual blocks within its network architecture to enhance learning efficiency and improve performance. Residual blocks, a component of the Darknet-53 network, allow the model to learn residuals or differences between the predicted and actual outputs. This approach helps in training deeper networks by mitigating issues like vanishing gradients and improving the network's ability to learn complex features.

· Bounding Box Regression: YOLO performs bounding box regression to predict the locations of objects within an image. This involves estimating the coordinates of bounding boxes (x, y, width, and height) that encapsulate detected objects. YOLO does this by dividing the image into a grid and predicting bounding boxes and class probabilities for each grid cell. This method ensures precise localization of objects in the detected regions.

· Intersection Over Union (IoU): IoU is used to measure the accuracy of the predicted bounding boxes by comparing them with the ground truth boxes. It calculates the ratio of the area of overlap between the predicted and actual bounding boxes to the area of their union. Higher IoU values indicate better alignment between predicted and true object locations, which is crucial for accurate object detection.

· Non-Maximum Suppression (NMS): To address the issue of multiple overlapping bounding boxes for a single object, YOLO employs Non-Maximum Suppression. NMS eliminates redundant boxes by keeping only the box with the highest confidence score and removing others that overlap significantly with it. This step helps in reducing false positives and ensures that each object is detected by a single, accurate bounding box.

Text to Speech

The primary objective of this project is to convert detected text into audio output for visually impaired users. This process involves two main components:

- Text Extraction with Tesseract-OCR: Text is extracted from images using the Tesseract-OCR engine, an open-source Optical Character Recognition (OCR) tool. Tesseract processes the images to identify and extract textual content.
- Audio Conversion with gTTS: The extracted text is then converted to speech using the gTTS (Google Text-to-Speech) module. gTTS is a Python library that interfaces with Google's Text-to-Speech API, converting text into spoken words efficiently. The gTTS module generates an MP3 file of the spoken text.
    - Execution on Raspberry Pi: Both text extraction and audio conversion processes are executed on the Raspberry Pi, leveraging its compact hardware design for portability and integration.
    - Accuracy and Performance: The combined text detection and audio conversion process achieves an accuracy of approximately 85% with random samples. gTTS provides rapid conversion to MP3 format, facilitating quick and efficient auditory feedback.
    - Audio Playback: The generated MP3 file is played back using the playsound library, an open-source command-line sound player compatible with Python and optimized for use on Raspberry Pi.

Advantages and Limitations of the Work

Advantages:

· Enhanced Independence: YOLOv3 object detection significantly boosts the independence of visually impaired individuals by enabling them to detect and recognize objects and obstacles in their environment. This capability allows for greater autonomy in daily activities and navigation.

· Improved Safety: By providing timely and accurate information about the presence of objects and barriers, the system reduces the risk of accidents and enhances the overall safety of users. This helps in preventing collisions and navigating safely through various environments.

· Better Quality of Life: The real-time feedback from the YOLOv3 model facilitates smoother execution of everyday tasks, such as locating objects and avoiding obstacles. This improvement in navigation aids in making the world more accessible and manageable for individuals with visual impairments.

· Faster Detection: YOLOv3's ability to process images quickly ensures that users receive immediate feedback about their surroundings. This rapid detection capability supports real-time decision-making and enhances situational awareness.

Limitations:

· Recognition Accuracy: The accuracy of object detection and recognition can be compromised in complex or cluttered environments. YOLOv3 may struggle to distinguish between objects in such scenarios, which can affect the reliability of the system.

· Lighting Conditions: Variations in lighting, such as low light or excessive brightness, can impact the system's performance. Inconsistent or poor lighting conditions may lead to inaccurate or unreliable object detection, affecting the user's experience and safety.

· Complex Environments: In environments with high visual complexity or multiple overlapping objects, YOLOv3 may face challenges in accurately identifying and distinguishing objects. This limitation can cause discomfort or confusion for users relying on the system for navigation.

13. Feasibility

The development and deployment of this project are highly feasible within a few months, considering the following factors:

· Technology Readiness: The core technologies (YOLOv3 for object detection, gTTS for text-to-speech, and Raspberry Pi for processing) are well-established and have proven efficacy in similar applications. These technologies are readily available and can be integrated with existing tools and libraries.

· Development Timeline: The project can be completed in stages—design, prototyping, testing, and deployment. Each stage involves manageable tasks that can be achieved within a few weeks to a couple of months, depending on resource availability and team efficiency.

· Hardware and Software Integration: Components such as the Raspberry Pi, camera modules, and the gTTS library are compatible and widely used in similar projects. Integration is straightforward with available documentation and community support.

· Prototype to Product Transition: Transitioning from a prototype to a full-scale product involves refining the design, ensuring scalability, and addressing user feedback. With a well-defined development plan, this transition can be accomplished efficiently.

14. Viability

The project is viable due to its alignment with market needs and potential impact:

· Market Demand: There is a significant demand for assistive technologies among visually impaired individuals. The project addresses key challenges in object detection and navigation, making it a valuable tool for enhancing independence and safety.

· Competitive Edge: The combination of real-time object detection with YOLOv3 and auditory feedback through gTTS offers a unique and innovative solution. This can differentiate the product from existing assistive technologies in the market.

· User Acceptance: The focus on user-friendly design and real-time feedback is likely to resonate well with the target audience, ensuring acceptance and adoption.

· Scalability: The technology stack used is scalable, allowing for future enhancements and adaptation to various use cases and environments.

15. Monetization

The project offers several monetization avenues:

· Direct Sales: Sell the wearable device or mobile application directly to consumers. This approach involves a one-time purchase fee for the hardware or app.

· Subscription Model: Implement a subscription model for ongoing access to software updates, additional features, or premium support. This could include monthly or annual plans.

· B2B Partnerships: Partner with organizations, schools, and healthcare providers that support visually impaired individuals. These partnerships can facilitate bulk sales and integration into existing assistive technology programs.

· Freemium Model: Offer a basic version of the product for free, with premium features or enhanced capabilities available through a paid upgrade. This model can attract a broad user base and convert a portion to paying customers.

· Grants and Funding: Seek grants and funding from organizations that support assistive technologies and disability-related innovations. This financial support can aid in product development and market entry.

· Advertising and Sponsorships: Explore opportunities for advertising or sponsorships within the app or device, provided they align with the users' needs and preferences.

Step 2: Prototype Development

Github Link: https://github.com/LearningBoy2002/VisionAid.git

Step 3: Business Model

Problem Statement

Visually impaired individuals face significant challenges in navigating and interacting with their environments due to their inability to see. This project aims to develop a system that uses YOLO (You Only Look Once) for object detection and recognition, providing real-time aural feedback to help visually impaired persons recognize objects and navigate their surroundings independently.

Market/Customer/Business Need Assessment

There are over 285 million visually impaired individuals globally, including 39 million who are blind. The primary need for this demographic is to gain independence and confidence in their daily activities. This solution will cater to small businesses or startups developing assistive technology for visually impaired users, providing an affordable and effective aid for object recognition and navigation.

According to the National Programme for Control of Blindness (NPCB), India has a significant population of visually impaired individuals, with approximately 8 million people who are blind and many more with low vision. This demographic seeks tools and technologies to improve their daily independence. Small to medium-sized businesses (SMBs) such as local tech companies, healthcare providers, and assistive technology firms are potential customers for this AI solution.

Target Specifications and Characterization

The primary target customers are:

- Visually impaired individuals
- Organizations supporting visually impaired people
  - Startups and small businesses in the assistive technology

sector Customer characteristics:

- Need for independence and safety
- Dependence on auditory information
- Requirement for real-time and accurate object detection

External Search

Research on similar technologies, existing patents, and scientific literature was conducted using online databases like IEEE Xplore, Google Scholar, and relevant industry reports.

Benchmarking Alternate Products

- OrCam MyEye: A wearable device that reads text and recognizes faces and objects using OCR and deep learning.
- Aira: A service that connects visually impaired users with remote agents who provide visual information through live video feeds.
- Seeing AI: A Microsoft app that narrates the world around the user by recognizing people, objects, and text.

Applicable Patents

- US10275917B2: "Systems and methods for object recognition"

Applicable Regulations

- Rights of Persons with Disabilities Act (RPWD), 2016: Ensuring our product meets accessibility standards set by the Indian government.
- GDPR (General Data Protection Regulation): If handling user data in the EU.
    - Central Drugs Standard Control Organization (CDSCO): If marketed as a medical device in India.
- FCC (Federal Communications Commission) regulations for electronic devices.

Applicable Constraints

- Limited computational power for real-time processing on portable devices.
- Budget constraints for small businesses or startups.
- Requirement for user-friendly interfaces that cater to visually impaired users.

Business Model

- Monetization Idea:
    - Direct Sales: Sell the device or application directly to customers.
    - Subscription Model: Monthly or annual subscription for continuous updates and support.
    - B2B Partnerships: Partner with organizations and schools for visually impaired individuals to distribute the technology.
    - Freemium Model: Basic version for free, premium features for a fee.

Concept Generation

- Identifying the need for real-time object recognition for visually impaired individuals.
    - Brainstorming possible solutions using existing technologies like YOLO and text-to- speech systems.
- Evaluating the feasibility of integrating these technologies into a user-friendly device.

Concept Development

- Develop a wearable device or mobile application using YOLO for object detection.
- Integrate text-to-speech (TTS) capabilities to provide real-time auditory feedback to users.
- Ensure the system is lightweight and portable to enhance user convenience.

Final Product Prototype (abstract) with Schematic Diagram

The final product will be a wearable device, such as smart glasses, equipped with a camera to capture the environment. The YOLO algorithm will process the video feed to detect objects, and the Pyttsx3 library will convert object names into speech, providing real-time auditory feedback to the user.

Schematic Diagram:

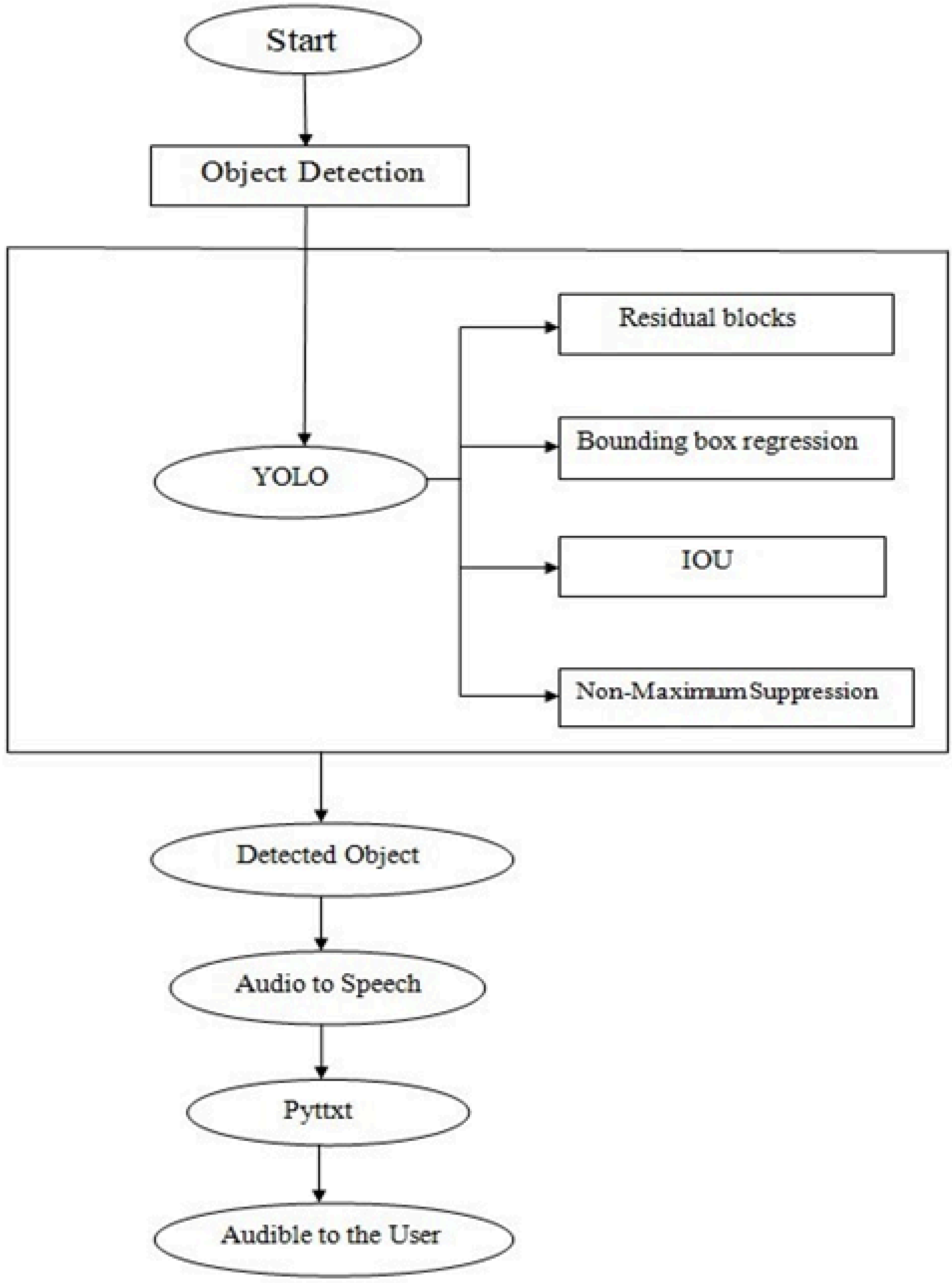

Fig.1. Workflow

Product Details

How does it work?

· The camera captures the environment in real-time.

· The YOLO algorithm processes the video feed to detect objects.

· Detected objects are converted to speech using Pyttsx3.

· The user receives auditory feedback through earphones or speakers.

Data Sources

· Pre-trained YOLO model weights from Kaggle

· Custom datasets for specific object detection scenarios.

Algorithms, Frameworks, Software

· YOLOv3 and YOLO-TINYv3 for object detection.

· Pyttsx3 for text-to-speech conversion.

· Flask framework for integration and deployment.

Team Required

· Data Scientist

· Software Developer

· UI/UX Designer

· Project Manager

Cost

- Development cost: ₹35,00,000
- Hardware cost (smart glasses): ₹22,500 per unit
- Ongoing maintenance and updates: ₹3,50,000 annually

Code Implementation/Explanation of the methods used

This article will illustrate object detection in Python using the YOLOTINYv3 deep learning algorithm and the ImageAI package.

A. ImageAI

An object recognition library created in Python is called ImageAI. Developers can create systems and applications with self-contained deep learning and computer vision with the aid of ImageAI. The ImageAI library can implement nearly all of the most recent deep learning algorithms, including YOLOv3 and TinyYOLOv3.The initial step consists of importing the required modules. After importing the ImageAI library and the Object detection class, the next step is to create an instance of the class. We can now call the various functions of the Object detection class after creating an instance of the class. Among the pre-trained models that can be called using the class's functions are Retina Net, YOLOv3, and TinyYOLOv3.In this article, we'll be utilizing the pre-trained TinyYOLOv3 model.

B. Yolotinyv3 architecture

The YOLO architecture resembles Google Net. It has a total of 24 convolutional layers, four max-pooling layers, and two fully connected layers, as shown in figure 1.

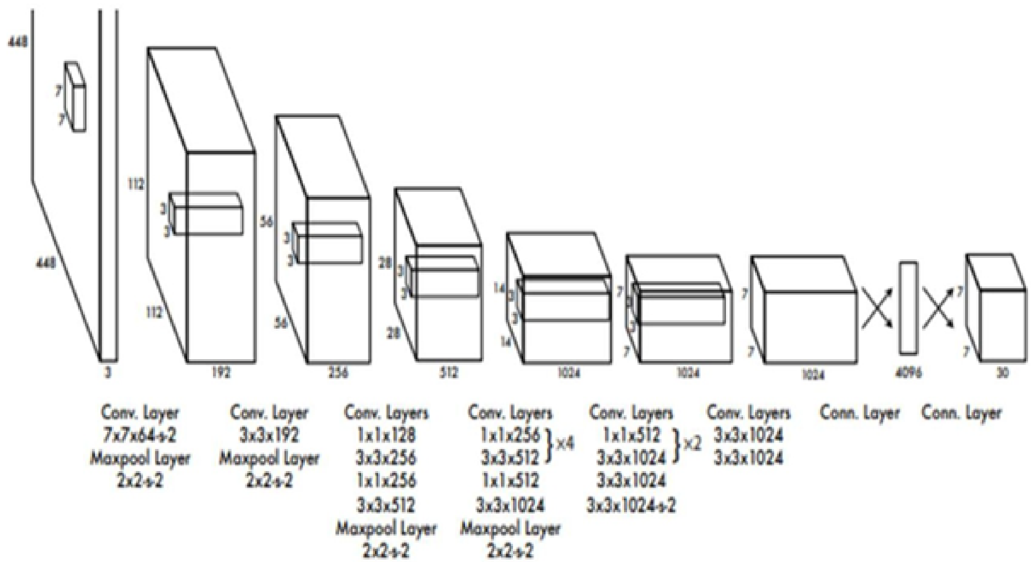

Fig. 2. YOLO Architecture

YOLO functions:

i. The input signal is shrunk to pixel size prior to being sent through the convolutional network.

ii. After an 11 convolution to reduce the number of channels, a 33 convolution is used to produce a cuboidal output.

iii. The activation function employed internally is a rectified linear unit, with the exception of the final layer, which uses a linear activation function.

iv. The model is normalized via other techniques, such as batch normalization and dropout, which also prevent the model from over fitting.

Based on the following four methods, the YOLO algorithm operates:

i. Residual blocks

ii. Bounding box regression

iii. Intersection Over Unions or IOU for short

iv. Non-Maximum Suppression.

C. Residual blocks:

Starting with the residual block is our first step where $n \times n$ grids will be created from my original image. N is an infinite number. The class of the object that each grid cell covers must be predicted locally, along with the probability/confidence value. The image below displays a 5 × 4 grid.

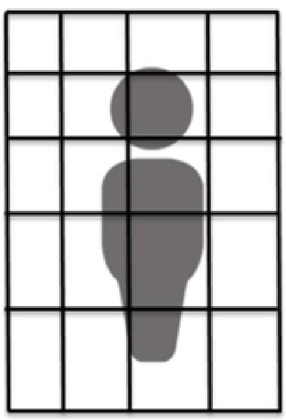

Fig.3. Image of 5 × 4 grid

D. Bounding box regression

A bounding box, where the anticipated area will be accessible, is the next step. The number of images in the anticipated area will be reflected in the bounding boxes. YOLO uses a single regression module to determine these bounding box's characteristics.

E. Intersection over unions or IOU

A single object in an image can frequently have multiple grid box candidates for prediction, even though not all of them are pertinent. The IOU's objective is to remove these grid boxes and keep only the pertinent ones. IOU will be very helpful in that situation if there are multiple images with numerous bounding boxes and our system is confused about which to select.

F. Non-max suppression or NMS

Establishing an IOU threshold is not always sufficient because an item may contain numerous boxes with IOU that exceeds the threshold, and leaving all of those boxes open could result in the inclusion of noise. Here, NMS can be employed to keep just the boxes with the highest likelihood of being discovered.

Through the YOLOTINYv3 steps, our algorithm has identified the objects, but that is not sufficient by itself. Our primary objective is to turn the detected object into speech because visually impaired people cannot see anything.

G. Pyttxt

Pyttsx3 is the name of a Python text-to-speech conversion library. This module can be installed in your working environment. Pyttxt was used in this article to turn text into speech. The speech is presently linked to the recognized item. It is now feasible to make an object visible to those with vision complexities.

Results and Discussion

These are a few of the objects that the YOLO algorithm can identify. Thus, these objects will be transformed into speech, such as "I see person" which can be used by people who are visually impaired, enabling them to carry out daily activities independently.

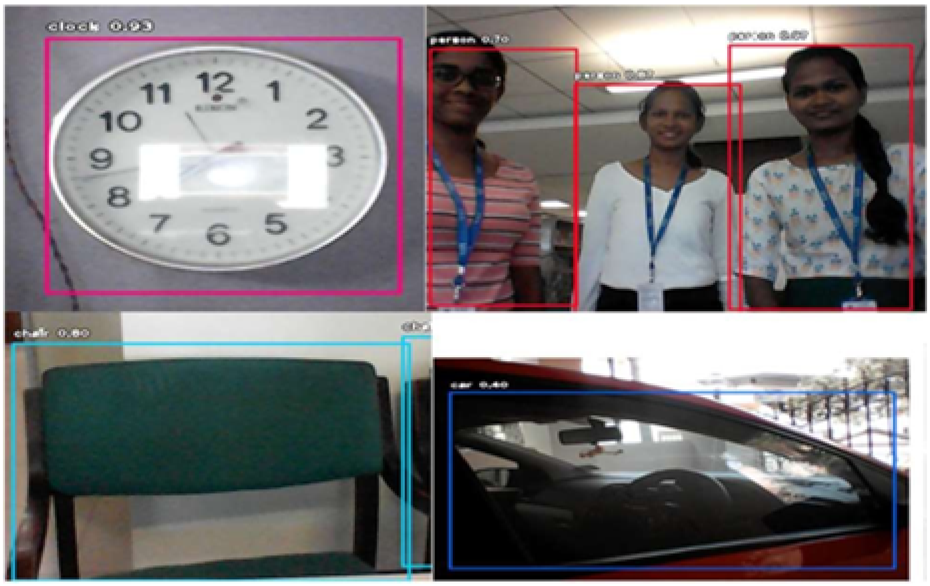

Fig.4. Detected Objects

Figure 4 illustrates the identified objects by the pre-trained YOLOTINYv3 algorithm so that these objects will now be transferred to speech and alert the people with vision complexities that there is a "car" in front of them so that they can move cautiously.

Advantages and Limitations of the Work

Object detection and recognition with YOLOTINYv3 will increase visually impaired individuals' independence, safety, and quality of life by recognizing things and obstacles in their environment quickly and precisely. This model will help visually impaired individuals by improving their capacity to accomplish everyday chores, reducing accidents, and offering faster and more accurate detection of objects and barriers, making the world more accessible to those with vision problems.

There are a few limitations to the study. Sometimes the recognition accuracy of object detection and recognition systems may occur, as these systems often struggle to identify objects in complex and cluttered environments. The object can be impacted by varying lighting conditions, which can lead to incorrect or inconsistent results. If the object is impacted, it creates a discomfort situation for people with vision complexities, as their visual perception may be impaired by low lighting or bright light sources.

Part 4: Financial Equation

To create a comprehensive financial equation for assistive technology for visually impaired users, we'll use the following elements:

1. Revenue Streams:
   - Subscription Fees (S)
   - Product Sales (P)
2. Costs:
   - Fixed Costs (F)
   - Variable Costs (V)
   - Costs related to Awareness and Outreach (A)
3. Market Growth:
   - Initial number of users ($x0$)
   - Growth rate (k)

Constants and Assumptions

Let's define hypothetical values for our constants and variables:

- Initial number of users $x0 = 1000$
- Growth rate $k = 200$ users per year
- Monthly subscription fee $S = 1,600$ INR
- One-time purchase price $P = 40,000$ INR
- Fixed costs $F = 16,000,000$ INR
- Variable cost per user $V = 4,000$ INR
- Awareness and Outreach cost per user $A = 2,000$ INR

Financial Equation:

The total profit $y(t)$ can be represented as:

y(t) = Revenue - Costs

Revenue:

Revenue = $(S+P).x(t)$

where $x(t) = x0 + kt$

Costs:

Costs = $F+(V+A).x(t)$

Final Financial Equation:

$y(t) = (S+P).(x0+kt) - [F+(V+A) \cdot (x0+kt)]$

Substitute the values:

$y(t) = (1{,}600 + 40{,}000) \cdot (1{,}000 + 200t) - [16{,}000{,}000 + (4{,}000 + 2{,}000) \cdot (1{,}000 + 200t)]$

Simplify the equation:

$y(t) = 41{,}600 \cdot (1{,}000 + 200t) - [16{,}000{,}000 + 6{,}000 \cdot (1{,}000 + 200t)]$

$y(t) = 41{,}600{,}000 + 8{,}320{,}000t - [16{,}000{,}000 + 6{,}000{,}000 + 1{,}200{,}000t]$

$y(t) = 41{,}600{,}000 + 8{,}320{,}000t - 22{,}000{,}000 - 1{,}200{,}000t$

$y(t) = 19{,}600{,}000 + 7{,}120{,}000t$

Final Simplified Financial Equation:

$\mathbf{y(t) = 19{,}600{,}000 + 7{,}120{,}000t}$

Interpretation:

- y(t): Total profit at time t (in INR)
- 19,600,000: Initial profit contribution from the existing user base and fixed costs
- 7,120,000: Profit contribution per year from the market growth

Considerations:

- Awareness and Outreach Programs: The cost A for awareness and outreach programs has been incorporated to address the limited awareness among individuals about assistive technologies.
- Growth Rate: The linear growth rate k reflects the increasing number of users due to various driving factors like increased awareness and funding by NGOs and healthcare agencies.

This equation provides a comprehensive model to forecast the financial outlook for assistive technology in a growing market, considering both revenues and costs, including factors for awareness and outreach programs.