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

Author: Amisha Gupta

Date: 27.04.2024

ABSTRACT

Vision is a crucial ability that enables people to view the world around them. Many visually impaired people struggle to feel and identify their surroundings, making it challenging for them to navigate and socialize in a new environment. This project will aid visually impaired persons by giving navigational assistance and aural feedback so they may more quickly discover specific products or pathways within a place, reducing their reliance on others. The major purpose is to help persons who are visually challenged, which helps everyone involved. This study addresses the notion of interpreting visual objects through hearing. The aural and optical senses have the capacity to find items in space. Therefore, this work utilized YOLOTINYv3 (You only look once) to recognize things within the bounding box, and the Pyttsx3 module will convert the object into speech.

Keywords: YOLO TINYv3, object detection, object recognition, visually impaired, pyttsx3, flask framework, text to speech

INTRODUCTION

In the science of computer vision, object detection and recognition is a crucial problem with several applications in fields including robotics, surveillance, and autonomous vehicles. However, for individuals with vision complexities such as low vision or blindness, the ability to detect and recognize objects in their environment can be particularly important for their daily independence and safety. Recent advancements in deep learning and computer vision have led to the development of powerful object detection algorithms such as YOLO (You Only Look Once). The YOLO technique is effective and precise for real-time object detection across several streams of photographs or footages. This makes YOLO a desirable option for systems that detect and identify objects for users with sophisticated vision.

However, developing a robust object detection and recognition system for people with vision complexities using YOLO is a challenging task. It requires a large and diverse dataset, to train the model with different object variations, lighting conditions, and backgrounds. Additionally, objects of interest for people with vision complexities may be different from general object detection tasks. The research in this field is relatively new, and there is a lot of room for improvement and innovation. For example, researchers can investigate how to improve the performance of YOLO for object detection in low-light and high-contrast environments, adapt YOLO for real-time object detection and recognition on resource-constrained devices such as smartphones or smart glasses, investigate the effectiveness of different techniques for providing feedback to individuals with vision complexities, and evaluate the impact of object detection and recognition systems on the daily lives of individuals with vision complexities.

Additionally, it is essential to consider the ethical and privacy implications when developing such systems, such as ensuring that the data collected is secure and protected. Overall, object detection and recognition for people with vision complexities using YOLO has the potential to greatly enhance the daily lives of individuals with vision complexities by providing them with more independence and safety.

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:

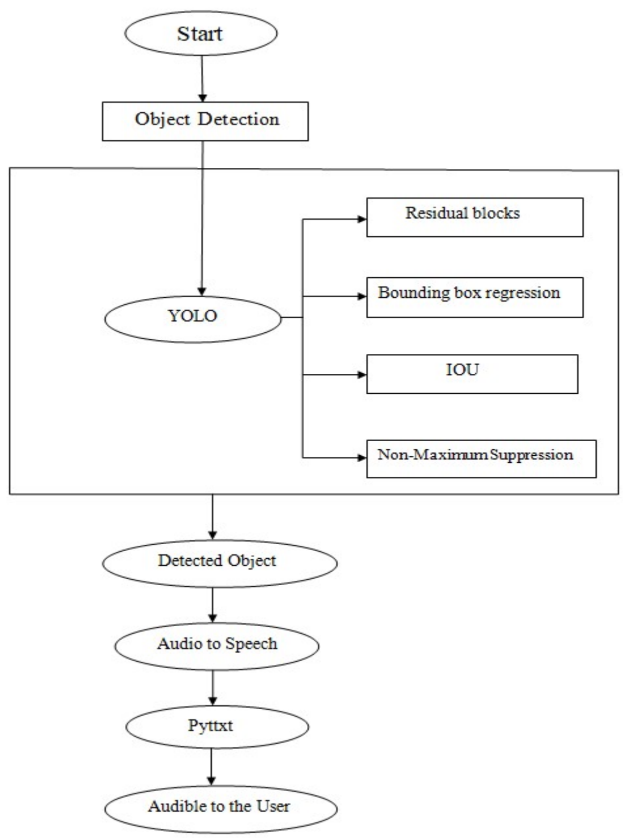

Fig.1. Workflow

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.

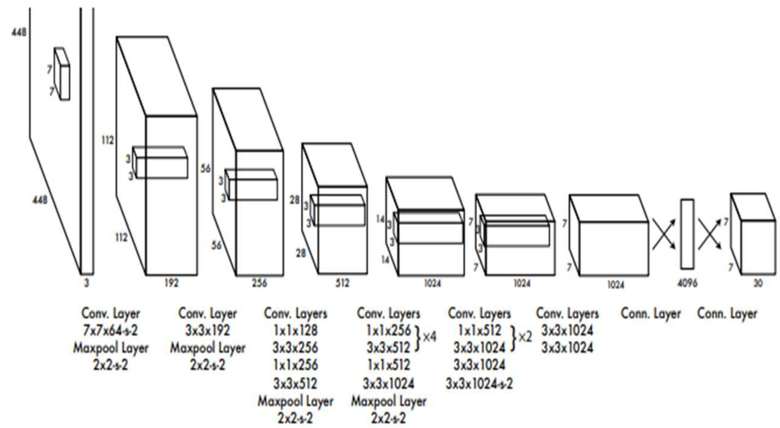

Fig. 2. YOLO Architecture

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.

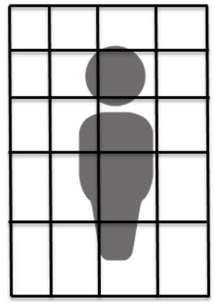

Fig.3. Image of 5×4 grid

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.

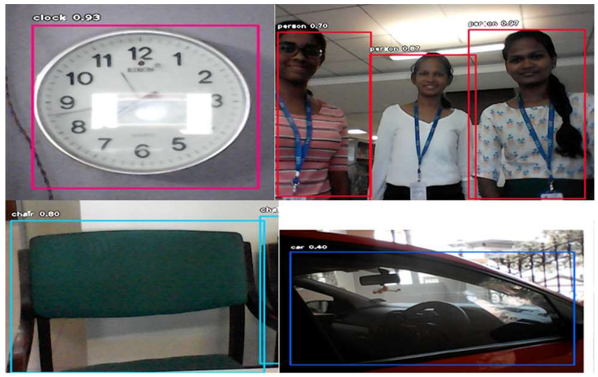

Fig.4. Detected Objects

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.

Conclusion

This project proposes a real-time object detection system using YOLO and text-to-speech technology to assist visually impaired individuals. The system enhances independence and safety, providing a valuable tool for daily navigation. Future work includes improving detection accuracy in varying lighting conditions and integrating the system into portable, user-friendly devices like smart glasses.

Objects in the actual world can now be recognized and identified by people with visual impairments. The recognized object will subsequently be converted into voice for persons who are visually impaired. With the help of this model, they can change the visual world into the auditory one. However, visually impaired persons may find it challenging to operate. Further, this work can be extended by using headphones to make it easy to recognize people and by using portable glasses to make it possible for the user to quickly detect objects.

References

1. Chandan, G., Jain, A., & Jain, H. (2018). Real-time object detection and tracking using Deep Learning and OpenCV. *2018 International Conference on Inventive Research in Computing Applications*.

2. Pintado, D., et al. (2019). Deep learning-based shopping assistant for the visually impaired. *2019 IEEE International Conference on Consumer Electronics*.

3. Nguyen, H., et al. (2020). Web-based object detection and sound feedback system for visually impaired people. *2020 International Conference on Multimedia Analysis and Pattern Recognition*.

4. Bhole, S., & Dhok, A. (2020). Deep learning-based object detection and recognition framework for the visually-impaired. *2020 Fourth International Conference on Computing Methodologies and Communication*.

5. Vaidya, S., et al. (2020). Real-time object detection for visually challenged people. *2020 4th International Conference on Intelligent Computing and Control Systems*.

6. Tosun, S., & Karaarslan, E. (2018). Real-time object detection application for visually impaired people: Third eye. *2018 International Conference on Artificial Intelligence and Data Processing*.

7. Shahira, K. C., et al. (2019). Obstacle detection, depth estimation and warning system for visually impaired people. *TENCON 2019 IEEE Region 10 Conference*.

8. Yadav, S., et al. (2020). Fusion of object recognition and obstacle detection approach for assisting visually challenged person. *2020 43rd International Conference on Telecommunications and Signal Processing*.

9. Iqbal, A., et al. (2022). A comprehensive assistive solution for visually impaired persons. *2022 2nd International Conference of Smart Systems and Emerging Technologies*.

10. Afif, M., et al. (2020). Indoor objects detection and recognition for an ICT mobility assistance of visually impaired people. *Multimedia Tools and Applications*.

11. Hisham, Z. A. N., et al. (2020). A Systematic Literature Review of the Mobile Application for Object Recognition for Visually Impaired People. *2020 8th International Conference on Information Technology and Multimedia*.

12. Dionisi, A., et al. (2012). Wearable object detection system for the blind. *2012 IEEE International Instrumentation and Measurement Technology Conference*.

13. Jabnoun, H., et al. (2015). Object detection and identification for blind people in video scene. *2015 15th International Conference on Intelligent Systems Design and Applications*.

14. Mainkar, V. V., et al. (2020). Raspberry Pi based intelligent reader for visually impaired persons. *2020 2nd International Conference on Innovative Mechanisms for Industry Applications*.