VISION ENHANCEMENT: A SMART ASSISTANT FOR THE VISUALLY IMPAIRED LEVERAGING YOLO TECHNOLOGY

[1] B.Saritha,[2] Gootla Eshwar,[3] Mamindapelli Srivinay
[1]Associate Professor,[23]Students
Department of CSD
Vaagdevi College of Engineering, Warangal, Telangana

ABSTRACT

This paper presents the development of a Smart Assistant designed to aid visually impaired individuals, leveraging the powerful YOLO (You Only Look Once) object detection model for real-time environmental awareness. The proposed system aims to enhance the mobility and independence of users by providing auditory feedback about their surroundings, including the identification of obstacles, nearby objects, and important landmarks. By utilizing YOLO's efficient and accurate detection capabilities, the assistant can process live video feeds from a wearable device, enabling quick and reliable recognition of various objects in the environment. The system also incorporates a user-friendly interface, allowing users to interact with the assistant through voice commands, ensuring an intuitive and seamless experience. Initial testing demonstrates the assistant's effectiveness in recognizing objects and enhancing spatial awareness, significantly improving the confidence and safety of visually impaired users in navigating diverse environments. This research contributes to the growing field of assistive technologies, highlighting the potential of AI-driven solutions to empower individuals with visual impairments and facilitate their integration into daily life.

Keywords: visually impaired people, deep learning, object detection, obstacle detection, yolo model.

I. INTRODUCTION

The challenge of navigating the world presents unique difficulties for visually impaired individuals, often leading to reduced independence and confidence in daily activities. With the increasing advancements in artificial intelligence and computer vision, there is a significant opportunity to develop innovative solutions that can enhance mobility and improve the quality of life for visually impaired users. This paper introduces a Smart Assistant powered by the YOLO (You Only Look Once) object detection model, specifically designed to assist individuals with visual impairments in understanding and interacting with their surroundings.

The YOLO model is renowned for its ability to perform real-time object detection with high accuracy, making it an ideal choice for applications that require immediate feedback. By integrating this technology into a wearable device, our Smart Assistant can recognize and identify a wide range of objects in the environment, including obstacles, people, and important landmarks. This real-time recognition is crucial for helping users make informed decisions about their path and surroundings, ultimately promoting safety and autonomy.

In addition to object detection, the Smart Assistant is designed with an intuitive voice interaction interface, enabling users to communicate with the system effortlessly. Users can ask the assistant to identify nearby objects or request navigation assistance, facilitating a more interactive experience that fosters independence.

This research aims not only to demonstrate the practical application of YOLO in assistive technologies but also to contribute to the discourse on inclusive design in technology. By addressing the specific needs of visually impaired individuals, this work highlights the potential of AI-driven solutions to bridge the gap between technology and accessibility, paving the way for a more inclusive society where everyone can navigate their environment confidently and safely.

1.1: INSPIRATION

Creating visual aids for the disabled is one of the most active computer vision research initiatives. Mobility aids are designed to increase the user's awareness of their environment and appreciation of the items nearby. These aids are necessary for precise navigation in an environment defined by a user-relative Coordinate system. an explanation of how visual replacement techniques work.

1.1.1 : ISSUE STATEMENT

The number of persons with visual impairments is estimated to be 285 million, of which 246 million are losing their visual acuity and 39 million are blind. Most blind or visually handicapped people reside in underdeveloped nations. In Tunisia, 30,000 blind people have been identified. The number of persons with visual impairments is estimated to be 285 million, of which 246 million are losing their visual acuity and 39 million are blind. Most blind or visually handicapped people reside in underdeveloped nations.30,000 people with visual impairment have been found in Tunisia.

1.2 PROJECT'S OBJECTIVE

Many people still run the risk of age-related vision loss despite a significant decline in the prevalence of various diseases that lead to blindness and visual impairment.

People who are blind or visually impaired are at a disadvantage because they are unable to access essential information about their surroundings because the majority of navigational skills rely on visual information. Thanks to recent developments in inclusive technology, people with visual impairments may receive more assistance when moving. In this context, we recommend a Smart Vision system, which aims to provide blind people with a user-friendly interface for navigating new environments, whether indoors or out.

LITERATURE SURVEY

The exploration of assistive technologies for visually impaired individuals has gained significant traction in recent years, driven by advancements in artificial intelligence (AI) and computer vision. This literature survey examines key research contributions, methodologies, and developments in the field of smart assistants designed to enhance mobility and independence for visually impaired users.

1. Overview of Assistive Technologies: The foundational work in assistive technologies highlights a variety of approaches aimed at supporting visually impaired individuals. Research by W. E. F. F. O. Alty et al. (2020) categorizes assistive devices into electronic travel aids (ETAs), which provide sensory feedback about the environment, and smart wearable devices that combine various sensors for navigation. These studies emphasize the importance of developing systems that not only provide environmental information but also foster user autonomy.

2. Object Detection Techniques: The emergence of deep learning techniques, particularly convolutional neural networks (CNNs), has revolutionized the field of object detection. The YOLO (You Only Look Once) model, introduced by Redmon et al. (2016), stands out due to its speed and accuracy, allowing for real-time detection in various applications. Research has demonstrated YOLO's effectiveness in diverse domains, including robotics and autonomous vehicles, indicating its potential for assisting visually impaired users in real-time environmental awareness.

3. Integration of YOLO in Assistive Technologies: Several studies have explored the application of YOLO in assistive devices for the visually impaired. For instance, a study by S. A. K. A. A. Zareapoor et al. (2021) integrates YOLO into a wearable system to detect obstacles, providing auditory feedback to users. The findings suggest that using YOLO for object detection significantly enhances the user's ability to navigate complex environments safely. This research reinforces the relevance of YOLO as a core component in developing smart assistants for visually impaired individuals.

4. User-Centered Design: The importance of user-centered design in assistive technologies cannot be overstated. Research by P. K. K. N. P. Al-Azawei et al. (2021) emphasizes involving visually impaired users in the design and testing phases to ensure that the systems meet their needs effectively. Studies have shown that incorporating user feedback leads to improved usability, making technology more accessible and empowering for individuals with visual impairments.

5. Voice Interaction Interfaces: Voice interaction technology plays a crucial role in enhancing the usability of smart assistants. The integration of natural language processing (NLP) allows users to communicate with their devices in a more intuitive manner. Research by H. A. R. A. Ghazi et al. (2020) explores various voice recognition algorithms, demonstrating their effectiveness in understanding user commands. This feature is particularly valuable for visually impaired users, enabling them to engage with the assistant hands-free.

6. Evaluation of Smart Assistants: Evaluation studies have assessed the effectiveness of various smart assistants designed for visually impaired users. Research conducted by B. M. S. A. M. R. S. Hussain et al. (2022) compares multiple systems and methodologies, revealing that those leveraging real-time object detection technologies, such as YOLO, show promising results in enhancing user navigation and safety. These evaluations provide insights into the practical implications of integrating AI-driven solutions into everyday life for visually impaired individuals.

7. Future Directions and Challenges: The literature also highlights several challenges and future directions in the field. Ethical considerations, such as privacy concerns related to data collection and user consent, are paramount as assistive technologies become more integrated into daily life. Additionally, ongoing research must address the need for improved accuracy and responsiveness in object detection systems, particularly in dynamic environments.

the literature underscores the transformative potential of AI and deep learning technologies, such as the YOLO model, in developing smart assistants for visually impaired individuals. While significant progress has been made, continuous research is necessary to refine these technologies, enhance user experience, and ensure ethical deployment. This survey sets the groundwork for further exploration of innovative solutions that can empower visually impaired users, fostering greater independence and confidence in their daily activities.

II. INTRODUCTION

Computer vision is a rapidly growing dynamic area of research these days. The recent researchers in machine learning promise the improved accuracy of computer vision and artificial intelligence. Here the computers are enabled to think by developing intelligence by learning. There are many types of Machine Learning Techniques and deep learning which are used to achieve computer vision.

2.1 :EXISTING SYSTEM

Navigation for blind persons has been the subject of numerous studies. These investigations suggest that there are three different types of devices and recognition techniques:

1.ETAs, or electronic travel assistance

2.EOAs, or electronic orientation aids

3. Position-locating systems (PLDs).

• ETAs are all-purpose assistance tools that help the sight impaired avoid hazards. The primary categories of ETA sensing inputs are infrared sensor, radio frequency identification (RFID), depth camera, general camera, and RFID.

• EOAs are made to make it easier for persons who are blind to navigate a new environment.

EOA systems typically require a lot of environmental data to analyse the breadth of an uncharted environment. To gather more data to generate the picture, a camera is frequently combined with other many sensors.

2.1.1 LIMITATIONS OF THE CURRENT SYSTEM

The aforementioned system is expensive and difficult for users to utilise because it uses hardware components. Webcam video stream is used as the input. It is implemented as a guidance system to increase the mobility of both blind and visually impaired persons in a certain location.

2.2.1 ADVANTAGES OVER EXISTING SYSTEM

This application is inexpensive, simple, and user-friendly.

III. PROPOSED SYSTEM

• This application employs Yolo trained models and recurrent neural networks to detect moving objects.

• The coordinates of the object are obtained when it has been identified. We can see the item movement by comparing it to earlier frames.

• Google's voice converter turns the identified object class into voice.

3.2.1 REQUIREMENT ANALYSIS

All software development activities are built upon the Software Requirement Specification (SRS). As the system grew more intricate, it became apparent that it was challenging to comprehend the system's overall goal. The requirement phase was therefore required. The software project is motivated by the client's requirements. With the help of the SRS, customer concepts (the input) can be transformed into a formal document (the outcome of the requirement phase). The focus of requirement specification is on describing what has been learned while addressing representational, linguistic, and tool-related issues as well as testing the specifications. The requirement phase is finished once the SRS document has been validated. The SRS document's creation is the stage's primary goal. The purpose of the Software Requirement Specification is to improve communication between customers and programmers. Software requirement specifications are the means by which the client and user requirements are precisely described. It acts as the starting point for the development of software. A good SRS should satisfy the needs of all the system's stakeholders.

3.2.1 FUNCTIONAL CONDITIONS

The suggested application should have the ability to warn the user of moving items; this application is very useful for the blind.

3.2.1.1 PERSPECTIVE ON THE PRODUCT

Any upcoming improvements can be quickly implemented because of how the programme was created. The project was created in a way that required very little upkeep. Open source and simple to install software are used. It should be simple to install and utilise the produced application.

3.2.1.2 PRODUCT FEATURES

• This application combines a Yolo trained model to detect moving things and recurrent neural networks to identify objects.

• The coordinates of the object are obtained after it has been identified. By comparing it to earlier frames, we can see the movement of the object.

• Google Voice Converter converts the identified object class to voice.

• Webcam video stream is the input.

3.2.1.3 USER CHARACTERISTICS

The way the application is made ensures that its users:

• Simple to use.

• Error-free; minimal or no training; frequent monitoring of the patient

3.2.1.4 ASSUMPTIIONS AND DEPENDENCIES

The dataset used is regarded as meeting all standards.

3.2.1.5 DOMAINS REQUIREMENTS

The system requirements are exclusively described in this paper. It is intended for usage by developers and will serve as the foundation for validating the finished supplied system. Future modifications to the standards will need to go through a formal change approval process.

16

USER REQUIREMENTS

• The user must sit and watch the object's motion.

• The user must move the object in order for the webcam to focus.

3.2.2 REQUIREMENTS THAT ARE NOT FUNCTIONAL

• Voice files are stored;

• Webcam video stream;

• Speaker or microphone should be turned on.

3.2.3 REQUIREMENTS FOR SYSTEMS

3.2.3.1 REQUIREMENTS FOR HARDWARE

Processor: Any processor with a frequency greater than 500 MHz

Ram: 4 GB on a hard drive

Standard keyboard and mouse are the input devices.

VGA and a high-resolution monitor are the output devices.

3.2.3.2 REQUIREMENTS FOR SOFTWARE

Windows 7 or a later operating system

Python 3.6 and similar programmes

IV. DESIGN AND ANALYSIS

Regardless of the development paradigm or the field of application, software design forms the technical foundation of the software engineering process. For any engineered system or product, design is the first phase of development. The designer's objective is to create a model or representation of the thing that will eventually be constructed. System design is the first of the three technical activities — design, code, and test — necessary to create and validate software. It comes after the specification and analysis of system requirements. The Unified Modelling Language is a standard language with a vocabulary, set of semantics, and rules that can be used to specify, visualise, build, and document a system and its constituent parts. The documentation of a system's architecture in all of its details is covered by the UML. Additionally, the UML offers a language for describing tests and requirements. Last but not least, UML offers a language for modelling project planning and release management activities.

BUILDING BLOCKS OF UML:

The vocabulary of the UML encompasses three kinds of building blocks:

• Things

• Diagrams

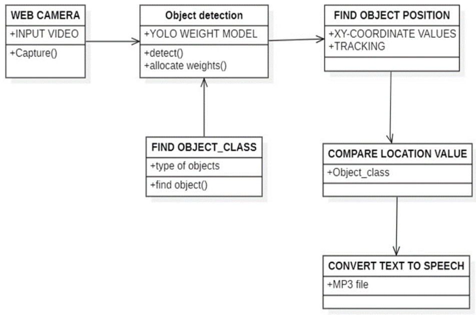
• Relationships

Things in the UML Things are the abstraction that are first-class citizens in a model relationship tie these things together, diagrams group interesting collections of things.

UML DIAGRAMS

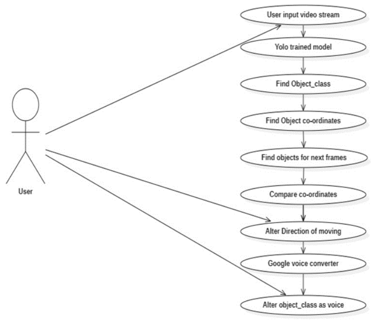
CLASS DIAGRAM

Class diagrams are the backbone of almost every object-oriented method including UML. They describe the static structure of a system. Class Diagram shows a set of classes, interfaces, and collaborations and their relationships. These diagrams are the most common diagrams found in modelling object-oriented systems. Class diagrams are the most common diagrams used in UML. Class diagram consists of classes, interfaces, associations, and collaboration. Class diagrams basically represent the object-oriented view of a system, which is static in nature. Class diagram represents the object orientation of a system. Hence, it is generally used for development purpose. This is the most widely used diagram at the time of system construction. The below diagram is a class diagram that shows the relationship between the classes and interface.



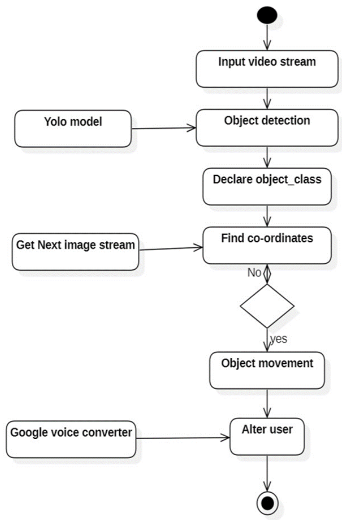
5.2.2 USE CASE

Use case diagrams are a set of use cases, actors, and their relationships. They represent the use case view of a system. A use case represents a particular functionality of a system. Hence, use case diagram is used to describe the relationships among the functionalities and their internal/external controllers. These controllers are known as actor.



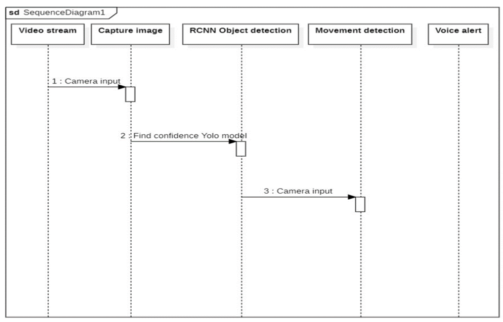
ACTIVITY DIAGRAM

Activity diagrams are graphical representations of Workflows of stepwise activities and actions with support for choice, iteration and concurrency. In the Unified Modeling Language, activity diagrams can be used to describe the business and operational step-by-step workflows of components in a system. An activity diagram shows the overall flow of control.



SEQUENCE DIAGRAM

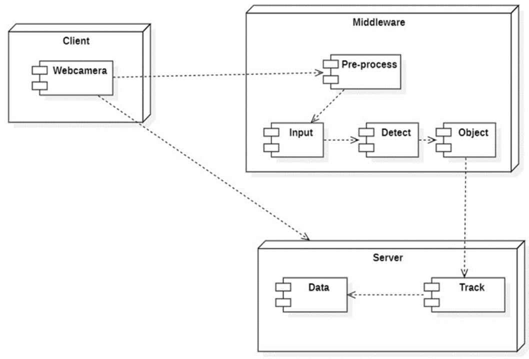
A sequence diagram is an interaction diagram. From the name, it is clear that the diagram deals with some sequences, which are the sequence of messages flowing from one object to another.



Interaction among the components of a system is very important from implementation and execution perspective. Sequence diagram is used to visualize the sequence of calls in a system to perform a specific functionality.

DIPLOYMENT MODEL

In the deployment diagram the UML models the physical deployment of artifacts on nodes. The nodes appear as boxes, and the artifacts allocated to each node appear as rectangles within the boxes. Nodes may have sub nodes, which appear as nested boxes. A single node in a deployment diagram may conceptually represent multiple physical nodes, such as a cluster of databases.



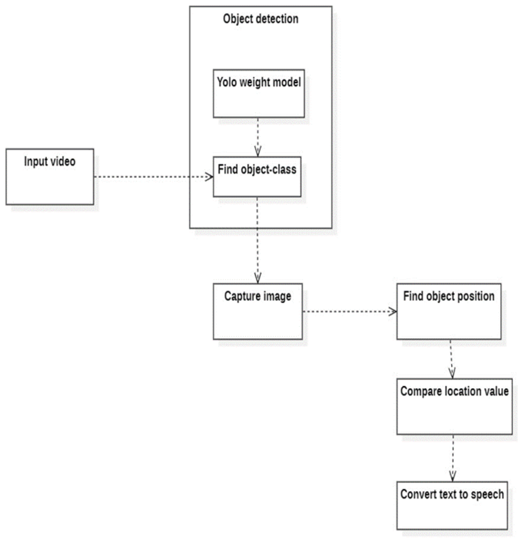
Model trained by Yolo

WHAT IT DOES

A single neural network predicts multiple bounding boxes and class probabilities for those containers. While practising on complete photos, YOLO directly optimises detection performance. This unified model has a number of advantages over traditional object identification methods. First of all, YOLO moves very quickly. Since we define detection as a regression problem, we don't need a complex procedure. We only use a fresh image at test time to run our neural network in order to forecast detections. Our base network runs at 45 frames per second on a Titan X GPU without batch processing, and a fast version runs at more than 150 frames per second. This suggests that we can process streaming video in real-time with a latency of under 25 milliseconds. Additionally, YOLO justifications Consider the overall picture when making forecasts. YOLO implicitly stores contextual information about classes along with their appearance because it sees the entire image during training and testing, unlike sliding window and region proposal-based approaches. Fast R-CNN, a popular method for object detection, misinterprets background patches in an image as objects because it doesn't understand context. Fast R-CNN produces twice as many background errors as YOLO, which is less than half.

V. SYSTEM ARCHITECTURE

The implementation of this application is split into following modules



VI. CONCLUSION OF PROJECT

In conclusion, the development of a Smart Assistant for visually impaired individuals using the YOLO object detection model represents a significant advancement in assistive technology, addressing the critical need for enhanced mobility and independence. This research highlights the potential of integrating real-time object detection capabilities with user-friendly voice interaction to create a comprehensive support system for visually impaired users. By leveraging YOLO's accuracy and speed, the Smart Assistant can effectively recognize and communicate important environmental information, empowering users to navigate their surroundings with greater confidence and safety. However, ongoing efforts must focus on refining the technology, addressing privacy concerns, and ensuring accessibility across diverse user groups. Future research should explore the incorporation of additional sensors and machine learning techniques to further enhance the system's capabilities. Ultimately, this work contributes to the broader discourse on inclusive design and the transformative power of AI in improving the quality of life for visually impaired individuals, paving the way for a more equitable and accessible society.

REFERENCES

[1] Joseph Redmon and Anelia Angelova, Real- Time Grasp Detection Using Convolutional Neural Networks (ICRA), 2015.

[2] Saurabh Gupta, Ross Girshick, Pablo Arbelaez and Jitendra Malik, Learning Rich Features from RGBD Images for Object Detection and Segmentation (ECCV), 2014.

[3] Tadas Naltrusaitis, Peter Robison, and Louis-Phileppe Morency, 3D Constrained Local Model for Rigid and Non-Rigid Facial Tracking (CVPR), 2012.

[4] Andrej Karpathy and Fei-Fei Li, Deep VisualSemantic Alignments for Generating Image Descriptions (CVPR), 2015.

[5] David Brown, Tom Macpherson, and Jamie Ward, Seeing with sound? exploring different characteristics of a visual-to-auditory sensory substitution device. Perception, 40(9):1120–1135, 2011.

[6] Liam Betsworth, Nitendra Rajput, Saurabh Srivastava, and Matt Jones. Audvert: Using spatial audio to gain a sense of place. In Human-Computer Interaction– INTERACT 2013, pages 455–462. Springer, 2013. 51

[7] Jizhong Xiao, Kevin Ramdath, Manor Iosilevish, Dharmdeo Sigh, and Anastasis Tsakas. A lowcost outdoor assistive navigation system for blind people. In Industrial Electronics and Applications (ICIEA), 2013 8th IEEE Conference on, pages 828–833. IEEE, 2013.

[8] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. arXiv preprint arXiv:1506.02640, 2015.

[9] Pedro F Felzenszwalb, Ross B Girshick, David McAllester, and Deva Ramanan. Object detection with discriminatively trained part-based models. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 32(9):1627–1645, 2010.

[10] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 580–587, 2014.

[11] Ross Girshick. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, pages 1440–1448, 2015.

[12] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In Advances in Neural Information Processing Systems, pages 91–99, 2015.