Real Time Object Detection for Blind And Mute Peoples Based on YOLO Technology

Shubham Kamble[1], Vishal Ghaytadak[2], Aditya Nadgouda[3], Aditya Chavan[4], Prof. Wagh. D. B[5]
Department of Computer Engineering
[1,2,3,4]UG Student, SVPM's College of Engineering Malegaon(bk), Baramati
[5] SVPM's College of Engineering Malegaon(bk), Baramati

Abstract- *In an innovative expansion, the project addresses communication challenges faced by mute individuals and visually impaired individuals by developing a system that combines computer vision and natural language processing. By integrating Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), the system interprets and translates hand gestures captured by the camera into meaningful messages. Optical Character Recognition (OCR) technology is integrated to convert text from images into machine-readable formats. The portable and user-friendly system captures real-time video feed through a camera-equipped device, processes it for object detection, and utilizes OCR for text extraction. Audio feedback, delivered through text-to-speech technology, assists blind individuals in comprehending printed materials and navigating their surroundings effectively. This dual-purpose system not only facilitates real-time text and object recognition for the visually impaired but also establishes a novel communication channel for mute individuals. Regular testing, a user-friendly interface, and ongoing user feedback are crucial for refining the system's accuracy and reliability in diverse real-world scenarios.*

This project addresses communication challenges faced by mute and visually impaired individuals through an innovative system combining computer vision and natural language processing. The system's architecture involves deep learning frameworks such as TensorFlow or PyTorch, utilizing models like YOLO or SSD for real-time object detection. This dual-purpose system not only facilitates real-time text and object recognition for the visually impaired but also establishes a novel communication channel for mute individuals. Regular testing, a user-friendly interface, and ongoing user feedback play a crucial role in refining the system's accuracy and reliability in diverse real-world scenarios. The system's architecture prioritizes portability, accuracy, and user-friendliness, aligning with the project's goal of creating an inclusive society where technology empowers individuals with diverse abilities to navigate the world autonomously and confidently.

Keywords *— Real Time Object Detection, Blind and Deaf individuals, CNN - RNN algorithm's, Single Shot Multi-Box Detector, Natural Language Processing, Optical Character Reorganization, YOLO technology, Computer vision, Natural Language Processing, Gesture Interpretation, Text-to-speech technology, DL frameworks, TensorFlow, PyTorch, Visually impaired, Mute individuals, Dual-purpose system, Audio feedback, Object recognition, System architecture, Methodology, Equality, Autonomy, Human-computer interaction, Assistive communication, Novel communication channel, Regular testing, Diverse real-world scenarios.*

I. INTRODUCTION

In the current era of rapid technological evolution, addressing the challenges faced by diverse communities is crucial. Individuals with visual and speech impairments encounter unique obstacles in daily life, navigating a world that can be particularly daunting for the visually impaired, compounded by difficulties in identifying text on handheld objects. Simultaneously, those with speech impairments face limited communication avenues. The project "Real-time Object Detection Using YOLO Technology" seeks to pioneer a comprehensive solution, aiding the visually impaired in real-time text recognition and introducing a groundbreaking feature for mute individuals' communication needs. The project addresses the lack of real-time text detection tools for handheld objects by integrating cutting-edge computer vision and natural language processing techniques, utilizing frameworks like TensorFlow or PyTorch. Models such as YOLO and SSD enable real-time object detection, complemented by Optical Character Recognition (OCR) technology to convert text from images into machine-readable formats. Beyond assisting the visually impaired, the project recognizes the communication challenges of mute individuals and introduces a hand gesture-based communication system. Using Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), this feature interprets and translates hand gestures into meaningful messages. The project envisions a more inclusive society, seamlessly combining real-time object detection and hand gesture-based communication.

In our current era, characterized by rapid technological evolution, the imperative to create inclusive solutions for diverse communities has become more pressing than ever. Among those confronting unique challenges in daily life are individuals with visual impairments and speech impairments. Navigating the world is especially formidable for the visually impaired, compounded by the difficulty of identifying and comprehending text on handheld objects. Simultaneously, those with speech impairments often find themselves contending with limited communication avenues. Recognizing the intersectionality of these challenges, the project "Real-time Object Detection Using YOLO Technology" seeks to pioneer a comprehensive solution that not only assists the visually impaired in real-time text recognition but also introduces a groundbreaking feature tailored to the communication needs of mute individuals.

The absence of real-time text detection and recognition tools specifically tailored for handheld objects exacerbates the difficulties faced by the visually impaired. By seamlessly combining real-time object detection and hand gesture-based communication, the project envisions a more inclusive society where technology acts as a bridge, addressing the diverse needs of individuals with visual and speech impairments.

Furthermore, the system provides audio feedback via text-to-speech technology, aiding blind individuals in understanding printed materials and navigating their surroundings effectively. It serves a dual purpose by not only facilitating real-time text and object recognition for the visually impaired but also establishing a novel communication channel for mute individuals.

Overall, the project prioritizes portability, accuracy, and user-friendliness, aiming to create an inclusive society where technology empowers individuals with diverse abilities to navigate the world autonomously and confidently.

This dual-purpose system is not merely a technological advancement; it symbolizes a commitment to social inclusivity. By providing instant and reliable information about the text content on handheld objects and offering a communication channel for mute individuals, the project aspires to contribute significantly to autonomy, inclusion, and accessibility. Regular testing, user feedback, and refinement are integral aspects of this journey, ensuring that the system evolves to meet the dynamic challenges of real-world scenarios. Ultimately, the project aims to be a beacon of technological innovation, demonstrating that advanced solutions can be both sophisticated and deeply human-centered, fostering a society where technology empowers individuals of all abilities.

II. OBJECTIVE

It aims to develop an advanced solution to address the specific challenges faced by individuals with visual and speech impairments. The primary goal is to enhance autonomy and inclusion for the visually impaired through the application of cutting-edge computer vision and natural language processing techniques.

In addition to aiding the visually impaired, the project extends its impact to mute individuals by introducing an innovative hand gesture-based communication system. The overarching goal is to create a dual-purpose system seamlessly integrating real-time object detection and hand gesture-based communication.

Through this approach, the project aims to significantly contribute to building a more inclusive society where technology serves as a bridge, catering to the diverse needs of individuals with visual and speech impairments. The project's success will be measured by its capacity to provide instant information, offer a reliable communication channel, and undergo continuous refinement based on user feedback, ensuring its practical applicability in real-world scenarios.

Developing a Comprehensive System:

The primary aim of the project is to pioneer the creation of a sophisticated system that addresses the unique challenges faced by individuals with visual and speech impairments. This involves integrating state-of-the-art computer vision and deep learning techniques, utilizing models such as YOLO or SSD for real-time object detection and CNN and RNN algorithms for hand gesture-based communication.

Conducting Rigorous Testing for Evaluation:

Rigorous testing is a critical component of the project, conducted under diverse real-world scenarios. This evaluation process aims to assess the accuracy, speed, and responsiveness of the system. Regular testing ensures that the developed features meet the dynamic challenges of real-world usage.

Ongoing Refinement Based on User Feedback:

A commitment to continuous improvement is embedded in the project's objectives. Ongoing refinement based on user feedback is prioritized, ensuring that the system evolves to meet the specific needs and preferences of its users.

This iterative refinement process is crucial for the system's adaptability and user satisfaction.

III. LITERATURE SURVEY

Title:

Real Time Text Detection and Recognition on Hand Held Objects to Assist Blind People

Author:

Samruddhi Deshpande, Ms. Revati Shriram

Description:

This paper presents Camera based system which will help blind person for reading text patterns printed on hand held objects. This is the framework to assist visually impaired persons to read text patterns and convert it into the audio output. To obtain the object from the background and extract the text pattern from that object, the 2 system first proposes the method that will capture the image from the camera and object region is detected. The text which are maximally stable are detected using Maximally Stable External Regions feature.

Title:

Real-Time Object Detection for Visually Challenged People.

Author:

Sunit Vaidya, Naisha Shah, Prof. Radha Shankar

Description:

The application uses image processing and machine learning techniques to determine real-time objects through the camera and inform blind people about the object and its location through the audio output. Inability to differentiate between objects has led to many limitations to the existing approach which includes less accuracy and lower performance results. The main objective of the system is to provide good accuracy, best performance results and a viable option for the visually impaired people to make the world a better place for them.

Title:

Real Time Object Detection with Audio Feedback using Yolo.

Author:

Mansi Mahendru, Sanjay Kumar Dubey

Description:

With the rapid development of deep learning many algorithms were improving the relationship between video analysis and image understanding. All these algorithms work differently with their network architecture but with the same aim of detecting multiple objects within complex image.

This paper proposes a system that will detect every possible day to day multiple objects on the other hand prompt a voice to alert person about the near as well as farthest objects around them.

Title:

YOLO v5 for SDSB Distant Tiny Object Detection

Author:

Xiaodong Yu, Ta Wen Kuan, Yuhan Zhang

Description:

In this work, three categorized target objects including, fallen leaves, speed bumps and manhole cover etc. are involved in training and validation phases. To reach SDSB with real-time object detection, wherein Yolo v5s in terms of its benefits on lower frame rate, high-accuracy, and high-speed characteristics for real-time object detection.

IV. METHODOLOGY

The methodology adopted for the "Real-time Object Detection Using YOLO Technology" project involved a systematic and purposeful approach, combining various research methods to achieve the project's overarching objectives.

The research primarily focused on the implementation and integration of advanced algorithms, including Convolutional Neural Network (CNN), Optical Character Recognition (OCR), and Single Shot Multi-Box Detector (SSD), each chosen for its specific strengths in real-time object detection, text recognition, and efficient processing.

The implementation of the Convolutional Neural Network (CNN) commenced with a comprehensive analysis of its architecture. Inspired by the organization of the Visual Cortex and designed with three layers (Input, Hidden, and Output), the CNN was selected for its ability to assign importance to different aspects/objects in an image.

The algorithm's capacity to learn filters/characteristics with training made it a suitable choice for real-time object detection and hand gesture-based communication.

For text recognition, Optical Character Recognition (OCR) was employed as a smart reader for images. Leveraging deep learning techniques, including CNNs and RNNs, OCR was chosen for its ability to accurately recognize characters in images. The process involved stages such as picture input, clean-up, finding words, breaking down, guessing letters, and assembling the recognized characters to form readable text.

In the realm of real-time object detection, the project utilized the Single Shot Multi-Box Detector (SSD). SSD's design for object detection in real-time, without the need for a region proposal network.

1. Convolutional Neural Network (CNN)

Objective - Implementing a CNN for real-time object detection and hand gesture-based communication.

Architecture Analysis - CNN's architecture mimics the connectivity pattern of neurons in the human brain, inspired by the organization of the Visual Cortex. Three layers, including input layers, hidden layers, and output layers, facilitate the learning and differentiation of various aspects in images.

2. Optical Character Recognition (OCR)

Objective - Utilizing OCR for converting text from images into machine-readable formats.

Working Flow - OCR operates as a smart reader, it involves steps like picture input, clean-up, finding words, breaking down, guessing letters, and putting it together to recognize characters and convert images to readable text.

3. Single Shot Multi-Box Detector (SSD)

Objective - Implementing SSD for efficient real-time object detection with a focus on speed and accuracy.

Working Flow - SSD operates by looking at the entire picture at once, predicting boxes around objects, naming objects, and achieving quick and accurate results. The algorithm uses multiscale feature maps, default boxes, predictions at multiple scales, and non-maximum suppression to handle objects of different sizes effectively.

4. System Integration and Multiscale Features

Objective - Integrating the developed algorithms into a cohesive system for real-time object detection.

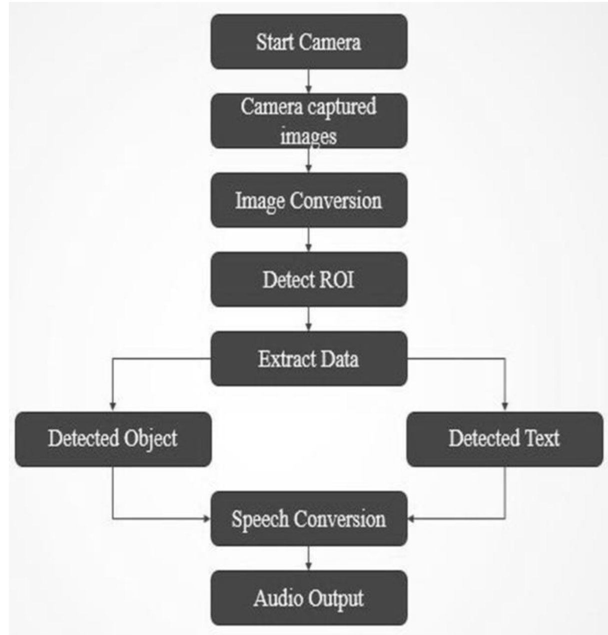
Implementation Details - The algorithms, including CNN, OCR, and SSD, are seamlessly integrated into a unified system. Multiscale feature maps from SSD capture information at various levels of granularity, allowing the system to detect objects of different sizes effectively.

5. User Testing and Feedback

Objective - Evaluating system performance through user testing and obtaining valuable feedback.

Iterative Refinement -. User feedback is actively solicited and used for iterative refinement, ensuring the system adapts to user needs and preferences.

V. SYSTEM ARCHITECTURE



Architecture

1. Input Layer:

The system begins with an Input Layer, where real-time video feed from a camera-equipped device is captured. This serves as the input to the subsequent layers and facilitates the recognition of objects, text, and hand gestures in the environment.

2. Integration Layer:

The outputs from the CNN, OCR, and SSD layers are integrated into a unified system. This integration ensures that the system combines the capabilities of real-time object detection, text recognition, and hand gesture-based communication seamlessly.

3. User Interface:

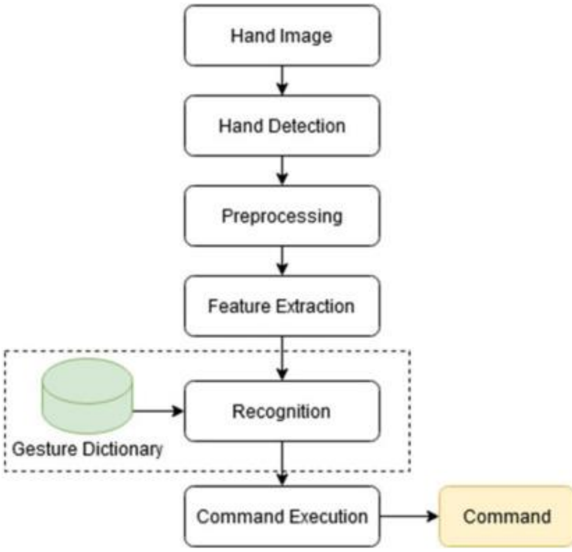
The system features a user-friendly interface designed for accessibility on low-cost devices. The interface facilitates easy navigation and interaction with the system, promoting autonomy for users with diverse abilities.

4. Text-to-Speech Module:

Audio feedback is provided through a Text-to-Speech (TTS) module. This module translates the recognized text into spoken words, aiding visually impaired individuals in comprehending printed materials and navigating their surroundings effectively.

5. Communication Module:

For mute individuals, the system includes a communication module that interprets and translates hand gestures captured by the camera into meaningful messages. This innovative feature establishes a novel communication channel, enhancing the project's scope for inclusivity.



Gesture Detection Flow

VI. PROPOSED SYSTEM

The proposed Real-time Text and Object Detection system for blind individuals aims to leverage computer vision and natural language processing techniques. Deep learning frameworks such as TensorFlow or PyTorch will be used to develop an object detection model, possibly YOLO or SSD, trained on a diverse dataset. Simultaneously, Optical Character Recognition (OCR) technology will be implemented to convert text from images into machine-readable text.

The system will be integrated with a camera-equipped device or specialized wearable, capturing real-time video feed. Video frames will be processed through the object detection model to identify objects, while OCR extracts text information. To assist blind individuals, the system will provide audio feedback using text-to-speech technology, vocalizing detected objects and recognized text in real-time. User-friendly interfaces and feedback mechanisms will enhance usability and accessibility.

Regular testing and user feedback will play a crucial role in refining the system, ensuring its accuracy and reliability in diverse real-world scenarios. Overall, the goal is to create a seamless experience for visually impaired users, enabling effective navigation through their surroundings.

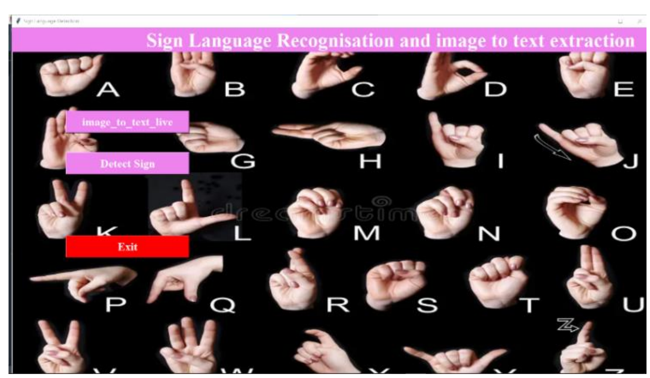
The proposed outcome of the Realtime Text and Object Detection for Blind Person project aims to empower visually impaired individuals by providing them with an innovative and accessible tool that leverages advanced computer vision technologies. Through real-time text recognition, the system will enable blind persons to comprehend printed materials, including books, menus, and signs, by converting text into speech or braille output. Additionally, the project incorporates object detection capabilities, allowing users to navigate their surroundings more independently by identifying and describing objects, obstacles, and people in their vicinity

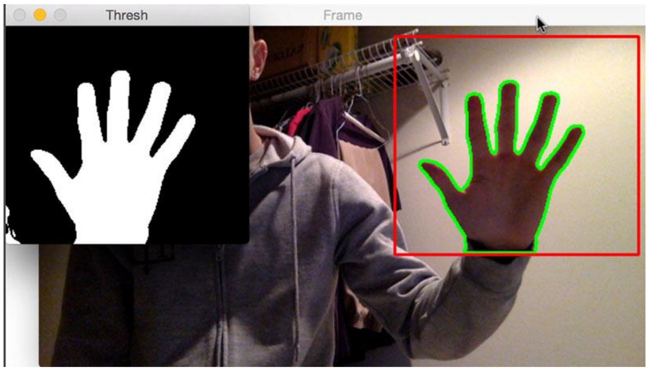
By seamlessly integrating these features into a user-friendly interface, the project intends to enhance the everyday lives of blind individuals.

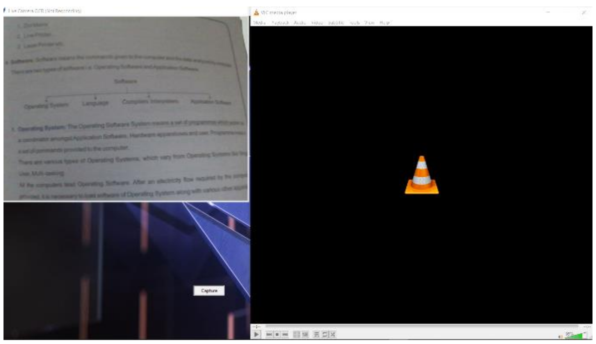
RESULT ANALYSIS

Beyond addressing the needs of the visually impaired, the project's motivation extends to mute individuals, recognizing the often-overlooked challenges they face in effective communication. Accuracy, user satisfaction, response time, error rate, conversation completion, user engagement, handling complexity, and compliance are all factors to consider when analyzing a Real Time Object Detection Using YOLO Technology for the blind and deaf individuals. It involves evaluating its accuracy in providing information, user experience, speed, and compliance with healthcare regulations.

User-friendly interfaces and feedback mechanisms have enhanced the overall usability of the system, ensuring a seamless and accessible experience for the target users. Regular testing and user feedback have played a crucial role in refining the system's accuracy and reliability, addressing challenges, and optimizing performance in various real-world scenarios.







CONCLUSION

In summary, the "Real-time Text and Object Detection" study shows promise in addressing the difficulties experienced by the deaf and blind. The project's primary goal is to empower visually impaired people by giving them access to real-time environmental information via sophisticated computer vision and natural language processing. With its smooth text-to-speech or Braille conversion and object-detection capabilities, the initiative not only makes daily life easier for blind people with vision impairments, but it also creates opportunities for social inclusion, work, and education.

Furthermore, the project’s commitment extends to mute individuals, introducing a hand gesture-based communication system. As the project aligns with ongoing advancements in machine learning and sensory technology, it not only represents a technological innovation but also a step towards fostering a more compassionate.

ACKNOWLEDGEMENT

We take this opportunity to thank our project guide Prof. Wagh D. B. and Head of the Department Prof. Dr. Sinkar Y. D. and Honorable Principal Prof. Dr. Mukane S.M. for their valuable guidance and for providing all the necessary facilities, which were indispensable in the completion of this report. We are also thankful to all the staff members of the Department of Computer Engineering of SVPM’s College of Engineering, Malegaon (Bk) for their valuable time, support, comments, suggestions, and persuasion. We would also like to thank the institute for providing the required facilities, Internet access and important books.

REFERENCES

[1] B. Khan, R. Naseem, F. Muhammad, G. Abbas, and S. Kim, “An empirical evaluation of machine learning techniques for chronic kidney disease prophecy,” IEEE Access, vol. 8, pp. 55012–55022, 2021.
[2] A. Ogunleye and Q.-G. Wang, “XGBoost model for chronic kidney disease diagnosis,” IEEE/ACM Trans. Comput. Biol. Bioinf.
[3] A. Khamparia, G. Saini, B. Pandey, S. Tiwari, D. Gupta, and A. Khanna, “KDSAE: Chronic kidney disease classification with multimedia data learning using deep stacked autoencoder network,” Multimedia Tools., vol. 79, nos. 47–48, pp. 35425–35440, Dec. 2022.
[4] D. Jain and V. Singh, “A novel hybrid approach for chronic disease classification,” Int. J. Healthcare Inf. Syst. Information., vol. 15, no. 1, pp. 1– 19, Jan. 2023,doi: 10.4018/IJHISI.2020010101.
[5] R. Parthiban, S. Usharani, D. Saravanan, D. Jayakumar, U. Palani, D. Stalin David, and D. Raghuraman, “Prognosis of chronic kidney disease (CKD) using hybrid filter wrapper embedded feature selection method,” Eur. J. Mol. Clin. Med., vol. 7, no. 9, pp. 2511–2530, 2021
[6] Xiaodong Yu, Ta Wen Kuan, Yuhan Zhang, Taijun Yan “YOLO v5 for SDSB Distant Tiny Object Detection” 2022 IEEE
[7] Mansi Mahendru, Sanjay Kumar Dubey,“Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3” 2023 IEEE Explore