IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

Smart Assistive System for Visually Impaired using PI

Prof. Mr. Vikas Gaikwaid, Mr. Pratik More, Ms. Sudhamani Bhagwat, Mr. Pratik Zende, Ms.Sakshi Bomble
Department of Artificial Intelligence and Data Science
Shree Ramchandra College of Engineering, Lonikand, Pune

Abstract: *Visually impaired individuals face significant challenges when navigating and engaging with their surroundings independently. Our solution, "Smart Assistive System for visually Impaired using pi" employs a Raspberry Pi and camera for real-time image capture, precise object classification (with over 90% accuracy), and auditory feedback. The project addresses a pressing need for greater inclusion and accessibility for the visually impaired, offering a cost-effective and innovative solution that converts visual information into non-visual cues. The "Caption-Speak" system holds the potential to significantly enhance the independence, mobility, and overall quality of life for visually impaired individuals..*

Keywords: Visual Impairment, Raspberry Pi, Image Processing, Object Classification, Auditory Feedback, Accessibility, Deep Learning, User Interface, Real-time Processing.

I. INTRODUCTION

Visual impairment impacts millions of individuals worldwide, limiting accessibility and independence in daily life. Navigating unfamiliar environments and recognizing objects pose significant challenges without sight. While some assistive technologies exist, they are often expensive and inaccessible to many. There is a need for an affordable, accessible smart assistive system to empower visually impaired individuals with greater mobility and independence.

This project aims to develop "Smart assistive system for visually impaired using pi", an image captioning system using a Raspberry Pi and camera module to provide contextual verbal descriptions of the surroundings to visually impaired users in real-time. The system captures images, analyzes the content using computer vision and deep learning algorithms, generates descriptive captions, and reads the text aloud to the user via headphones.

Current assistive technologies are limited in accessibility, real-time capability, and accuracy of environment understanding.

Our proposed system aims to overcome these limitations through optimised deep neural networks for precise context-aware captioning, efficient integration with the Raspberry Pi for real-time processing, and a usercentric design focused on accessibility and affordability.

The report provides a detailed overview of the problem, justification, proposed solution architecture, implementation plan, and expected outcomes. We survey relevant literature comparing existing systems and our innovations. The project aims to empower visually impaired individuals with greater awareness and understanding of their surroundings for increased safety, independence, and quality of life.

II. BACKGROUND

Visually impaired people face significant mobility and accessibility challenges in unfamiliar indoor and outdoor environments due to their inability to independently sense and understand the visual surroundings. The absence of affordable, accessible intelligent assistive technologies restricts their awareness, navigation, and participation in everyday activities.

There is a need for an effective computer vision-based assistive system that can automatically analyze visual environments in real-time and convey contextual understanding to visually impaired users through auditory feedback. This can greatly empower them with more independent mobility and improved awareness, safety, and quality of life.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

Existing assistive devices are limited in capabilities to understand and describe visual environments and objects. They also lack optimized integration with affordable hardware platforms to make them accessible to wider populations.

This project aims to develop an accessible, affordable image captioning system called "Caption-Speak" to address these gaps. It uses a Raspberry Pi and camera to capture images, analyze visual content using deep learning algorithms, generate contextual descriptions, and provide audio feedback to users - converting visual inputs to audio outputs.

III. LITERATURE SURVEY

The paper "Smart Assistive System for Visually Impaired People Obstruction Avoidance Through Object Detection and Classification" by Masud et al. (2021) focuses on developing a smart assistive aid for navigation and obstacle avoidance. The system uses a Raspberry Pi camera module for image capture and a TensorFlow object detection model toidentify obstacles with over 90% accuracy. Ultrasonic sensors provide proximity alerts. This assists visually impaired individuals in independently navigating while avoiding collisions.

The paper “Text to image synthesis for improved image captioning” by Hossain et al. (2021) explores generating relevant images from captions to improve image captioning systems. They employ a conditional GAN architecture called GAN space to synthesize images matching input text descriptions. This allows controlling image generation through latent space interpolation for high-quality results. The interpolated synthetic images help train improved captioning models.

The paper “An accurate generation of image captions for blind people using extended convolutional atom neural network” by Tiwary and Mahapatra (2022) focuses on image caption generation specifically for assisting blind users. The ECANN model combines CNN for feature extraction and LSTM for caption generation. It is optimised using an Adaptive Atom Search algorithm and achieves over 99% accuracy in generating human-readable captions.

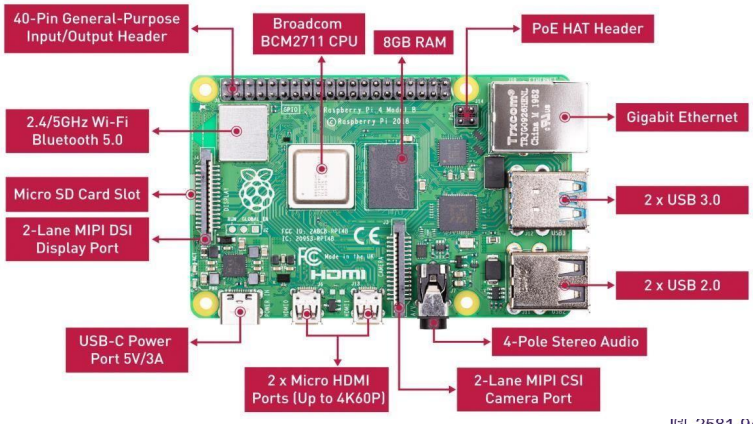
The survey paper “An Overview of Image Caption Generation Methods” by Wang et al. (2020) provides a comprehensive review of various techniques for image captioning using deep learning. It summarizes progress in encoders like CNN, decoders like LSTM, attention mechanisms, evaluation metrics and datasets. It also covers limitations of current methods and future improvements.

The paper “Obstacle and Fall Detection to Guide the Visually Impaired People with Real Time Monitoring” by Rahman et al. (2020) develops a system using sensors and smartphones to detect obstacles and notify users. It also monitors for falls and automatically alerts emergency contacts for safety. The system uses Bluetooth for connectivity and achieves 98% accuracy in obstacle detection.

IV. REQUIREMENT ANALYSIS

The requirements for the image captioning system were derived from a comprehensive consultation process involving vision impairment experts, potential end-users, academic guides, and extensive literature review. These requirements have been meticulously analyzed to ensure that the system fulfills the needs of visually impaired individuals effectively.

RASBERRY PI MODEL



IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

Processor: Broadcom BCM2711 quad-core Cortex-A72 (ARM v8) 64bit SoC @ 1.5GHz.
Memory: 8GB LPDDR4-3200 SDRAM
Connectivity: 2.4 GHz and 5.0 GHz IEEE 802.11b/g/n/ac wireless LAN
Bluetooth 5.0, BLE
Gigabit Ethernet

Ports:
2 × USB 3.0 ports
2 × USB 2.0 ports
2 × micro HDMI ports (up to 4Kp60 supported)
1 × 3.5mm audio jack (analog audio and composite video)
1 × CSI camera port
1 × DSI display port
MicroSD card slot for operating system and data storage
40-pin GPIO header (fully backward-compatible with previous Raspberry Pi boards)

Video & Audio:
H.265 (4Kp60 decode), H.264 (1080p60 decode, 1080p30 encode)
OpenGL ES 3.0 graphics
DSD512 audio support

Power:
5V DC via USB-C connector (minimum 3A)
5V DC via GPIO header (minimum 3A)
Power over Ethernet (PoE) enabled (requires separate PoE
HAT)

Dimensions: 85mm × 56mm × 17mm.

Operating temperature: 0°C to 50°C ambient.

AUDIO FEEDBACK:

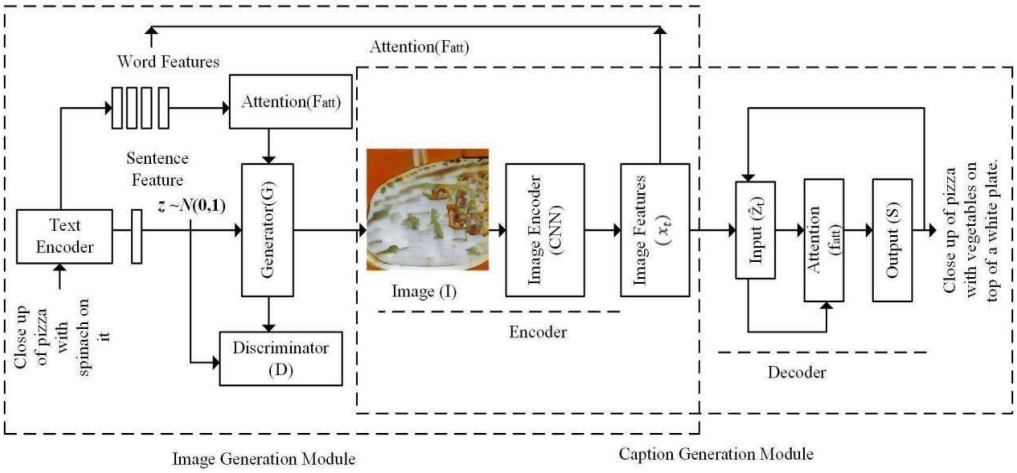
- **Caption-to-Audio:** The generated captions should be seamlessly converted into high-quality audio feedback. The system should allow customization of voice and speech rate to suit the preferences and needs of the visually impaired user.
- **Hands-Free Operation:** To minimize interaction barriers, hands-free operation modes, including voice commands, should be implemented, allowing users to interact with the system without physical input.

V. OBJECT DETECTION AND IMAGE CAPTIONING

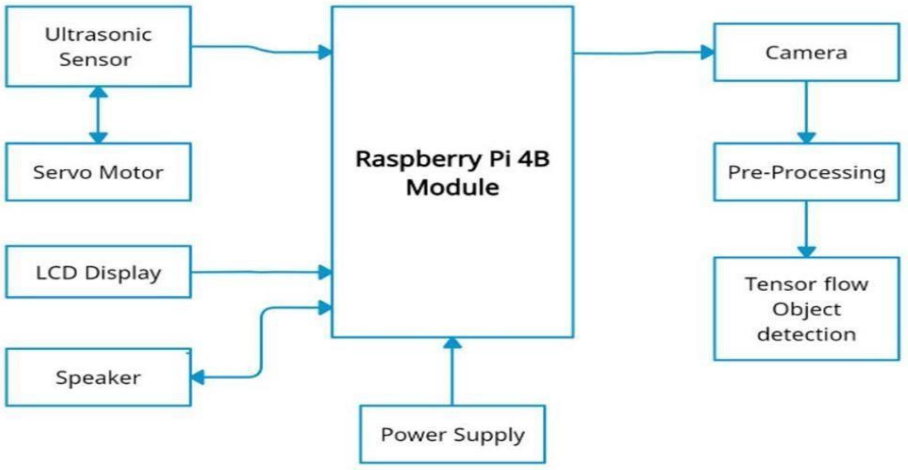
- **Object Identification:** The system should leverage an optimized deep neural network such as MobileNet to identify objects in the video frames. These objects should include, but are not limited to, people, cars, roads, and buildings, with minimal latency.
- **Caption Generation:** Specialized techniques, including Recurrent Neural Networks (RNNs) and attention mechanisms, should be incorporated into the system. These techniques will be used to generate multi-sentence captions that describe the scene, detected objects, their attributes, and any actions with contextual accuracy

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

VI. ARCHITECTURE



MATHEMATICAL MODEL:



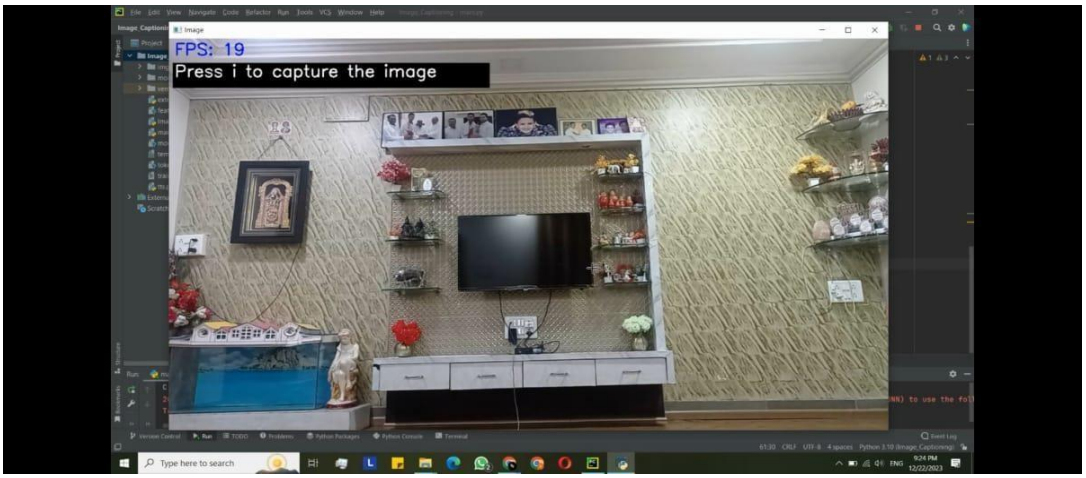
ACCURACY AND LATENCY:

- **System Accuracy:** The system must demonstrate a minimum accuracy of 95% for real-world deployment, ensuring the reliability of object identification and caption generation.
- **Latency:** The end-to-end latency for real-time interactivity should be kept under 500 milliseconds to provide users with prompt and responsive feedback

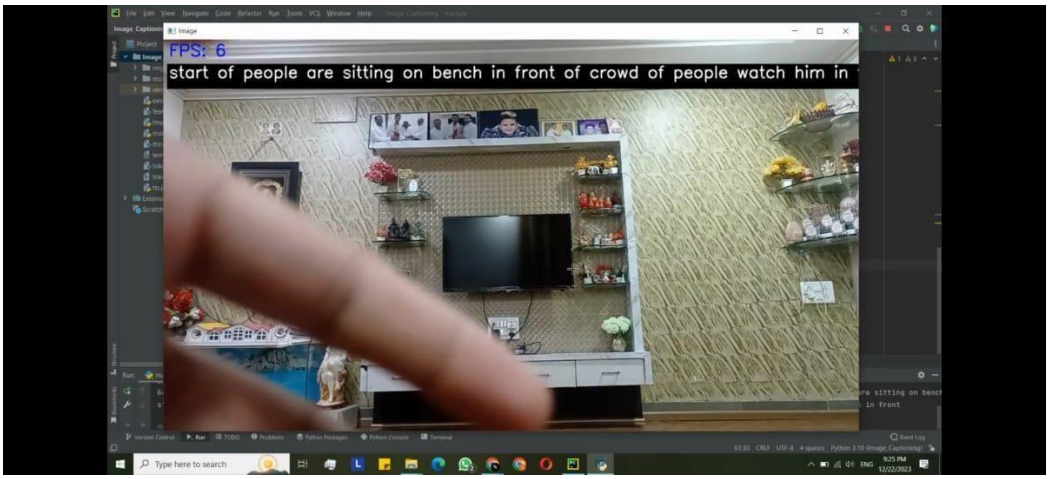
VIDEO FEED CAPTURE AND PREPROCESSING:

Video Capture: The system must capture real-time video feed from the environment. This is to be achieved using a Raspberry Pi camera module capable of recording video in 1080p resolution at 30 frames per second (fps).

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024



Preprocessing: Each video frame should undergo preprocessing techniques, including normalization and resizing, to enhance image quality, making it suitable for object detection and image captioning.



VII. CONSIDERATIONS FOR SYSTEM DESIGN

We use a Raspberry pi camera to take pictures of the scenario, whether it is an object or a person. This camera is mounted on the stick at some height so that it can capture the video as in real time. We have consulted an eye specialist and according to his recommendations we placed ultrasonic sensor at ground level as it helps detection in bumps, stairs, and hurdles in the pathway, while camera is positioned in the middle of human body. As we know that a normal lens camera captures the scene at 120 degrees from all sides. So, at this position, it will capture the images perfectly and also helps avoiding any contact with user's hands or body.

The servo motor is attached under the ultrasonic sensor so that it rotates in such a way that it tracks and left and right of the person in order to clear the path

The battery is also used as a power source in order to provide the required voltage for the operation of the system which can be recharged and have impressive amount of battery timing.

The Raspberry pi board behaves like a focal preparing unit for observing and controlling gadgets joined to it.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, May 2024

VIII. RESULT AND ANALYSIS

- **Performance**: The performance of real-time image caption generation on Raspberry Pi depends on the efficiency of the implemented algorithms and models. Optimization techniques such as model quantization, pruning, and hardware acceleration (if available) can improve performance.
- **Accuracy**: The accuracy of generated captions relies heavily on the quality of the pre-trained models used for image recognition and natural language processing. Fine-tuning or adapting models to specific domains or tasks can enhance caption accuracy.
- **Usability**: Real-time image caption generation on Raspberry Pi can have practical applications in areas such as assistive technologies, surveillance, and education. User-friendly interfaces and integration with other systems can enhance usability.
- **Scalability**: While Raspberry Pi serves well for prototyping and small-scale deployments, scaling up to handle a larger number of images or more complex tasks may require additional computational resources or distributed systems.

Overall, implementing real-time image caption generation on Raspberry Pi involves addressing technical challenges related to resource constraints, real-time processing, and model optimization while considering performance, accuracy, usability, and scalability requirements for practical applications.

IX. CONCLUSION

"Smart Assistive System for visually Impaired People Using Pi" represents a significant leap towards addressing the challenges faced by visually impaired individuals in navigating and understanding their surroundings. By harnessing the capabilities of Raspberry Pi, image processing, and deep learning, this smart assistive system not only enhances accessibility but also fosters greater independence and quality of life for its users. The commitment to ongoing improvement, usercentric design, and inclusivity highlights the project's potential for making the world a more accessible place.

REFERENCES

[1]. Github link:- https://github.com/PratikMore99/image-captioning-with-pi