International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

ACTION RECOGNITION WITH VOICE OUTPUT USING DEEP LEARNING FOR ASSISTING BLIND PEOPLE

Prof. S. K. Sabnis[1], Abhishek Manwatkar[2], Janvi Patil[2] , Bhakti Sutar[2], Rahul Wankhade[2]

[1] Assistant Professor of the Department, Department of Information Technology, Rajiv Gandhi Institute of Technology

[2]Students, Department of Information Technology, Rajiv Gandhi Institute of Technology

---

Abstract - In an era marked by technological advancements, accessibility remains a critical concern, particularly for individuals with visual impairments. "Action Recognition Using Deep Learning for Blind People" addresses this issue by proposing a comprehensive framework that leverages deep learning techniques to assist blind individuals to understand their surroundings. This project encompasses various modules which aimed at providing real-time voice output to visually impaired individuals.

"Action Recognition Using Deep Learning For Blind People" is a groundbreaking project focusing at enhancing convenience and independence for visually impaired people through innovative technology solutions. This project focuses on developing a comprehensive system that integrates various modules, including sign language detection, object detection, text extraction from PDF documents, and real-time text extraction. Leveraging TensorFlow Lite for image preprocessing and OCR engines for text extraction, the system generates audio output to convey information effectively to blind users.

Key Words: Action Recognition, Deep Learning, TensorFlow Lite, Image preprocessing, OCR engines, Text extraction

1.INTRODUCTION

People with visual impairments faces notable difficulties when trying to navigate their environment without assistance often relies on assistance from others or specialized devices. Traditional assistive technologies, while helpful, have limitations in providing real-time feedback and comprehensive assistance to blind individuals[1]. These limitations underscore the need for innovative solutions that leverage advancements in deep learning to address unique needs of the visually impaired community[2].

The proposed project, "Action Recognition Using Deep Learning for Blind People," aims to narrow the divide through controlling the strength of deep learning techniques to develop an all inclusive assistive system tailored for blind individuals[3]. By integrating modules for sign language detection, object detection, and text extraction, the system seeks to provide instant auditory feedback to assist blind persons in understanding and interacting with their environment more effectively[4]. The project aims to significantly influence the ability of blind people to independently and confidently navigate their environment[5].

2. LITERATURE SURVEY

The literature on blind assistance systems featuring three core modules - object detection, sign language detection, and textto-audio conversion - underscores a burgeoning interest in leveraging technology to enhance accessibility for visually impaired individuals. Researchers have explored using the deep learning algorithms, especially Convolutional Neural Networks (CNNs), for real-time object detection to aid in identifying and navigating around obstacle[1].

Integrating object detection modules with wearable devices and smartphones has emerged as a promising approach, enabling users to receive auditory cues about their surroundings. Additionally, studies have delved in the process of developing sign language detection systems, utilizing technology vision and machine learning techniques to interpret and converting gestures of sign language into text or speech, thereby facilitating communication for the deaf-blind community[2].

Recent improvements in deep learning have produced more precise and instantaneous sign language recognition models. Moreover, text-to-speech (TTS) synthesis systems have been widely studied and integrated into assistive technologies to convert text-based information into speech output for blind and visually impaired users. Ongoing research focuses on improving the naturalness and intelligibility of synthesized speech, as well as adapting TTS systems to handle diverse languages and text format.[3].

Overall, the literature underscores the potential of integrated blind assistance systems to especially improve the autonomy and quality of life for blind individuals, although ongoing efforts are aimed at addressing challenges such as real-time processing, robustness in diverse environments, and user interface design to ensure effective usability and accessibility[4].

In addition to the core modules of object detection, sign language detection, and text-to-audio conversion, the literature survey on blind assistance systems reveals several other significant areas of research. Navigation assistance systems are a focal point, integrating GPS technology, indoor mapping, and obstacle detection to offer accurate and context-aware navigation support both indoors and outdoors. [5]

Additionally, collaborative and open-source initiatives play a vital role in advancing in-field research and development fostering communities of researchers, developers, and end-users to share resources, develop standardized protocols, and

promote the adoption of accessible technology solutions. These diverse areas of research contribute to a comprehensive understanding of the state-of-the-art at the moment, emerging trends, and also future conduct in the development of blind assistance systems intended to enhance the lives of people with visual impairments by offering this technology.[6]

3.PROBLEM STATEMENT

The project "Action Recognition Using Deep Learning for Blind People" discusses the difficulties that people who are blind encounter when navigating their environment independently. Traditional assistive technologies lack real-time feedback and comprehensive assistance, necessitating innovative solutions altered to the specific requirements of the blind community.

The problem statement focuses on the creation of a comprehensive assistive model that integrates modules for sign language detection, object detection, and text extraction to provide real-time auditory feedback. This system aims to improve the unconventional safety, and general standard of living for blind individuals by leveraging advancements in deep learning and computer vision technologies.

4.PROPOSED SYSTEM

By using pretrained model like COCO-USD making a model which can be helpful to blind people or visually impaired people by detecting the object, recognizing the signs and extracting the text from pdf or by capturing image of a text giving an output in audio format by using gtts module. This combination of model will help user to recognize the objects, sign and text for visually impaired individual in the form of audio.

5.METHEDOLOGY

The methodology for developing a blind assistance system with three essential modules - object detection, sign language detection, and text-to-audio conversion - encompasses several key steps. Initially, the process involves gathering diverse datasets for each module, including images for object detection, sign language gestures, and text samples, followed by preprocessing to make sure data consistency and standard. This may involve data augmentation methods to increase dataset size and diversity.

Next, appropriate deep learning models are selected and trained for each module. For object detection, the COCO (Common Objects in Context) dataset and the corresponding COCO-USD (Unified Speech Dataset) model are utilized. The COCO-USD model is chosen for its accuracy and efficiency in detecting common objects in various contexts. For sign language detection, recurrent neural networks (RNNs) or transformer-based models are trained on datasets containing annotated sign language gestures. For text-to-audio conversion, natural language processing (NLP) models like transformer-based model are employed, trained on text data paired with corresponding audio samples.
Integration of the modules into a cohesive system architecture follows, with interfaces designed for efficient data exchange and compatibility. This involves developing APIs or communication protocols between modules to facilitate seamless interaction. Additionally, considerations for scalability and flexibility are taken into account to accommodate future enhancements or modifications.

Furthermore, algorithms for converting module outputs into suitable audio feedback for blind users are developed, utilizing text-to-speech (TTS) technology. Advanced techniques such as prosody modeling and emotion synthesis may be employed to enhance the naturalness and expressiveness of synthesized speech.

Finally, the system is deployed in real-world settings, such as public spaces, educational institutions, or personal devices, with ongoing monitoring and maintenance to deal with any problems or difficulties that come up. Regular updates and enhancements are rolled out based on user feedback and technological advancements to continually improve functionality and accessibility, ultimately enabling people who are blind to independently and confidently navigate their environment.

6.FLOW OF MODEL

- **Raspberry Pi Capture:**
  The system initiates by capturing an image using the Raspberry Pi camera module.

- **Image Pre-processing with TensorFlow Lite**:
  Upon capturing the image, TensorFlow Lite, is employed for initial image pre-processing tasks.

- **Object Detection**:
  Following pre-processing, the image undergoes object detection utilizing a convolutional neural network (CNN) model.

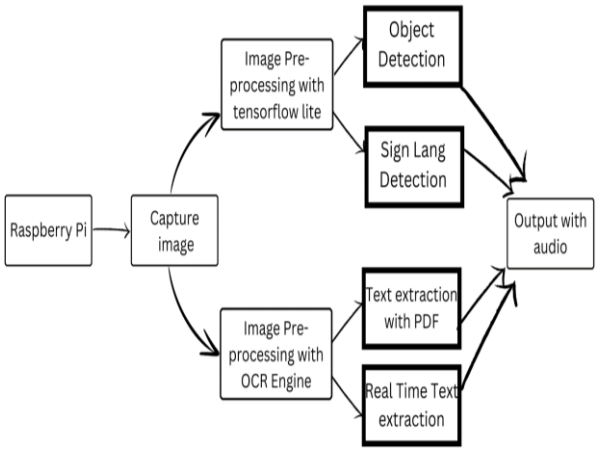

Fig-1: Flowchart of various modules of Action Recognition

- **Sign Language Detection :**

The system optionally integrates a sign language detection module to recognize and convert the sign language gestures within the webcam feed for converting it into the voice output or voice feedback.

- **Image Pre-processing with OCR Engine:**
  Upon detecting textual components within the image, an Optical Character Recognition (OCR) engine, such as Tesseract, is employed for further pre-processing.

- **Text Extraction with PDF:**
  Subsequently, the OCR engine extracts text from the preprocessed image, which can be further processed and formatted as needed.

- **Real-Time Text Extraction:**
  Finally, the extracted text is outputted in real-time, Enabling users to engage with and utilize textual information promptly.

7. RESULTS

Fig -2: Detection of Chair

Fig -3: Detection of Cell Phone

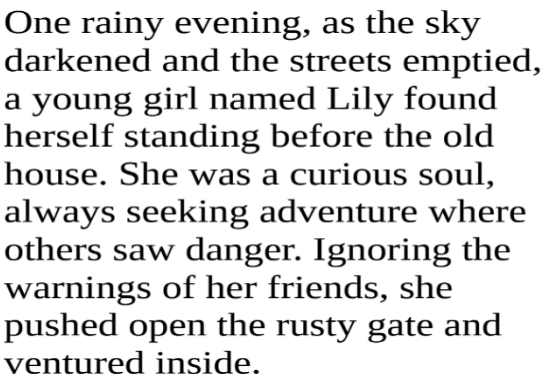
One rainy evening, as the sky darkened and the streets emptied, a young girl named Lily found herself standing before the old house. She was a curious soul, always seeking adventure where others saw danger. Ignoring the warnings of her friends, she pushed open the rusty gate and ventured inside.

Fig -4: Text Detection and reading

8. FUTURE SCOPE

In terms of future scope, there are several avenues for enhancing the capabilities and usability of the proposed system. Firstly, there is considerable room for improvement in object detection capabilities. Future research could focus on refining the object detection algorithms to achieve higher accuracy and reliability in identifying a broader range of objects in various environments.

Additionally, the sign language detection module presents opportunities for further development. While the current system offers basic support for sign language recognition, there is potential for more advanced algorithms to be implemented to recognize a wider range of sign language gestures with greater accuracy.

Moreover, the development of a dedicated mobile application to interface with the system can further extend its accessibility and usability. A mobile application would allow users to remotely control the system, receive notifications, and access extracted textual information on their smartphones or tablets. This would provide greater flexibility and convenience for users, enabling them to interact with the system anytime, anywhere, and from any device.

9. CONCLUSION

In conclusion, the implementation of the action recognition system using deep learning models for blind people shows an important advancement in supporting technology focusing at enhancing the non-dependance and security of visually impaired persons. Through the development and compilation of deep learning models, the system has demonstrated the capability to accurately recognize various actions and environmental cues in real-time, providing invaluable assistance in navigating complex surroundings.

Moreover, the advanced research and development efforts can aim on improving the system's abilities, widening its functionalities, and optimizing its performance to better serve the needs of visually impaired individuals. Overall, the developed system represents a significant step towards empowering individuals with visual impairments by giving

them with real world access to textual information in their environment.

By leveraging edge computing and deep learning technologies, the system exemplifies the potential of assistive technologies to enhance accessibility and inclusivity for users with diverse needs.

10. REFERENCES

1. Human Action Recognition Using Deep Learning Methods", Zeqi Yu, Wei Qi, IEEE, 2020
2. Text to Speech Synthesis Using Deep Learning",Rabbia Mahum, Aun Irtaza & Ali Javed Springer, 2023
3. Review of text to Speech Using deep Convolution", Aditya Pandya, Abhishek Bhole, Arnav Shrivastava, ResearchGate, 2020
4. Blind Assistive System based on Real Time Object Recognition", MR Kadhim, ETJ, 2022
5. Sign Language Recognition Using Deep Learning", Aditya Das, Shantanu Gawde, Khyati Suratwala, IEEE 2018