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: 7.301
Volume 3, Issue 5, May 2023

Introducing Next Generation Assistance: The Cutting-Edge Smart Cap for the Visually Impaired

Girish BG[1], Mohammed Bilal Zafer[2], Mohammed Thouqir[3], Javeriya Taj[4], Niveditha G S[5]
Department of Computer Science and Engineering
Sri Jagadguru Chandrashekaranatha Swamiji Institute of Technology, Chikkaballapura, Karnataka, India

Abstract: Natural and manmade disasters pose a myriad of challenges, which are more severe for individuals with disabilities. Ordinarily to perform daily activities, the disabled get support from assistive technological devices and services; these are commonly disrupted during and after disasters. A proposed solution to support those with visual impairment is a cost-effective wearable 'Smart Cap'. The aim of Smart Cap is to make the life of visually impaired people easier, comfortable and independent. There is a need of cap which is affordable, portable and user-friendly. In this paper, we design and implement a system using Raspberry Pi which helps the blind and also the visually impaired people to navigate freely by experiencing their surroundings. It provides features like face recognition, image captioning, text detection and recognition, and online newspaper reading using Internet of things and deep learning.

Keywords: Assistive technological devices and services, Resilience, Disaster, Smart Cap, Internet of Things, Deep Learning

I. INTRODUCTION

According to the estimations of the World Health Organization, worldwide, there are approximately 285 million visually impaired people, among them 246 million have low vision, and 39 million are blind. Vision is one amongst the very essential human senses, and it plays the important role in human perception about surrounding environment. The vision problems can occur due to various factors, such as aging, disease, injury, or genetics, which can have a significant impact on our quality of life.

The Internet of Things (IoT) plays a critical role in enabling smart systems to function effectively. IoT technology allows devices to connect and communicate with each other over the internet, creating a network of interconnected devices that can work together to gather and share data. In smart systems, IoT technology is used to connect various sensors, devices, and other components, allowing them to collect and share data in real-time. This data is then analysed using advanced algorithms and machine learning models to derive insights and make data-driven decision. One example of a smart system is a smart cap. As same through deep learning which is a subset of machine learning that uses artificial neural networks with many layers to analyse and learn from complex data. It is a powerful technology that has become increasingly popular in smart systems due to its ability to extract insights from large and complex datasets.One example of a smart system using both Internet of Things and deep learning is a smart cap.

The Dlib is a popular open-source software library written in C++ that provides a range of computer vision and machine learning tools. One of its most widely used modules is the face recognition module, which provides an implementation of deep metric learning for face recognition.The face recognition task based on dlib's face recognition module can be broken down into the following steps as face detection, alignment, feature extraction, and identification. The first step is to detect the face(s) in the input image or video frame which uses a histogram of oriented gradients (HOG) descriptor and a linear support vector machine (SVM) classifier to detect faces in images.

Once the faces are detected, the face alignment algorithm is used to align the face(s) so that they are in a standardized orientation and size. This is done to ensure that the facial features are in the same location and scale for all faces, making it easier to compare them. The next step is to extract a set of features from each face that will be used to identify it later. To identify a face, the feature vector extracted in the previous step is compared to a database of feature vectors for known faces. The image captioning module consists of attention-based CNN-LSTM

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: 7.301
Volume 3, Issue 5, May 2023

encoder-decoder architecture [15]. The Resnet-101 model [16] is used as an encoder, while LSTM [15] decoder with attention, coupled with beam search, is used to generate the best possible caption for the input image. Google's Vision API is used for text detection and recognition.

II. RELATED WORKS

Several related works have been developed in the field of assistive technology for the visually impaired. One of theresearchpresents the development of a "Smart Cap" system that uses deep learning and IoT technologies to assist visually impaired individuals in their daily activities. The system is designed to help users navigate their surroundings, recognize objects, and perform other tasks that may be challenging for people with visual impairments.One of the most prominent is the use of smart glasses equipped with cameras and audio feedback mechanisms to provide real-time audio descriptions of the user's surroundings. However, smart glasses can be bulky and heavy, making them uncomfortable for extended periods of use.Another related work is the use of handheld devices with camera-based object recognition systems to provide audio feedback to the user. However, these devices can be cumbersome to carry and use, and they can also be difficult to use in outdoor environments where lighting conditions may vary.Another research focuses on the use of a Smart Cane.The Smart Cane is designed to assist users in navigating their surroundings, avoiding obstacles, and identifying objects.The Smart Cane system consists of a standard white cane with sensors and cameras that capture information about the user's surroundings. The data is transmitted to a deep learning model, which processes the information and provides the user with audio feedback through a small speaker attached to the cane. Another important work presents a systematic review of wearable devices for visually impaired individuals. The authors conducted a comprehensive review of existing literature to identify and evaluate wearable technologies designed to assist visually impaired individuals in their daily activities.The review covers various types of wearable devices, including smart glasses, smartwatches, and other wearable sensors. The authors evaluated each technology based on its functionality, usability, and effectiveness in assisting visually impaired individuals.

III. PROPOSED SYSTEM

The system helps the blind to navigate independently using real time object detection and identification. The proposed system consists of a Raspberry Pi-3 processor which is loaded with a pre-trained Convolutional Neural Network model (CNN) developed using TensorFlow. The processor is connected to a Pi camera.

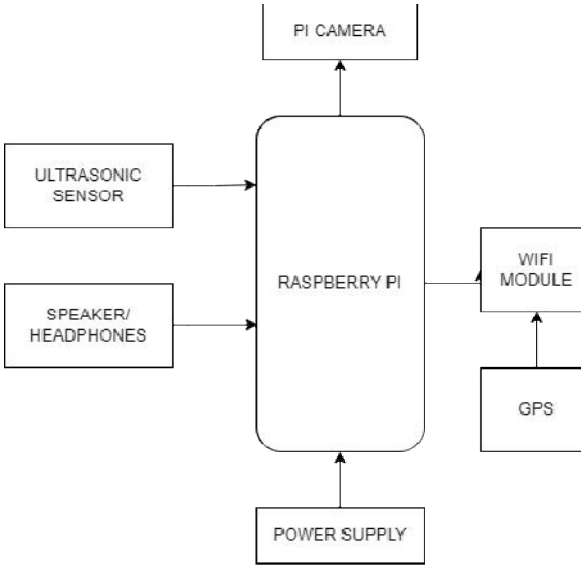

Figure 1: System Architecture

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: 7.301
Volume 3, Issue 5, May 2023

The processor is coded in python. The Pi camera captures the image in real time and will be provided to the Raspberry Pi-3 processor for processing it. The python code is used to detect and classify the objects. It will draw boundary boxes around the detected and will also show the category index of the object. The category index of the detected objects will be stored in a text file. The category index consists of the class name and class id of the detected object. After the process of Object detection, the ultrasonic sensors measure the distance of the object detected. This information is stored in a text file.

The contents of the text file is converted to voice using the Text to Speech Synthesizer (TTS) software e-Speak. This system is portable and the user can easily carry it. PI-CAM: It connects to a computer and internet and captures picture or motion video of user or another object and it allows face to face communication. UltrasonicSensor(HC-SR04): It emits sound waves at a frequency too high for humans to hear, they wait for sound to be reflected back and calculates the distance. The GPS is interfaced to find location of the user & sent through iot application to the user, their beloved ones and ultrasonic for detecting range of upcoming objects and alerting the user.

IV. SYSTEM DESCRIPTION

Face Recognition Module

The dlib library is a popular open-source software library that provides machine learning algorithms for various tasks, including face recognition. The dlib face recognition module uses a deep learning-based approach for face recognition, which is based on a 128-dimensional face embedding representation that is learned from a large dataset of face images. To use the dlib face recognition module in Smart Cap, the first step would be to install the dlib library and any necessary dependencies. Once installed, the module can be used to train a face recognition model using a set of known faces and their corresponding embeddings. This model can then be used to recognize faces in real-time from images captured by the Smart Cap's camera. The dlib face recognition module is known for its high accuracy and speed, making it a popular choice for face recognition applications in various industries. However, it may require some expertise in deep learning and computer vision to set up and optimize the module for specific use works.

Image captioning module

The image captioning model in Smart Cap is a deep learning-based model that generates textual descriptions of images captured by the Smart Cap's camera. This model consists main components: an image encoder, a text decoder, beam search and training.

Image encoder:

The image encoder in the image captioning module of Smart Cap is a convolutional neural network (CNN) that is responsible for extracting high-level visual features from an input image. The image encoder plays a critical role in the image captioning process, as it converts the input image into a fixed-length vector of feature representations that can be used by the text decoder to generate a textual description of the image.The image encoder typically consists of a series of convolutional and pooling layers, which are used to extract visual features of increasing complexity from the input image. The convolutional layers apply a set of learnable filters to the input image, producing a set of feature maps that capture different visual aspects of the image, such as edges, corners, and textures. The pooling layers then down sample the feature maps, reducing their spatial dimensions while preserving their important features.

Text Decoder:

The decoder in the image captioning module of Smart Cap is a recurrent neural network (RNN) that generates textual descriptions of images based on the visual features extracted by the image encoder. The decoder takes in the fixed-length vector of feature representations generated by the image encoder as input and generates a sequence of words that describe the image.The decoder typically consists of a series of recurrent layers, such as Long Short-Term Memory (LSTM) or Gated Recurrent Unit (GRU) layers, which are used to capture the temporal

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: 7.301
Volume 3, Issue 5, May 2023

dependencies between the words in the generated sequence. At each time step, the decoder takes in the previous word in the sequence as input, along with the fixed-length vector of feature representations generated by the image encoder, and generates a probability distribution over the next word in the sequence.

Beam Search:

Beam search is a decoding algorithm that can be used in the image captioning module of Smart Cap to generate more diverse and accurate captions. The algorithm works by maintaining a fixed number of candidate captions at each time step and selecting the captions with the highest probability scores. The number of candidate captions is referred to as the "beam size" and is typically set to a small value, such as 5 or 10.At each time step, the beam search algorithm generates new candidate captions by expanding the existing candidate captions. For each existing candidate caption, the algorithm generates a set of new candidate captions by appending each possible word in the vocabulary to the end of the caption. The probability scores of the new candidate captions are then calculated based on the probabilities assigned by the decoder model.

Training:

The goal of the training process is to optimize the model parameters to accurately generate textual descriptions of the input images. The quality of the trained model depends on the quality of the visual features extracted by the image encoder, the complexity and design of the text decoder, and the size and diversity of the training dataset.During the training process, the model is presented with an image and asked to generate a textual description of the image. The model generates a predicted caption, which is compared to the ground-truth caption using a loss function. The model parameters are then updated using backpropagation and gradient descent to minimize the loss function.The hyperparameters of the model, such as the learning rate and batch size, are tuned to optimize the performance of the model. The model is evaluated on a separate test set to measure the quality of the generated captions using metrics such as BLEU, ROUGE, and METEOR.The training process is computationally intensive and may require specialized hardware, such as GPUs or TPUs, to speed up the training process.

Text Recognition Module

The text recognition module in Smart Cap is designed to recognize text from images and convert it into machine-readable format. This module can be used by visually impaired individuals to read printed text, such as signs, labels, and documents.The text recognition module uses a combination of image processing techniques and machine learning algorithms to recognize text from images. The module typically consists of the following steps:

1. Image Pre-processing: The input image is pre-processed to enhance the quality of the image and remove noise. This may involve operations such as image binarization, skewing, and noise reduction.
2. Text Detection: The pre-processed image is analyzed to detect regions containing text. This may involve using techniques such as edge detection, morphology, and deep learning-based object detection.
3. Text Segmentation: The detected text regions are segmented into individual characters or words. This may involve using techniques such as connected component analysis, contour detection, and deep learning-based text recognition.
4. Text Recognition: The segmented characters or words are recognized using machine learning-based algorithms such as convolutional neural networks (CNNs) or recurrent neural networks (RNNs). These algorithms learn to recognize patterns and features in the input images and output the recognized text in machine-readable format.
5. Post-processing: The recognized text may undergo additional processing to improve its accuracy, such as language modeling, spell checking, and error correction.

The text recognition module can be trained using a dataset of labeled images and corresponding recognized text. The training process typically involves optimizing the parameters of the machine learning algorithms using a loss function that measures the difference between the recognized text and the ground truth.The performance of the text recognition module can be evaluated using metrics such as accuracy, precision, and recall. The module can be further optimized by fine-tuning the parameters, improving the preprocessing and post-processing steps, and

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: 7.301
Volume 3, Issue 5, May 2023

incorporating domain-specific knowledge.Overall, the text recognition module in Smart Cap is a powerful tool that enables visually impaired individuals to read printed text and access information that would otherwise be inaccessible to them.

V. RESULTS

The proposed system has the potential to be a useful tool for visually impaired individuals, and it may help to address some of the challenges that they face in their daily lives. If the system is implemented successfully, it could provide several benefits to visually impaired individuals, such as increased independence, improved navigation, and better situational awareness. The deep learning model, combined with IoT sensors and bone-conducting transducer, could help users to identify and avoid obstacles, recognize objects in their environment, and navigate unfamiliar spaces more confidently.

The success of the system would depend on several factors, including the accuracy and reliability of the system components, the ease of use and comfort of the hardware, and the degree to which the system meets the needs and preferences of visually impaired individuals. It is essential to involve visually impaired individuals in the development and testing of the system to ensure that it is useful, usable, and acceptable to them. In conclusion, while there may not be any actual results or analysis available yet for Smart Cap, the proposed system has the potential to be a valuable tool for visually impaired individuals. Further research and development are needed to evaluate the effectiveness of the proposed system, optimize its performance, and ensure that it meets the needs and preferences of visually impaired individuals.

VI. CONCLUSION

The system has a simple architecture that transforms the visual information captured using a camera to voice information using Raspberry Pi. Unlike other systems available in the market, the subject needs only to wear the cap and doesn't require any particular skills to operate it. Smart Cap is a technology solution that uses deep learning and IoT to assist visually impaired individuals in their daily lives. The cap is equipped with sensors and a camera that captures images of the surrounding environment, which are then processed using deep learning algorithms to identify objects, people, and text. The system then provides audio feedback to the user, enabling them to navigate their surroundings more easily. Overall, Smart Cap has the potential to significantly improve the quality of life for visually impaired individuals by providing them with greater independence and autonomy.

REFERENCES

[1]. Abidi, S., & Mahmood, T. (2020). Smart Cap: A Deep Learning and IoT Based Assistant for the Visually Impaired. IEEE Access, 8, 179022-179031

[2]. Singh, A., & Singh, A. (2021). Smart Cap for Visually Impaired People using Raspberry Pi. International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), 6(1), 135-140.

[3]. Niu, Q., Zhang, Y., Liu, Y., & Liu, J. (2020). Development of an Intelligent Glasses System for the Visually Impaired Based on IoT and Deep Learning. IEEE Internet of Things Journal, 7(12), 12408-12417.

[4]. Sharma, P., Arora, M., & Soni, M. K. (2021). Smart Assistive System for Visually Impaired People using IoT and Deep Learning. Journal of Ambient Intelligence and Humanized Computing, 12(8), 9077-9092.

[5]. Khan, A., & Elahi, N. (2021). IoT and Deep Learning-Based Smart Cane for the Visually Impaired. Sensors, 21(9), 3003.

[6]. Fernandes, S., & Rodrigues, J. J. (2020). Wearable Devices for Visually Impaired People: A Systematic Review. Journal of Ambient Intelligence and Humanized Computing, 11(6), 2459-2475.

[7]. Dara, R., & Gupta, G. (2020). Assistive Technology for the Visually Impaired: A Survey. International Journal of Computer Applications, 177(3), 11-17.

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: 7.301
Volume 3, Issue 5, May 2023

[8]. Balakrishnan, K., & Balakrishnan, R. (2020). An Overview of Assistive Technologies for Visually Impaired Persons. International Journal of Control and Automation, 13(1), 165-176.

[9]. Priya, M. G., &Marimuthu, R. (2021). Review on Assistive Technologies for Visually Impaired. Journal of Ambient Intelligence and Humanized Computing, 12(5), 5315-5333.

[10]. Zhang, Y., Li, J., Yang, J., & Wang, H. (2020). Assistive Technologies for Visually Impaired People: A Survey. IEEE Transactions on Human-Machine Systems, 50(5), 470-482.

[11]. Lee, S. G., & Choi, S. (2020). Artificial Intelligence Technologies for the Blind and Visually Impaired. Journal of Ambient Intelligence and Humanized Computing, 11(7), 2859-2874.

[12]. Khetre, A. (2021). Smart Assistive System for Visually Impaired People using Deep Learning and IoT. Journal of King Saud University-Computer and Information Sciences.