

Smart Blind Stick

Ms.G.Thilagavathi¹, Mr. R. Guru Ragavendaran², Mr.K.Jananitharan³ and Mr. R. Pratheep⁴

¹Assistant Professor, Department of Information Technology, Coimbatore Institute of Technology

E mail: thilagavathi.g@cit.edu.in

²⁻⁴Student, Department of Information Technology, Coimbatore Institute of Technology,

Email: gururagavendaranr@gmail.com, 71762107203@cit.edu.in, 71762107204@cit.edu.in

Abstract— A smart stick with an external camera and a Convolutional Neural Network (CNN) algorithm helps visually impaired people navigate their environment more autonomously. Users are able to more easily and confidently traverse their environment on their own because to the system's ability to recognise items in front of them and offer audio output to identify them. A Raspberry Pi computer, the Tensor Flow library, the Coco Dataset, and a Pi camera are used to perform the suggested fix. With a quick response time, minimal power consumption, and a lightweight, foldable form that is simple for users to handle and use, the smart stick is made with the user in mind. The proposed approach could, in general, considerably enhance the standard of living of individuals who are blind by boosting their mobility and autonomy. The smart stick's external camera takes pictures of the things the user is looking at, which are subsequently processed using the CNN algorithm. This algorithm can identify many items, including people, automobiles, and barriers, and it can voice-output the information to the user so they can recognise it. This makes it easier for users to move around and avoid potential dangers. The camera's photos can be processed effectively uses of the by using the Coco Dataset, which increases the algorithm's precision in detecting and classifying various objects.

Index Terms— Convolutional Neural Networks, Image Recognition, Coco Dataset, Image Classification.

I. INTRODUCTION

A. Object Detection

Object identification, is process of image processing and computer vision of image, involves identifying specific objects or features within digital images and videos. This technology is used in a wide range of applications, including image finding and motion surveillance. One of the most well-researched areas of object detecting is object and face detection. By identifying people and faces in the images and videos. Computer vision algorithms can be used to analyze the behaviour of human, track movements, and recognize individuals.

B. Image Captioning

Image captioning, also known as image tagging. A key method utilised in both computer vision and natural language processing is image tagging. It entails giving each image a descriptive tag that perfectly expresses the visual information. Image tagging is essential for a number of applications, such as content-based image search, object detection, and image retrieval. It entails creating a textual description of a picture, like a screenshot or a photograph, that accurately represents the image's context and content. This task is difficult because the model must comprehend the visual characteristics, semantic linkages, and linguistic patterns of the natural language used

to describe the objects, scenes, and actions in the image

C. Convolutional Neural Networks

A Deep Learning method used for picture categorization and detection tasks is called a convolutional neural network. In general, they provide weight to different features of the image, i.e., detect patterns, and can tell one image from another. In comparison to traditional Deep learning algorithms, CNNs require a great deal less pre-processing. CNNs can learn any type of image with enough practise and refinement. The configuration of neurons in the human brain closely resembles that of a convolutional neural network.

II. LITERATURE SURVEY

A. Real-Time Object Detection for Visually Challenged People – 2021 IEEE Access

In order to navigate, communicate, and obtain information in our daily lives, the sense of sight is essential. Unfortunately, there are millions of people who are visually impaired, which can make it challenging for them to carry out daily activities securely and independently. By warning blind people of items in their way, a suggested study seeks to solve this problem by transforming the visual world into an auditory one. They will be able to go on their own without assistance from outside sources thanks to this. The study will make use of a tool that recognises things in the field of view of the camera using machine learning and image processing.

B. A Smart Voice-Enabled Blind Stick with An Emergency Trigger - 2020 IEEE Access

Blind or partially sighted people occasionally struggle with transportation safety and frequently require assistance from others to complete even the most basic tasks. We propose a solution to this issue in the form of a system. Our technique uses ultrasonic sensors to detect obstacles and infrared sensors to detect raised surfaces like stairs. We also have a panic button for sending crisis signals to predetermined emergency contacts, as well as ISD1820 for verbal alerts when barriers are met.

C. Smart Electronic Stick for Visually Impaired using Android Application and Google's Cloud Vision -2019 IEEE Access 7 Journal [4]

Blind people encounter several difficulties in their daily lives, such as mobility and orientation issues. Simple sticks have always been the go-to tool for blind people to help them navigate their surroundings, but they might not always be enough to accomplish daily tasks on their own. However, a voice-activated Android software and an e-stick module have been developed as a potential remedy. In order to keep the traditional stick's lightweight and manageable construction, our method aims to provide a simple, affordable alternative.

D. Embedded Real-Time Object Detection and Measurement of its Size – 2018 IEEE Access

In a number of industrial industries, real-time object recognition and dimensioning have gained importance. The use of computer vision technologies is essential in overcoming these difficulties. In this study, a better method for quickly measuring and classifying objects from video streams is suggested. The method measures objects in real-time from video using Open CV libraries, a complex edge recognition algorithm, as well as erosion and dilation methods. Four processes are involved in the procedure: detecting and sorting contours, locating the object to be measured using a sophisticated edge detection approach, filling in the gaps between edges using morphological operators like erosion and dilation procedures, and measuring the object's dimensions. Real-time object detection and dimensioning, which are essential in many applications, can be solved with this technology in an efficient and effective way.

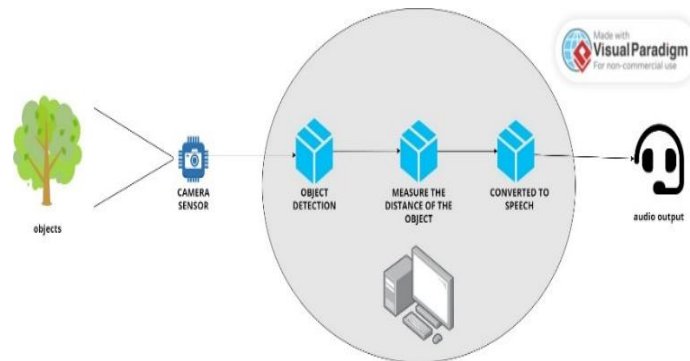
E. Moving object detection and tracking Using Convolutional Neural Networks – 2018 IEEE Access

Object tracking and object detection are two common steps in computer vision systems. While object tracking might be problematic when items are hidden by other objects, strong object detection can be difficult owing to fluctuations in situations.

This study suggests a method that uses the Tensor Flow object detection API to recognise moving objects in order to get around these challenges. An object tracking technique is used to determine the location of an object after it has been discovered. Even in difficult situations like low lighting or occlusion, a new CNN-based object tracking technique is used to achieve precise object detection. This method may produce self-generated visual sequences that can be applied in a variety of ways across multiple sectors.

III. SYSTEM ARCHITECTURE

The user data will be pre-processed before being entered into the algorithm to ensure accurate predictions. This could entail scaling the image, pixel value normalisation, and other data cleaning procedures.



ARCHITECTURE

Fig 1 System Architecture

A. Webcam or Any Camera Image Capturing Module:

The Webcam or Any Camera Image Capturing Module forms the fundamental interface between the system and the physical environment. This module harnesses the capabilities of the OpenCV library in Python to interact with the device's camera, be it a webcam or any other camera source. Users enjoy the flexibility to initiate image capture manually or set the module to operate automatically at predefined intervals, accommodating diverse user preferences and scenarios. The module's significance extends beyond mere image capture; it excels in image processing. Captured images undergo a sequence of transformations and analyses to extract crucial details about the object of interest, encompassing attributes like size, shape, and location within the frame. This extracted information serves as the bedrock for further analysis, where various approaches can be applied to gain deeper insights into the object's characteristics. The outcome of this image processing is versatile, with applications spanning from object localization and tracking to categorization and anomaly detection, empowering visually impaired individuals with real-time environmental awareness.

B. Image to Speech Converter

The Image to Speech Converter module stands as a pivotal component in bridging the gap between visual information and auditory output, catering to individuals with vision impairments. The process commences with meticulous image analysis, as the captured image undergoes scrutiny to identify and extract relevant text or objects, employing advanced image processing techniques, with OpenCV at its core. Once the necessary textual information is extracted, it is seamlessly relayed to the Pyttsx text-to-speech engine. Pyttsx, a robust Python utility, is responsible for the conversion of extracted text into audible speech. Users benefit from personalization options, allowing them to fine-tune their auditory experience by adjusting parameters such as speech pace, volume, and voice type through Pyttsx's configuration settings. This module translates visual data into a format comprehensible to those with visual impairments, granting them access to and interaction with visual content in a whole new dimension. Crucially, developers retain the flexibility to tailor the text-to-speech engine to meet specific user preferences, ensuring a uniquely customized experience for each individual.

C. Training Module

The Training Module resides at the core of the system's proficiency in recognizing and interpreting objects within images. Embracing Convolutional Neural Networks (CNNs) as the primary modeling technique, this component is chosen due to its prowess in handling image data and overcoming issues like overfitting and subpar test set accuracy. To train the CNN model effectively, pixel data from images is meticulously converted into numerical representations, facilitated by an image data generator. This data preprocessing step is pivotal, ensuring the model receives meaningful input. The initial model training employs a conventional CNN architecture with specific attributes, including three layers, 64 convolutions, and an activation layer, establishing a robust foundation. Further enhancements in accuracy are pursued through research and experimentation, which involves exploring pre-trained models such as Dense Net, VGG-16, Efficient Net, and InceptionV3. Importantly, these pre-trained models serve as a basis for fine-tuning the model's performance, without relying on pre-existing weights. Throughout the

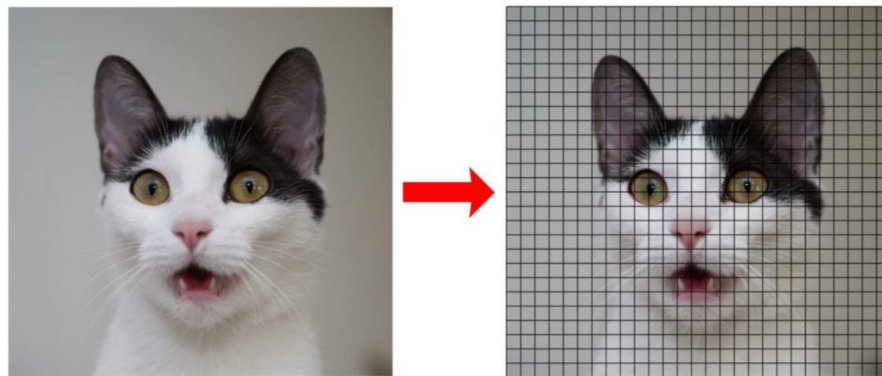
training process, the input data consists of pixel arrays, with labels corresponding to the associated filenames, thus enabling the model to learn the associations between visual information and specific objects or categories. The utilization of CNNs and pre-trained models substantially elevates the precision and effectiveness of object recognition within the system, empowering visually impaired individuals to navigate their environment with increased independence and confidence.

IV. WORKING

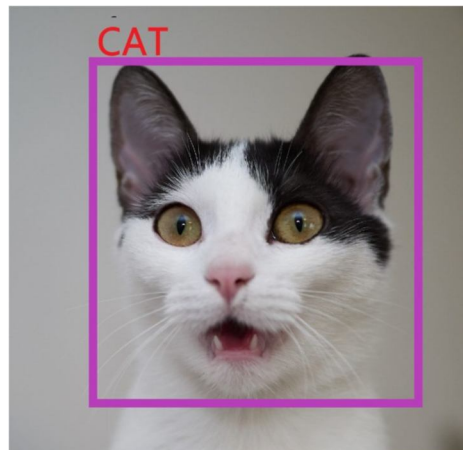
A. Working Of Object Detection

A computer vision technology called object detection makes it possible to recognise items in pictures and videos. To create object detection algorithms, there are two methods: pre-training and learning from scratch. Pre-trained models are frequently employed, and their weights are occasionally changed to satisfy certain specifications. Three phases are commonly involved in the object detecting process.

- The image is initially transformed into an array form. Each rectangle in the grid is checked to see if it contains a valid object before being extracted as a featured row or column.



a)

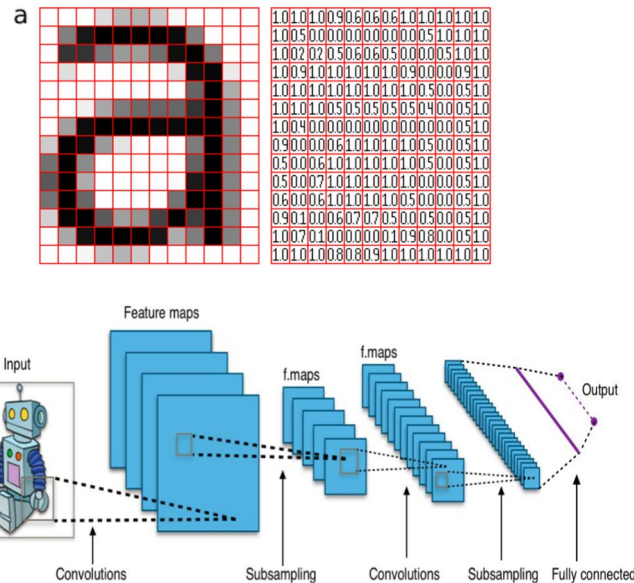


b)

B. CNN (Convolutional Neural Networks)

In data processing, image analysis and recognition, image classification, and edge detection in visual data, convolutional neural networks are the neural network that are frequently employed. In a CNN, each layer of neurons applies a non-linear operation—a linear transformation—to the results of the layer before it. Convolutional and pooling layers are two examples of the several types of layers included in the CNN architecture. CNNs can efficiently detect numerous relationships in a picture by utilising filters. To ensure a deep learning algorithm works effectively, the complexity of the CNN architecture needs to be adequately addressed. In CNNs, convolution is mostly utilised to extract fine details from the image, such as edges. Multiple layers may be stacked on top of one

another in CNNs. The first layer often records coarse data like edges, colours, and gradient direction. The architecture aids in the discovery of increasingly complex and high-level features as we delve further into the network, which is helpful for training our deep learning model.



B. Convolutional Layers

The filters that comprise a convolutional layer in a CNN are crucial, and their parameters must be learned. In comparison to the input volume, these filters are lighter and smaller in height. Each filter is convolved with the input volume in order to compute the activation map. As a result, a neural activity map is produced. The activation maps of all the filters are stacked to generate the convolutional layer's output volume. After that, the network's next layer receives this output volume for additional processing. The stacking of the activation maps enables the detection of intricate patterns in the image by capturing various input volume features.

C. Pooling Layers

A pooling layer is frequently inserted between two subsequent convolutional layers in CNNs to simplify and reduce the amount of parameters. With the help of the pooling layer, representation of previous convolutional layer is scaled down. A representation can be downsampled by the pooling layer using a variety of functions, including average or maximum pooling. However, because it works better in real-world situations, the max pooling function is more frequently utilised. Max pooling aids in preserving the most important aspects of the input volume while rejecting the less significant ones by taking the maximum value in a pooling region. As a result, the input volume can be represented in a way that is more effective and efficient for the network's next layer to handle.

D. Fully Connected Layers

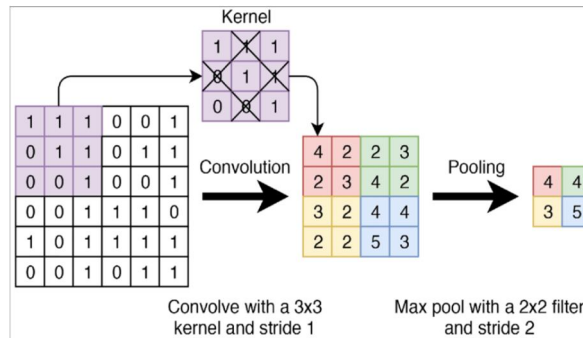
All of a layer's inputs are linked to the activation unit of the layer above them in a fully connected neural network. Dense layers refer to the final few levels of a completely linked network, where each layer of neurons is coupled to every other layer of neurons. In order to produce the final output, the network might then merge the data collected from earlier levels. Fully connected layers demand more calculation time than convolutional layers since there are many more parameters involved in the connections between the layers. Consequently, training a fully connected neural network may take longer and cost more money than training a CNN.

E. Receptive Field

The receptive field in a CNN refers to the particular region of the input space that affects a given network unit. Both an input to the network and an output from other network nodes are possible for this location. It is possible to determine the receptive field in relation to both the input we are considering and the receiving unit. In other words, we can identify the network feature maps that correlate to the size and location of the receptive field in the

input space. The capacity of the network to detect and represent complex patterns and structures in the input data is strongly influenced by the size of the receptive field.

F. Weights

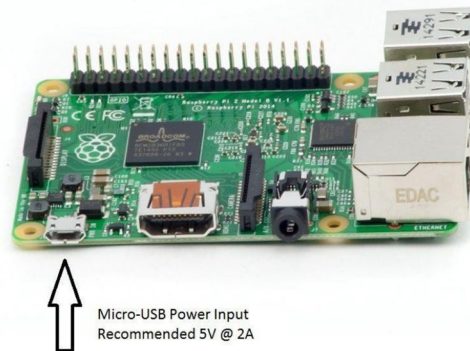


The neural network's weights are changed throughout the training phase to enhance the network's performance on a particular task. The importance of the input signal transmitted across a connection is determined by its weight. An input signal's significance is diminished when it is transferred through a link because it is multiplied by the relevant weight value. The next layer of the network then combines the combined input signals that have been altered, and so on until the final output is produced.

V. HARDWARE AND DETAILS

A. Raspberry PI

A well-liked and adaptable single-board computer around the size of a credit card is called the Raspberry Pi. Despite its diminutive size, it is a strong tool with a wide range of applications. It is simple to use for people of all technological levels because it can be connected to a computer display or TV and is operated with a standard keyboard and mouse. Low power consumption is one of the Raspberry Pi's main benefits. It uses a little amount of power, typically 5V @ 2A, making it an energy-efficient choice for a variety of uses. The Raspberry Pi is also intended to be inexpensive, with a variety of versions offered at varying price ranges to suit varied budgets.



(a)

B. Setup The Raspberry PI

- Step1: Insert an SD card of desire size to the laptop
- Step2: Then go and visit the official Raspberry Pi website and get the imager
- Step3: Then using this Raspberry Pi imager software select the OS which is to be installed on the SD card
- Step4: Then select the SD card in which the OS is going to be installed
- Step5: Then finally click on the write to download and install the OS to the SDcard.
- Step6: Place SD card in Raspberry Pi's SD card port, then connect the charger to turn it on.
- Step7: The Raspberry Pi will be booting into the OS.

C. Camera

The system's built-in camera module can record high-definition still photos and videos in addition to still images. A fixed-focus lens and an image sensor with a native resolution of five megapixels are included in the camera module. This makes it possible to record clear, detailed images, enabling exact and accurate item analysis and identification in the visual environment. As a whole, the camera module is an essential part of the system since it offers accurate and high-quality visual data for modelling and analysis.



(b)

D. Speaker

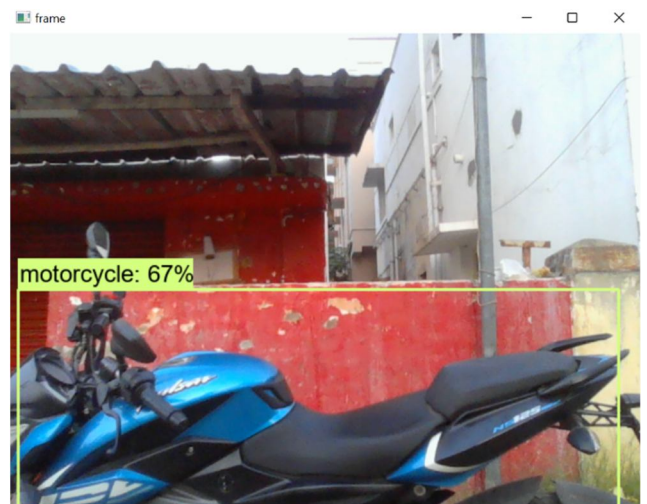
Speaker is an output device that gives audio as an output, mostly used with computers and phones to listen to the mp3 and other audio files.

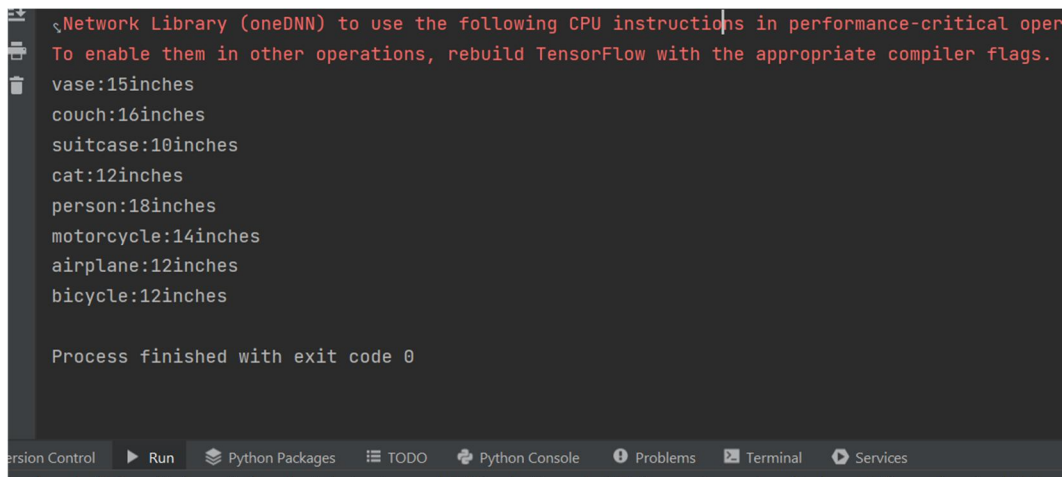
E. Raspberry PI Speaker Connection

A Raspberry pi can be connected to the speaker in different ways like through the inbuilt Bluetooth , through the inbuilt audio jack or through the displays which has inbuilt speaker on them.



VI. RESULT





```
<Network Library (oneDNN) to use the following CPU instructions in performance-critical oper
To enable them in other operations, rebuild TensorFlow with the appropriate compiler flags.
vase:15inches
couch:16inches
suitcase:10inches
cat:12inches
person:18inches
motorcycle:14inches
airplane:12inches
bicycle:12inches

Process finished with exit code 0
```

In the outcome of this project, an audio output system plays a pivotal role by providing detailed auditory information regarding objects positioned directly in front of the smart blind stick. This audio feedback is multifaceted in its utility; firstly, it serves as a real-time object detection mechanism, alerting users to the presence of obstacles, pedestrians, or other objects obstructing their path. Additionally, it goes a step further by offering distance measurements, allowing users to ascertain not only what's in front of them but how far away it is. This feature is immensely empowering for individuals with visual impairments as it significantly bolsters their navigational capabilities. By using this audio feedback, they can make informed decisions about how to proceed, whether it involves adjusting their route, slowing down, or stopping altogether. The system becomes their virtual eyes, helping them explore their environment with a newfound sense of confidence and autonomy. Furthermore, this functionality enables blind individuals to actively search for specific objects or landmarks, such as doors, staircases, or bus stops, enhancing their independence in daily life. The ability to precisely locate and identify objects in real-time provides a level of situational awareness that is transformative for the visually impaired. This project's output, therefore, represents a ground-breaking advancement in accessibility technology, offering a lifeline of information and independence to those who rely on it.

VII. CONCLUSION

Providing assistance to people with visual impairments is a significant difficulty in contemporary society, but recent technological developments are offering fresh approaches. A transportable model created to aid blind people in navigating and comprehending their surroundings is one such remedy. This device has the ability to recognise items in its environment, spell out their names, and provide details about their relationships. The project makes use of cutting-edge technology including computer vision, natural language processing, and natural language generation to complete these goals. In particular when compared to other boards on the market, using the Raspberry Pi as the main board for this project has shown to be a cost-effective and efficient alternative. This device's voice-activated technology, which enables users to activate the system with a straightforward keyword command, is a crucial feature. The device enters a standby state until the next activation after completing the intended task. People who are blind or visually handicapped will find this to be helpful and easy to use. Overall, this ground-breaking tool offers a straightforward and practical solution for people who are blind or visually impaired to better comprehend their environment and manage their surroundings more easily and independently.

REFERENCES

- [1] Prof. Supriya Mangale, Shraddha Mane, "Moving object detection and tracking Using Convolutional Neural Networks", 2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS 2018) Part Number: CFP18K74-ART; ISBN: 978-1-5386-2842-3
- [2] Prof. Radha Shankarmani, Sunit Vaidya, Niti Shah, Naisha Shah, "Real-Time Object Detection for Visually Challenged People", International Conference on Intelligent Computing and Control Systems (ICICCS 2020) Part Number: CFP20K74-ART; ISBN: 978-1-7281-4876-2
- [3] Nashwan Adnan OTHMAN, Mehmet Umut SALUR, Mehmet KARAKOSE, Ilhan AYDIN, "An Embedded Real-Time Object Detection and Measurement of its Size", 2018, IEEE TRANSACTIONS, 2012 978-1-4673-0311-8/12/\$31.00 ©2018 IEEE.

- [4] Devshree Bharatia, Priyanka Ambawane, Piyush Rane, "Smart Electronic Stick for Visually Impaired us in Android Application and Google's Cloud Vision", 2019, 2019 Global Conference for Advancement in Technology (GCAT Bangalore, India. Oct 18-20, 2019.
- [5] Priyanka Abhang, Shambhavi Rege, Shrishti Kaushik, Shriya Akella, Manish Parmar, "A Smart Voice-Enabled Blind Stick with An Emergency Trigger", 2020, ICICES 2014. 10.1109/ICICES.2014.7033988.
- [6] Mukesh Prasad Agrawal, Atma Ram Gupta, "Smart Stick for the Blind and Visually Impaired People", 2018, 2nd International Conference on Inventive Communication and Computational Technologies (ICICCT 2018) IEEE Xplore Compliant - Part Number: CFP18BAC-ART; ISBN:978-1-5386-1974-2
- [7] Nishant Banat James , Ashish Harsola, "Navigation Aiding stick for the Visually Impaired", 2015, 978-1-4673-7910-6/15/\$31.00 © 2015 IEEE.
- [8] N. Loganathan, K. Lakshmi, Chandrasekaran, S. R. C. Bisakaravathi, "Smart Stick for Blind People", 2020, IEEE TRANSACTIONS, 2012 978-1-4673-0311-8/12/\$31.00 © 2020 IEEE.
- [9] Vanitha Kunta, Charitha Tuniki, U. Sairam, "Multi-Functional Blind Stick for Visually Impaired People", 2020, IEEE TRANSACTIONS, 2012 978-1-4673-0311-8/12/\$31.00 © 2020 IEEE.
- [10] Himanshu Sharma, Meenakshi Tripathi, Amit Kumar, Manoj Singh Gaur, "Embedded Assistive Stick for Visually Impaired Persons", 2018, IEEE TRANSACTIONS, 2012 978-1-4673-0311-8/12/\$31.00 © 2018 IEEE.
- [11] Apoorva Raghunandan, Mohana, Pakala Raghav, H. V. Ravish Aradhya, "Object Detection Algorithms for Video Surveillance Applications ", 2018, IEEE TRANSACTIONS, 2012 978-1-4673-0311-8/12/\$31.00 © 2018 IEEE.
- [12] Shaji Thorn Blue, M. Brindha, "Edge detection based boundary box construction algorithm for improving the precision of object detection in YOLOv3", 2019, 10th ICCCNT – 2019 July 6-8, 2019, IIT
- [13] Lushen Wu, Zhen Liu, Yuwen Li, "Moving Objects Detection Based on Embedded Video Surveillance", 2012, International Conference on Systems and Informatics (ICSAI 2012).
- [14] M. S. Chelva Dr. S. V. Halse Dr. B. K. Ratha, "Object Tracking In Real Time Embedded System Using Image Processing", 2016, International conference on Signal Processing, Communication, Power and Embedded System (SCOPES)-2016.
- [15] Souhail Guennouni, Ali Ahaitouf, Anass Mansouri, "Multiple Object Detection using OpenCV on an Embedded Platform", 2014, 978-1-4799-5979-2/14/\$31.00 © 2014 IEEE