



Smart Cane for Visually Impaired

Mr. J. Mohamed Farhan¹, Mr. S.V. Sakthivel¹, Mr. R. Vishal¹, Mrs. S. Rajasree²

¹Student, Coimbatore Institute of Technology

²Assistant Professor, Coimbatore Institute of Technology

ABSTRACT

We study on the hypothesizes of a smart cane that alerts visually-impaired people over obstacles. The aim of the paper is to address the development of a cane that could communicate with the users through voice alert and haptic vibrations. We research emphasizes on Internet of Things (IoT) based affordable platforms to take proper and timely measures for disabled people. The paper introduces an IoT based real time analysis and alerting system for the disabled people.

Keywords: Visual Geometry Group(VGG16), Real time Image Classification, Machine Learning, HC-SR04 Ultrasonic sensor, MPU6050 accelerometer, Raspberry pi.

1 INTRODUCTION

1.1 Rationale

The increasing number of people with disabilities in India attracts the concern of researchers to invent various technologies, hoping that these technologies can assist the disabled people in carrying out their tasks in everyday life like normal people. India is home to around 12 million blind people, which accounts for one-third of the world's total blind population. A cane is a device used by many people who are blind or visually impaired. A cane primarily allows its user to inspect their surroundings for obstacles or orientation marks, but is also helpful for onlookers in identifying the user as blind or visually impaired and taking appropriate care.

1.2 Internet of things

The Internet of Things (IoT) is a system of interrelated computing devices, mechanical and digital machines, objects, animals or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction. In the Internet of Things, all the things can be put into three categories: 1. Sensors that collect information and then send it. 2. Computers that receive information and then act on it. 3. Things that do both. Sensors can quantify temperature, movement, dampness, air quality, light, and nearly whatever else you can consider. Sensors, when combined with a web association, permit us to gather data from the climate which, thusly, assists settle on with bettering choices. The Internet of Things, or IoT, is about extending the power of internet connectivity beyond computers to a whole range of other things, processes, and environments. Those connected, smarter, things are used to gather information, send information, or both.

1.3 YOLO V5

YOLO V5 is a family of compound-scaled object detection models trained on the COCO dataset, and includes simple functionality for Test Time Augmentation (TTA), Model ensembling, hyperparameter evolution, and export to ONNX, CoreML. YOLO v5 is different from all other prior releases, as this is a PyTorch implementation rather than a fork from original Darknet. Same as YOLO v4, the YOLO v5 has a CSP backbone and PA-NET neck. The major improvement includes mosaic data augmentation and auto learning bounding box anchors.

1.4 Visual Geometry Group (VGG16)

VGG 16 is a convolutional neural network that is 22 layers deep. The network trained on ImageNet classifies images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. VGG 16 is a leading architecture for any object-detection task and may have huge applications in the computer vision sector of artificial intelligence and IoT based sectors. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes.

2. LITERATURE SURVEY

2.1 IoTEnabled Automated Object recognition for the Visually Impaired.- Rahman, Md Atikur, and Muhammad Sheikh Sadi- (2021)Elsevier

This paper uses ultrasonic sensor, accelerometer and four laser sensors are used in the system to detect the objects in the direction of the front, left, right and ground. The proposed system uses Single Shot Detector (SSD) model with MobileNet and Tensorflow-lite to recognize objects along with the currency note in the real-time scenario in both indoor and outdoor environments. The overall accuracy of the proposed system in object detection and recognition is 99.31%.

2.2 Smart cane for Visually Impaired with Obstacle, Water Detection and GPS.-Nazri, Nur Maisarah Ahmad, et al.- (2020)ResearchGate

This paper ultrasonic sensor, IR sensor, water sensor to detect the object in front of the user and detects the water and informs the user with a speech warning message. The proposed system uses Node MCU chip and ESP8266 Wi-Fi board is a system-on-chip (SOC) with an integrated TCP/IP protocol stack that can give some secondary microcontroller access to your Wi-Fi network. The accuracy of water detection and GPS tracking is 93.71%.

2.3 Smart Stick For the Blind Using Arduino.- Romadhon, A. S., and A. K. Husein, et al.-(2020)IOP SCIENCE

This paper uses an ultrasonic transducer that is based on a transmitter and receiver and mainly used to determine the distance from the target object with the wave length about 20kHz- 20 MHz. GPS NEO-6M is one in a million GPS Module that one of the GPS UBLOX Series. It can track up to 22 satellites on 50 channels and achieves the industry's highest level of sensitivity i.e. -161 dB tracking, SIM800l is a GSM module that can be applied with Arduino or AVR as a sender or receiver of data which will be processed by an Android device or microcontroller.

3.1 Architecture

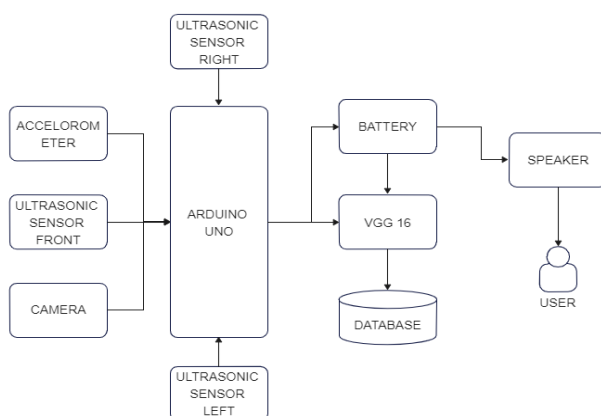


Fig 3.1 system architecture

The IoT devices are integrated with the software VGG16 connected with a battery and alerts the user with a voice playback on the smart cane.

Ultrasonic sensor Ultrasonic sensors sends an ultrasonic pulse out at 40kHz which travels through the air and if there is an obstacle or object, it will bounce back to the sensor. By calculating the travel time and the speed of sound, the distance can be calculated. The transducer of the sensor acts as a microphone to receive and send the ultrasonic sound. The ultrasonic sensors use a single transducer to send a pulse and to receive the echo. The sensor determines the distance to a target by measuring time lapses between the sending and receiving of the ultrasonic pulse.

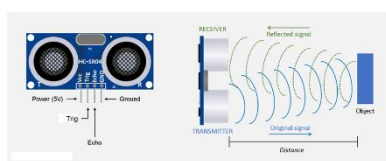


Fig 3.2 Ultrasonic pulse hit the object and bounce back

Accelerometer Freefall is a type of dynamic acceleration. Using the MCU6050 accelerometer to monitor the accelerations of a system, different freefall motions can be detected, called freefall signatures. We have developed an algorithm to monitor the components of freefall signatures to determine when a freefall condition is met.

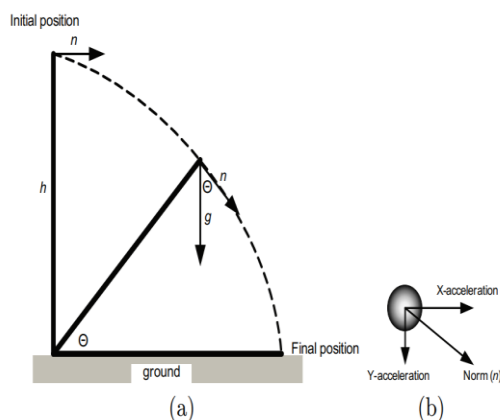


fig 3.3 The (a) side view and (b) top view

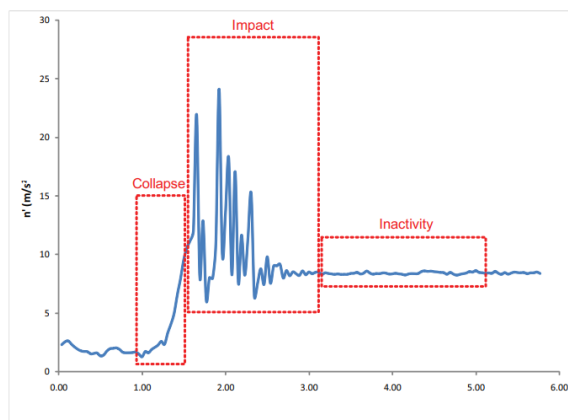


fig 3.4 normal acceleration of the typical falling motion of a cane'n'based on the model

Image Capturing module using Webcam / Any Camera module. This module captures the image of the object(trash) using python's OpenCV library. The function it uses is video capture. This image capturing process can either be automated (Automatic capturing of the object every 5 seconds/any desired time) or manually done by pressing 'Q' button in the keyboard.

Raspberry pi 64 bit. Raspberry pi is a capable little device that enables people of all ages to explore computing, and to learn how to program in languages like Scratch and Python. Raspberry Pi is good for developing software applications using Python, while Arduino is good for interfacing Sensors and controlling LEDs and Motors. It is integrated with the camera module to make it portable with the cane.



Fig 3.5 Raspberry pi

Training Module. Convolutional Neural Networks are used to train and test the model. The main reason behind this will be the poor test set accuracy and overfitting issue with the traditional Machine Learning Algorithms. Trains the network with the pixel data that is converted into numerical values using Image Data generator. Ordinary Convolutional neural network with 3 layers and 64 convolutions each and with tanh activation layer is used for initial training. Further experiments were conducted on pre-trained models without utilizing the existing weights. Pre-trained models like EfficientNet, DenseNet, VGG-16 and InceptionV3 are utilized for building the desired model. Input as pixel arrays and labels as their respective filenames are used for training.

Labeling LabelImg is a graphical image annotation tool. It is written in Python and uses Qt for its graphical interface. Annotations are saved as XML files in PASCAL VOC format, the format used by ImageNet. Besides, it also supports YOLO formats. The set of images as been labelled based on the classes specified and stored as a XML file.

4 SYSTEM DESIGN AND IMPLEMENTATION

4.1 IoT Integration

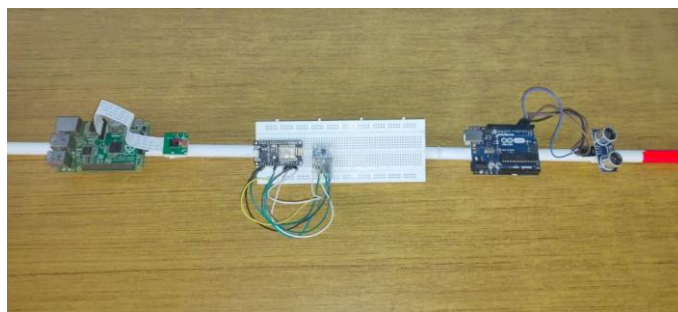


Fig 4.1 Smart cane Integration

Shown above is the fully connected and working model of smart cane interconnected with accelerometer, ultrasonic and raspberry pi camera module.

4.2 Dataset

The dataset comes from a real time images of car, bike, dog, bus etc., for research purposes. All the images are captured in different background, different lighting conditions to keep the diversity in the dataset alive. In addition to this dataset, we included some more images taken from our own camera. Labels of the images was divided into 7 classes like car, bike, cycle, bus, snakes, cow and dog. These classes are further subdivided into harmful and unharmful objects.

Class	Number of images
Dog	121
Cow	119
Snakes	141
Bike	120
Car	120
Cycle	112
Bus	120

Total images: 853 images



Fig 4.2 dataset images

These are the sample images of the dataset .

Train dataset. The training dataset consists of a data from all the labelled images. Each class have an average of 117 images.

Test dataset. The testing data consists of randomly selected images for each class.

Raspberry pi. The respective python file accessing camera is written and manipulated using raspberry pi, it is used for real time data storage and retrieval. It produces a righteous image/video quality and the lapse can also be mentioned in the code.

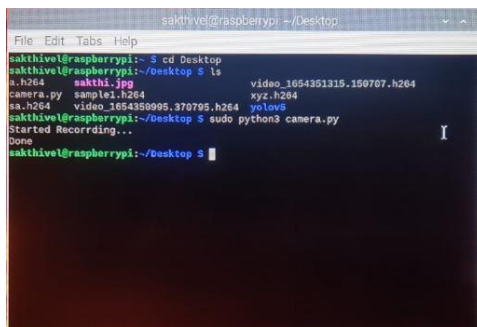


Fig 4.3 raspberry pi Camera module

Data Pre-processing. Machine learning models cannot use real-time data which contains missing values, noisy data and it also may be in an unsuitable format. To rectify the data, a data pre-processing method is used. The repetitive images which contain the redundant values will be deleted which helps to upgrade the efficiency and accuracy of the model. The Python libraries that we have used to perform data pre-processing are pandas, NumPy and matplotlib. Every object was selected tentatively and labelled separately using LabelImg and stored in Separate drive folders for efficient retrieval. Each object was specifically selected and computed to get accurate results.

Prediction. It is all about fitting a shape that gets as close to the data as possible. In this project we have included close to 1000 images and trained it separately for obtaining a constructive output, for this reason we get an exact output both in VGG and YOLO V5 algorithms. It also prevents overfitting and underfitting because every image is considered separately.

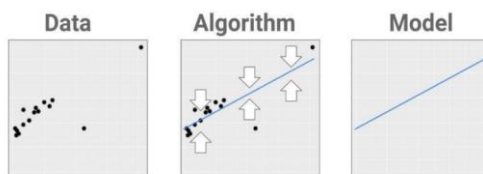


Fig 4.4 Prediction working

VGG16. VGGNet-16 consists of 16 convolutional layers and is very appealing because of its very uniform Architecture. Similar to Alex Net, it has only 3x3 convolutions, but lots of filters. It can be trained on 4 GPUs for 2–3 weeks. It is currently the most preferred choice in the community for extracting features from images. VGG 16 is a convolutional neural network that is 22 layers deep. The network trained on ImageNet classifies images into 1000 object categories, such as keyboard, mouse, pencil, and many animals. VGG 16 is a leading architecture for any object-detection task and may have huge applications in the computer vision sector of artificial intelligence and iot based sectors. The model achieves 92.7% top-5 test accuracy in ImageNet, which is a dataset of over 14 million images belonging to 1000 classes.

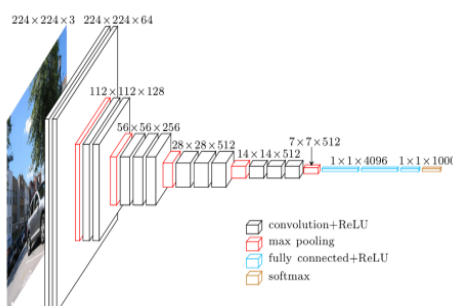


fig 4.5VGG 16 Architecture

YOLO V5. YOLO is an abbreviation for You Only Look once. We are using version 5, launched by Ultralytics in June 2020 and it is now the most advanced way to identify what is available. A convolutional neural network (CNN) that detects real-time objects with great accuracy. This method uses a single neural network to process the whole image, then divide it into sections and predict binding boxes and opportunities for each component. These binding boxes are weighted with expected chances. The method "looks once" in the image in the sense that it makes predictions after a single continuous stream running on the neural network. It then delivers the findings after a non-severe suppression (which ensures that the acquisition algorithm identifies each item only once).

Open CV. In OpenCV, the CV is an abbreviation form of a computer vision, which is defined as a field of study that helps computers to understand the content of the digital images such as photographs and videos. The purpose of computer vision is to understand the content of the images. It extracts the description from the pictures, which may be an object, a text description, a three-dimensional model, and so on. For example, cars can be facilitated with computer vision, which will be able to identify different objects around the road, such as traffic lights, pedestrians, traffic signs, and so on, and act accordingly. Computer vision allows the computer to perform the same kind of tasks as humans with the same efficiency. There are a two main task which are defined below:

Data Labelling and Training. For VGG 16 we first have to split and make the data scalable, then training the dataset will have an adverse effect on the classification and finally evaluating the module will give a structured model. When Yolo V5 comes into play we have to label the images separately. The labelled image will generate a text file containing the axis where the image has been marked and the class number to be stored. Then we separate the text or XML files in a separate file and the image in a separate folder to train the model. The labelling can be accessed through any python compiler, We particularly have used Anaconda Prompt to activate and store the labelled Images.

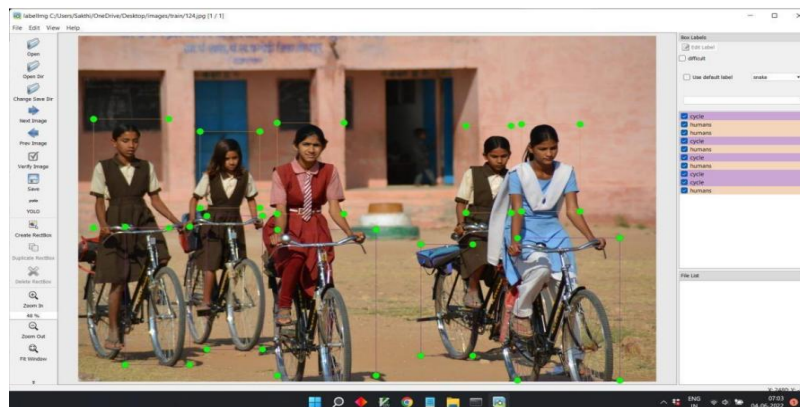


Fig 4.6 labelling images

Object Classification - In the object classification, we train a model on a dataset of particular objects, and the model classifies new objects as belonging to one or more of your training categories.

Object Identification - In object identification, our model will identify a particular instance of an object. VGG16 is a convolution neural net (CNN) architecture which is used in various applications where object identification is a constraint. It is considered to be one of the excellent vision model architectures till date. Most unique thing about VGG16 is that instead of having a large number of hyper- parameters they focused on having convolution layers of 3x3 filter with a stride 1 and always used the same padding and maxpool layer of 2x2 filter of stride. Yolo V5 utilizes OpenCV and real time object detection including videos and live time streaming. It is a family of compound-scaled object detection models trained on the COCO dataset, and includes simple functionality for Test Time Augmentation (TTA), model ensembling, hyperparameter evolution, and export to ONNX, CoreML and TFLite. Model. size.

Example:



fig 4.7 Open CV Example for our model



Fig 4.8 YOLO V5 video output

As the model has been trained to identify the objects in real time when a video is passed, accurate results are displayed on the video file.

5 COMPARISON WITH THE STATE OF THE ART

Finally, we compare our findings to the state of the art. Our "VGG" team placed 2nd in the ILSVRC-2014 challenge's classification task (Russakovsky et al., 2014), with 7 published as a conference paper at ICLR 2015.

An ensemble of seven models yielded a test error of 7.3 percent. We used an ensemble of two models to reduce the error rate to 6.8% after the submission. As seen in Table 7, our very deep ConvNets surpass the previous generation of models, which won the ILSVRC-2012 and ILSVRC-2013 competitions with the best results.

Combined ConvNet models	Error		
	top-1 val	top-5 val	top-5 test
ILSVRC submission			
(D/256/224,256,288), (D/384/352,384,416), (D/[256;512]/256,384,512) (C/256/224,256,288), (C/384/352,384,416) (E/256/224,256,288), (E/384/352,384,416)	24.7	7.5	7.3
post-submission			
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), dense eval.	24.0	7.1	7.0
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), multi-crop	23.9	7.2	-
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), multi-crop & dense eval.	23.7	6.8	6.8

Our performance is similarly competitive with the winner of the classification job (Google Net with 6.7 percent error) and beats the ILSVRC-2013 winning submission Clarifai, which achieved 11.2 percent with outside training data and 11.7 percent without it. This is particularly impressive given that our top result is obtained by merging only two models, which is much less than the number of models used in most ILSVRC submissions. In terms of single-net performance, our architecture outperforms a single Google Net by 0.9 percent (7.0 percent test error). Notably, we did not abandon LeCun et al. (1989)'s classic ConvNet architecture, but rather improved it by significantly increasing the depth.

In ILSVRC categorization, a comparison with the state of the art is made. "VGG" is the abbreviation for our technique. The results produced without the use of external training data are the only ones reported.

Method	top-1 val. error (%)
VGG (2 nets, multi-crop & dense eval.)	23.7
VGG (1 net, multi-crop & dense eval.)	24.4
VGG (ILSVRC submission, 7 nets, dense eval.)	24.7
GoogLeNet (Szegedy et al., 2014) (1 net)	-
GoogLeNet (Szegedy et al., 2014) (7 nets)	-
MSRA (He et al., 2014) (11 nets)	-
MSRA (He et al., 2014) (1 net)	27.9
Clarifai (Russakovsky et al., 2014) (multiple nets)	-
Clarifai (Russakovsky et al., 2014) (1 net)	-
Zeiler & Fergus (Zeiler & Fergus, 2013) (6 nets)	36.0
Zeiler & Fergus (Zeiler & Fergus, 2013) (1 net)	37.5
OverFeat (Sermanet et al., 2014) (7 nets)	34.0
OverFeat (Sermanet et al., 2014) (1 net)	35.7
Krizhevsky et al. (Krizhevsky et al., 2012) (5 nets)	38.1
Krizhevsky et al. (Krizhevsky et al., 2012) (1 net)	40.7

CONCLUSION

In this proposed system, we proposed a smart cane model by integrating IOT, Machine Learning and CNN for efficient object detection and identification and have managed to give constructive and efficient accuracy in the same. The dataset was carefully chosen to be real time to make the project productive as a whole. This can avoid major accidents happening in roads and will save many innocent lives and give a safe and secure environment to the visually impaired. This project can be further developed by establishing an gps tracker and sending a help message to the nearest social activists and sending immediate messages to hospitals. As for the software alexnet can be used to enhance the performance.

REFERENCES

- [1] Wickramasinghe, B. K. T. P., et al. "Smart Cane for Staircase and Water Detection." (2017).
- [2] Zhou, Shifeng. "A Smart Cane to Help the Blind People Walk Confidently." IOP Conference Series: Materials Science and Engineering. Vol. 439. No. 3. IOP Publishing, (2018).
- [3] Rahman, Md Wahidur, Rahabul Islam, and Md Harun-Ar-Rashid. "IoT based blind person's stick." International Journal of Computer Applications 975 (2018): 8887.
- [4] Ghani, Faiz Abdul, and AsnazulfadhliZariman. "Smart Cane based on IoT." International Journal of Education, Science, Technology, and Engineering 2.1 (2019): 12-18.
- [5] Jin, Yongsik, et al. "Smart cane: Face recognition system for blind." Proceedings of the 3rd International Conference on Human-Agent Interaction. 2015.
- [6] Romadhon, A. S., and A. K. Husein. "Smart Stick For the Blind Using Arduino." Journal of Physics: Conference Series. Vol. 1569. No. 3. IOP Publishing, 2020.
- [7] Pawaskar, Pratibha Shantaram, D. G. Chougule, and A. S. Mali. "Smart cane for blind person assisted with android application and save our souls transmission." International Journal of Engineering and Management Research (IJEMR) 8.3 (2018): 235-240.
- [8] Adil, Md, et al. "An IoT based Voice Controlled Blind Stick to Guide Blind People." (2020).
- [9] Soares, Johnny Marcos S., et al. "Smart Cane: Laser guide as an inclusion tool for the visually impaired." Proceedings of the XX Brazilian Symposium on Human Factors in Computing Systems. 2021.
- [10] Rahman, Md Atikur, and Muhammad Sheikh Sadi. "IoT Enabled Automated Object Recognition for the Visually Impaired." Computer Methods and Programs in Biomedicine Update (2021):100015
- [11] Dhanalakshmi, K. S., et al. "Development of Smart Stick for Visually Challenged People." Annals of the Romanian Society for Cell Biology (2021): 17078-17090.
- [12] Abdel-Jaber, Hussein, et al. "Mobile Based IoT Solution for Helping Visual Impairment Users." Advances in Internet of Things 11.4 (2021): 141-152.
- [13] Branig, Meinhardt, and Christin Engel. "SmartCane: an active cane to support blind people through virtual mobility training." Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments. 2019.
- [14] Gaikwad, Arun G., and H. K. Waghmare. "Ultrasonic smart cane indicating a safe free path to blind people." Int. J. Adv. Comput. Electron. Technol.(IJACET) 2.4 (2015): 12-17.
- [15] Sharmila, K., and D. Boomika. "Smart Walking Cane Using IOT." International Journal of Advanced Engineering Science and Information Technology 4.4 (2021)