

NEURAL EYES

REVATHI S

*Department of Computer Science and Engineering
New Horizon College of Engineering
Marathahalli, Bangalore, Karnataka, India
revathis.0289@gmail.com*

C.V. SAMPATH RAJU

*Department of Computer Science and Engineering
New Horizon College of Engineering
Marathahalli, Bangalore, Karnataka, India
sampathvarma666@gmail.com*

Dr. S J Subhashini

*Department of Computer Science and Engineering
New Horizon College of Engineering
Marathahalli, Bangalore, Karnataka, India
drsjsnhce@gmail.com*

DEEPAK C

*Department of Computer Science and Engineering
New Horizon College of Engineering
Marathahalli, Bangalore, Karnataka, India
deepakc9535@gmail.com*

Abstract – Blind people need help as they face major challenges because it is difficult for them to navigate the real world as there can be many obstacles in each step. According to the study, about 40% of visually impaired people report headaches at least once a year. The proposed paper represents work using in-depth reading models that help the visually impaired in real time. The method we proposed using the convolutional neural network (CNN). we have used coco128 data sets to train a model that also helps to distinguish real-time objects by providing user-friendly speech output. With CNN we gain more than 90% of the acquisition of the item, so we have used CNN to better sort things out. Image collection is trained through CNN to identify objects.

Keywords –CNN, Coco128, visually impaired

I. INTRODUCTION

The process and use of machine learning design have impacted in a huge way and changed the way we look at computer vision problem. Normally Graphics methods were used to process any of these difficulties faced by the AI system. It proved to be difficult in using these same methods for detection in live or real time detection. This changed after we learned about the new way of deep learning in classifying of these images. The ability of finding objects in real time has had a profound change in more usability or advantages.

In deep learning, mapping is done using representation-learning algorithms, which are considered simple representations. In other words, in-depth reading demonstrates the concept of an image by combining simple concepts such as dots and lines defined as terminals. The system can be trained to achieve the output you want using a variety of algorithms, a set of measurement data and appropriate label packages. An important aspect of in-depth education in image classification is the use of convoluted architecture.

Vision plays an important role in gathering information in the environment. The act of noticing things without seeing is a

difficult task. Blind people are primarily dependent on their other senses for this purpose. The act of knowing what is in front of a blind person without touching it (with hands or using another tool) is very difficult. In some cases, physical contact between a person and an object can be dangerous, even fatal. A computer system that detects objects by isolating them using CVM algorithms and providing audio or Braille feedback can help to a great extent by providing a more accurate understanding of the environment and eliminating the threats involved in this process. The study is based on the TensorFlow categories. Flow tensor is a built-in goggle used in numerical calculations Flow Tensor provides excellent interaction displaying of machine learning algorithms and an application implementation of these algorithms [1].

II. Related work

A. Acquiring Classifiers

We tend to describe the way or the method in which the classifiers based on human visual. We first examined the best-known method used by normal humans to process the task and the processed into the system

B. Classification Images

The name classification of images states the simple method used by humans to determine or identify the objects[2].

The main objective is to classify or identify these images like a normal human observes to distinguish between two different things easily.

The key point of this classification of images is that we must perform many trials to estimate the decision function $f(.)$ that can distinguish between two different images or objects as $f(.)$ can also make the faults we can obtain good functional model. This model that is obtained can help us view how the human visual system differentiates between images or objects.

Generally, these classifications of these images are thought to be linear but then science proves that human determining or

object detection is nonlinear but we seek only linear method as to implement and obtain a nonlinear method it requires more training and which in turn becomes more expensive but as practice makes linear models perfect, we also wish to share the efforts put in the study of nonlinear models that classify images.[2]

C. Convolutional Neural networks (CNN)

CNN is similar to other architectural forms such as ANN [4] using a perceptron and a unit of machine learning algorithm that helps analyse data. Interneuron connectors have synaptic loads that can be adjusted in training to collect data, so a fully trained network can detect the image correctly when the same object is taken. The network consists of multiple layers such as Convolutional layer, integration layer and fully integrated layers. each layer contains many neurons that respond to different combinations of previous layers. For better recognition training CNN typically uses 5 to 25 defined layers. A set of “labelled” data also uses methods to define the weights of the central and end-to-end neurons. On linearity such as the task of trapping and finding inputs and loads uses a dot product. [4].

Convolution layers is really important in CNN because it helps to extract different features. Kernel weights are chosen for training purpose. Regional can also be extracted by convolutional layer to not allow receptive field local hidden layer. The advantage of CNN with compare to predecessors is it automatically capture the important features with no need of human supervision. We use this feature to train the algorithm by giving many pictures of cats, dogs and many day-to-day life pictures. It is also Computationally efficient.

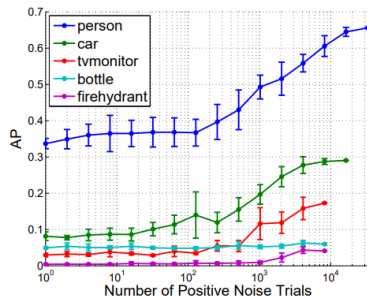


Figure 1: Noise trails

Figure 1: This represents the object classification performance on PASCAL VOC with CNN features with number of positive noise trails as these trails increase performance of these trails increase.

CNN has convolutional layers, ReLU layers, pooling layers and a fully connected layer this type of network is really good for processing the 2D images and if we compare the classification method the CNN is best among them all. Most the time 90% accuracy is achieved by using CNN method to capture and classify the image. It basically extracts features from the image eliminating the manual method and because it can have many hidden layers in each layer increases the complexity of learning features.

CNN first takes input image and starts applying the filters to get the feature map and then later applies Relu function to improve non linearity and gradually starts applying pooling layer for each feature map also flattening pool images into a long vector and later adds this vector into artificial neural network helping the image to be process in the network. Basic CNN Architecture contains following parts shown in the below figure 2.

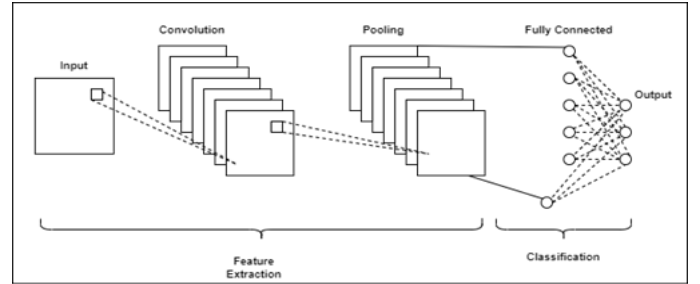


Figure 2 : Basic CNN Architecture

D. TensorFlow

Tensor flow is a open library that computes large machine learning. It makes sure that it combines both machine and deep learning that are in models and algorithms. The main script utilized by it is python as its very useful and easy to use as front end for constructing of applications that includes framework while running the applications. It can also run and help the networks get familiarized or also called as trained for image as well as digit as well as word to help classify also proficient in RNN also sequence to sequence for the translation of machine and natural language process. Tensor flow supports production prediction at scale with same models used for training [1].

E. Feature Extraction

The main purpose of this section is to talk about extracting a feature from the obtained zebra-crossing ROI which indicates that it is important to detect an object. This is also important because it is part of the computer vision. For good recognition is resized to 128x128 pixels. Then the modified circuit crossing zebra pattern Uniform Local Binary [5] similarities are later extracted then it will send to the Vector Support Machine [5] for reference. The fixed time of eight nearby pixels also radius aren't similar binary pattern of the area used. Thus, 59 i.e., Px (P-1) +3. Is image size element of the local Binary pattern.

F. Image recognition using deep learning

The main thing discussed is that it uses algorithms that is developed using machine learning using which the model is trained that develops cognitive abilities. This technique begins by labelling the special things that are present in each dataset, in them all the image databases and to recognize model machine learning is used to train. Visual image algorithms made with the usual method have many limitations. These first-time processing methods require a lot of time as they extract features and at the same time search for the same. Then there is the endless element of the element where the categories of objects increase in size [5]. Also, using manual emissions may reveal

flawed features and this may reduce visual acuity. But more time is needed to examine all the factors. Generally, algorithms are trained to be able to distinguish a fixed number of objects. According to the identification provided our system should be able to label and detect many types of objects or objects. Using this recognition method each item has its own image recognition algorithm. This increases the complexity of nutrition [6]. To attain the goal of seeing more things, a deeper studying approach is used that can automatically learn different aspects and extract them later. The best image recognition can be achieved through Convolutional Neural Networks (CNNs) networks, which are used and proven to be more reliable than the current in-depth study method. CNN is very useful as it is good for reducing image quality and making the network much easier. Our model seems to be successful in solving the common problem of image classification using CNN, Google, Microsoft, VGG seems to have greatly improved the techniques and the more accurate learning models. Through those courses, we have a better and more advanced learning model and the main goal is to speed up the recognition process. Later, the problem of awareness changed from retreating to fragmentation problems. These models offer more complete accuracy than traditional methods.

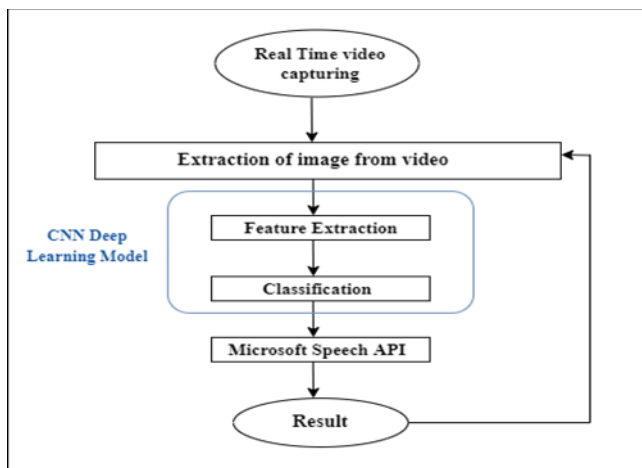


Figure 3: Data Flow Diagram

Figure 3: Using a camera of 5MP spec, a video is captured, and image frames are extracted from this video using Open CV Library. The extracted image is then inserted into the CNN model from which the high resolution, crisp image is resized without losing any critical features. The framework extracts the high-level features such as the edges, from the loaded image by dividing it into grids. This process is called Feature Extraction. Once the extraction is completed the image classification and localization are applied on each grid using a SoftMax classifier. The model now identifies the objects by predicting the bounding boxes and the corresponding class probabilities for objects. Finally, the recognized words are converted into the corresponding speech using pyttsx3 library.

000000000071 - Notepad				
File	Edit	View		
2	0.752648	0.525833	0.043359	0.033216
2	0.835727	0.538498	0.038047	0.028169
2	0.700875	0.521455	0.036437	0.020235
2	0.937398	0.559777	0.042422	0.033592
6	0.452477	0.576408	0.755359	0.280892
2	0.794984	0.538779	0.036812	0.02615
2	0.953469	0.514906	0.026125	0.012723
7	0.539727	0.464061	0.037328	0.023897
7	0.58518	0.471397	0.037984	0.016596
2	0.982555	0.572371	0.034891	0.030047
2	0.767367	0.531279	0.037391	0.029695
2	0.617961	0.476373	0.025016	0.01007
2	0.588344	0.471491	0.030781	0.018333
2	0.560102	0.471303	0.030234	0.011995
2	0.796047	0.521796	0.029969	0.014577
2	0.734094	0.523427	0.029375	0.02723

Figure 4: data labelling

Figure 4: The first number in each line is the category number of the target, the second is the abscissa of the center point of the frame, the third is the ordinate, the fourth is the width of the frame, and the fifth is the height of the frame. The coordinates of the center point and the width and height are the normalized results.



Figure5: Image segmentation and object detection

G. Image To Text Converter

Camera attached using raspberrypi first gets the image input that needs to read and these images file are stored in folder. A software called tesseract [5] added to the Raspberry pi gadget using a command module. Commands are used to install it is: `sudo apt get install tesseract-ocr`, This software converts the captured image file and converts it into a text file by extracting text by the image and saving of text file of the .txt extension.

H. Algorithm

You Only Look Once(YOLO) [12] is an abbreviation for You Only Look Once. In the year 2016, it is proposed. Rather than employing a classifier to do the detection, the

entire process is carried out in a single network. In a single study, single neural network generates bounding boxes and class of probabilities, bringing its performance to a whole new level. Because the entire process is a single pipeline, it may be further optimized.

For training after every 1000 epochs of training, on the validation dataset, the model would be evaluated to calculate precision, remember, and imply common precision (mAP). This method tests for the following:

- When accuracy stops growing, the version no longer desires extra schooling to save you over fitting.
- Depending on overall performance, hyper parameters must receive changes to optimize for assessment metrics.

After the model has finished education, use the testing and validation datasets to evaluate the community's performance. The version attracts a bounding box around correctly detected objects with a self-assurance rating of .50 or better.

The following performance metrics were applied to provide consequences.

- Precision and Recall represents whether the model successfully detects the Main crop and weed in an image.
- $Recall = \frac{TP}{TP+FN}$
- $Precision = \frac{TP}{TP+FP}$

I. Text To Speech Conversion

This speech program or process interface is developed by microsoft to recognise the speech and its synthesis in the OS. There are many versions of it released over the years which are within the OS itself or given as a seperate SDK.

Applications that use this are MS office, MS agent and MS speech server.

All process or its versions are scripted in a way that any user can develop and utilize it to speech recognize and produce it using a set of standards that can be accessed by different scripts. Text-to-speech work with or get accustomed to existing models of it. These same engines can be defined and used replacing the MS supplied engine[5].

III. LITERATURE SURVEY

Garima Jain, Samkit Shah, Mayur Ghevariya, Sarosh Dastoor, JayrajBandariya, [2], Talk about using a CNN model trained using the COCO 2017 database and using Haar cascade which

is the fastest rated algorithm that can be fitted with an ultrasonic detection sensor. the distance between an object and a device that uses text-to-speech conversion to provide information can also work.

Neel Parikh, Ishita Shah and SafvanVahora [4], talked about the acquisition of in-depth readings using a different CNN model but the limit of using this method is using a high-end server with GPU processing via an internet connection that can always interrupt if necessary. features not met.

Saleh Alghamdi, Ron van Schyndel and Ibrahim Khalil [6] They talk about using a headmount to take pictures and process these images using the SURF object recognition algorithm that provides a safe safe for the safest that a user can take from the environment that provides 90% of human detection and 80% of the associated artifacts found to use this type of System reliably. especially if there are other pedestrians which can make the user less confident to walk on the streets when there are fewer people walking.

Deepthi Jain B, Shwetha M Thakur and K V Suresh, [7] They Talk about using headmount by visually impaired people who receive speech commands recognizing sign boards and moving objects around them using a raspberry pi using python language as artificial intelligence. using the Haar cascade algorithm and because it takes the speech input it becomes more reliable for the user to navigate his position.

Devashish Pradeep Khairnar, Apurva Kapse, Dr. Geetanjali Kale, Prathamesh Jadhav [8], The system that they speak about is using CNN architecture to assist the blind people as it shows the type of the object and the distance object around them and it is also user friendly. The system can be iproved using Dijkstra's algorithm to find the shortest path.

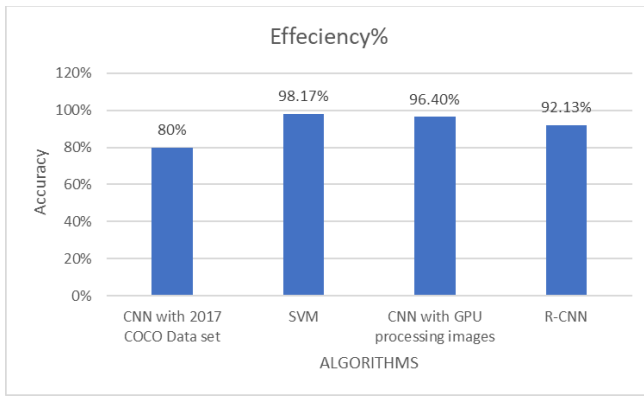
Dhruv Dahiya, Hardik Gupta, Malay Kishore Dutta[9], They speak about using a CNN based algorithm that helps the blind people navigate through their surroundings the CNN algorithm was trained using a set of 450 images that gives a accuracy of 92.13% because of using a such a process it provides good real time detection of the surroundings.

Faruk Ahmed, Md Sultan Mahmud, and Mohammed Yeasin [10], They speak about using a interactive and portable footpath assistive device that uses CNN algorithm. It works because of previously trained DNN models to the footpath obstacle detection one of the limitation of this model is not having vocabulary speech to increase to usability of the device.

Carl Vondrick, Hamed Pirsivash, Aude Oliva, Antonio Torralba [11], They speak about using SVM algorithm main goal of this paper was to transfer human bias to machine to help object detection generalize across datasets.

TABLE I. SURVEYED ON ALGORITHMS AND METHODOLOGY

SNO	Year of Publication	Author	Study	Algorithm/ Methodology	Advantages	Limitations
1	2019 IEEE Xplore	Garima Jain Et(al.)	This paper talks about a technique which used to assisting visually impaired people. They speak about using a CNN model that is trained using COCO 2017 dataset and also uses Haar cascade which is the fastest algorithm.	Algorithm: Object detection using python, Haar Cascade	User friendly, 80% accuracy	can be fitted with a ultrasonic sensor to gauge the distance between the object and also a device that uses text to speech conversion to provide information can also be implemented.
2	2019 IEEE Xplore	Mahinul Islam Meem Et(al.)	Their reasearch paper speaks about aiding the blind using a SVM system to help the blind cross zebra crossings	Efficiently detects zebra crossing using binary pattern recognition.	98.17% accuracy in detecting Zebra crossing	This proposed method is only highly accurate in detecting zebra cross and does not function better with unusual shapes in zebra crossing such as round zebra crossing.
3	2018 IEEE Xplore	Neel Parikh Et(al.)	They speak about a deep learning based object detection using different CNN model	CNN based algorithm, GPU for processing the images	Highest racogniti on rate of 96.4% was achieved using inception V3 CNN model	The system doesn't work if there is no active internet connection and good high end smart phones would provide better assistance.
4	2020 IEEE Xplore	Dhruv Dahiya Et(al.)	They speak about using a CNN based algorithm that helps the blind people navigate through their surroundings the CNN algorithm trained using an amount of 450 images that gives a accuracy of 92.13% because of using a such a process it provides good real time detection of the surroundings.	R-CNN algorithm	92.13% accuracy using set of 450 images to train	Detects 11 out of 140 images that are used for training.
5	2018 IEEE Xplore	Deepthi Jain B Et(al.)	Here, a method of assisting the visually impaired is used. The plan proposed here is simple and friendly to the build, letting the user feel independent in his home.	Haar Cascade Algorithm: Open CV library Python has functions specifically to detect objects.	User friendly,	Less accuracy, algorithm can be updated.



Graph 1: Representation of efficiency

Parameters	Haar Cascade	CNN	HOG+SVM
Accuracy	Poor	Good	Average
Performance	Poor	Good	Poor
Computational time	Good	Poor	Good
Computation Complexity	Good	Poor	Average

Table II: Comparison Between Different Object Detection Algorithms

The graph 1 shows the comparison of different algorithm used and the accuracy. SVM and CNN algorithm with GPU processing the images shows the highest accuracy achieved with other algorithm as R-CNN and CNN neural network trained using 2017 COCO Dataset also showed good accuracy.

IV. RESULTS

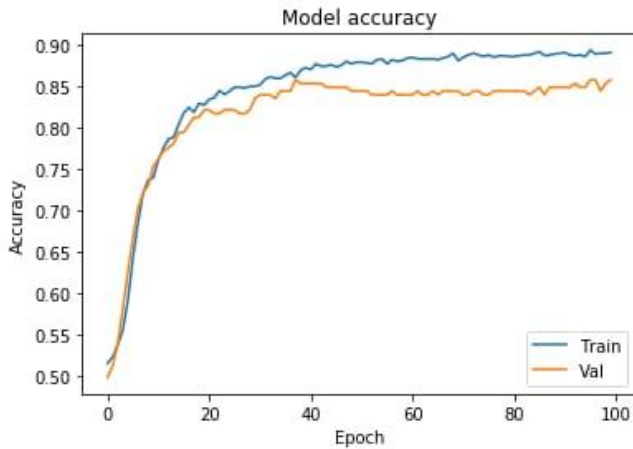


Figure6 : Train-Validation Accuracy

Figure 6: As the graph shows in training dataset as the epochs is increasing accuracy is also increasing proportionally but in validation accuracy was highest at 4 epochs and did not show any significant increase in accuracy past 4 epochs so 4 epochs is appropriate for training the module to have maximum accuracy.

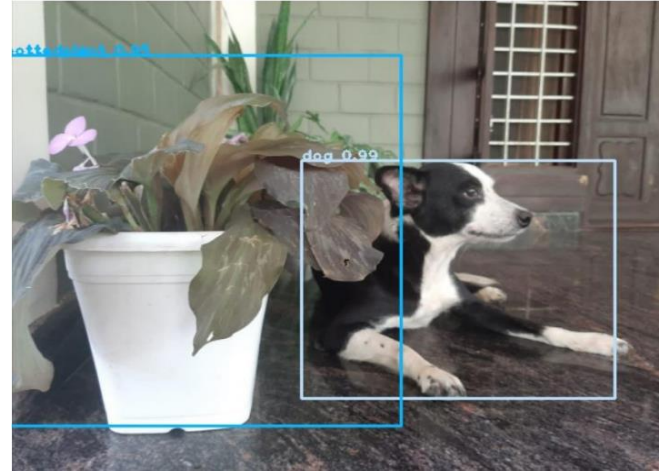


Figure7: Obtained Results after training the model

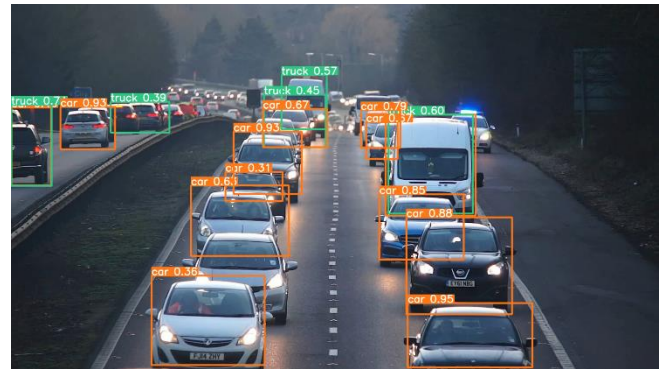


Figure8: Obtained Results after training the model .

V. CONCLUSION AND FUTURE WORK:

The system proposed here is for obstacle detection and identification to implement in the computer vision domain, the system mainly shows how the progress of CNN based schemes have made it most suitable for processing the images. TensorFlow is used to achieve object detection for a live scene. Unlike other existing models, it does not require a large database because of the pre-trained Cognitive Neural Network model. The CNN shows to drop in 2% of detection rate. As the graph above represents system can detect initially faster with simple still images as the algorithm gets trained more the CNN network would be able to detect the object even more efficiently it would even detect movable objects giving a 96.40% by implementing CNN.

The system can be made portable in future and with the aid of more complex processors and further improvements, the system may recognize any kind of object irrespective of its nature and scope.

REFERENCES.

- [1] PriyalJawale, H.P.Chaudhary, et.al. “ Real-Time Object Detection using TensorFlow” *International Research Journal of Engineering and Technology (IRJET)*., vol. 07 Issue: 08 Aug 2020 e-ISSN: 2395-0056 p-ISSN: 2395-0072.
- [2] Garima Jain ,Samkit .S, et.al ” CNN based Auto-Assistance System as a Boon for Directing Visually Impaired Person” *Proceedings of the Third International Conference on Trends in Electronics and Informatics (ICOEI 2019) IEEE Xplore Part Number: CFP19J32-ART; ISBN: 978-1-5386-9439-8*.
- [3] Mahinul Islam Meem, Pranab Kumar.D, et.al.” Zebra-Crossing Detection and Recognition Based on Flood Fill Operation and Uniform Local Binary Pattern,” *2019 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 7-9 February, 2019.
- [4] Neel Parikh, IshitaS, et.al ”Android Smartphone based Visual Object Recognition for Visually Impaired using Deep Learning,” *International Conference on Communication and Signal Processing*, April 3-5, 2018, India.
- [5] Rithika.H, B. Nithya santhoshi,”Image Text To Speech Conversion In The Desired Language By Translating With Raspberry Pi,” 978-1-5090-0612-0/16/\$31.00 ©2016 IEEE.
- [6] Saleh Alghamdi, R.V.Schyndel, et.al ” Safe Trajectory Estimation at a Pedestrian Crossing to Assist Visually Impaired People,” *IEEE Engineering in Medicine and Biology Society. Conference · August 2012*.
- [7] Deepthi Jain B, Shwetha M Thakur, et.al “Visual Assistance for Blind using Image Processing,” *International Conference on Communication and Signal Processing*, April 3-5, 2018.
- [8] D.Pradeep Khairmar, et.al ” PARTHA: A Visually Impaired Assistance System,” *2020 3rd International Conference on Communication System, Computing and IT Applications (CSCITA)*.
- [9] Dhruv Dahiya, Hardik G, Malay Kishore.D, “A Deep Learning based Real Time Assistive Framework for Visually Impaired,” *2020 International Conference on Contemporary Computing and Applications (IC3A) Dr. A.P.J. Abdul Kalam Technical University, Lucknow. Feb 05-07, 2020*.
- [10] Faruk Ahmed, Md M.Sultan, et.al,”An Interactive Device for Ambient Awareness on Sidewalk for Visually Impaired,” 978-1-5386-5959-5/18/\$31.00 © 2018 IEEE.
- [11]Carl Vondrick, H.Pirsiavash ,et.al.”Acquiring Visual Classifiers from Human Imagination”.
- [12]S Prasad, M Thangatamilan , M Suresh., HPanchal., C Rajan , C Sagana., B Gunapriya., A Sharma , T Panchal , K K Sadasivuni. “An efficient LoRa-based smart agriculture management and monitoring system using wireless sensor networks”, *International Journal of Ambient Energy*,2021.