ISSN [Online]: 2583-2654

An Innovative approach to provide Communication Interface between Deaf & Dumb people

[1]Thara Devi M, [2]Gangothri KR, [3]Imaad Uwaiz, [4]Keerthana R, [5]Lakshmi V
Department of Computer Science and Engineering,
Dr T Thimmaiah Institute of Technology,
Kolar Gold Fields, Karnataka,
India

Abstract: The deaf and dumb people face problems in communicating with others. Addressing the problems and issues of individuals with Hearing and Vocal Impairment through single aided system may be a powerful job. The paper focuses on finding a unique technique that aids the visually impaired by letting them hear what is represented as text, and letting them visualize what is been represented as voice command, it is achieved by the technique that captures the video through a camera and convert it text form and visualize which is in audio form by speech to text conversion technique. Our goal is to design a desktop human computer interface application that facilitates communication among such people with normal ones. This can be achieved by making use of YOLO algorithm. This algorithm out performs the other detection methods, including DPM and R-CNN, when generalizing from natural images to other domains.

Keywords: Hand Gestures, Image Processing, Annotation, Label Image, Feature Extraction, Training, Testing, Tensor Flow Object Detection.

I. INTRODUCTION

One of the most precious gift to a human being is an ability to see, listen, speak and respond according to situation. We aim for developing prototype model for dumb and deaf people by employing in single compact device. Sign language is an expressive way of communication between normal and dumb-deaf people in order to improve the life style of dumb and deaf people. The deaf and dumb people are not involved with the social world because of their disabilities. Unintentionally, they are treated in an unusual manner by the rest of the society. Sign language is a communication skill that is used to convey a meaning of a speaker‟s thought using gesture. It is a well- structured code gesture; each gesture has a meaning assigned to it.Gesture recognition is gaining importance in many application areas such as human interface communication, multimedia and security. Normal person face problem in communicating with disabled people because they cannot understand sign language. There are not many sign language institutions in our society. So, many of dumb people use usual sort of sign language to communicate and they do not have a customized sign language. It is also not possible for the many people to learn sign language. Therefore, a large communication gap still exists between dumb, deaf and normal people.

II. SYSTEM ARCHITECTURE

System architecture is the conceptual model that defines the structure, behavior, and more views of a system. An architecture description is a formal description and representation of a system, organized in a way that supports reasoning about the structures and behaviors of the system The purpose of system architecture activities is to define a comprehensive solution based on principles, concepts, and properties logically related to and consistent with each

ISSN [Online]: 2583-2654

other. The solution architecture has features, properties, and characteristics which satisfy, as far as possible, the problem or opportunity expressed by a set of system requirements and life cycle concepts.

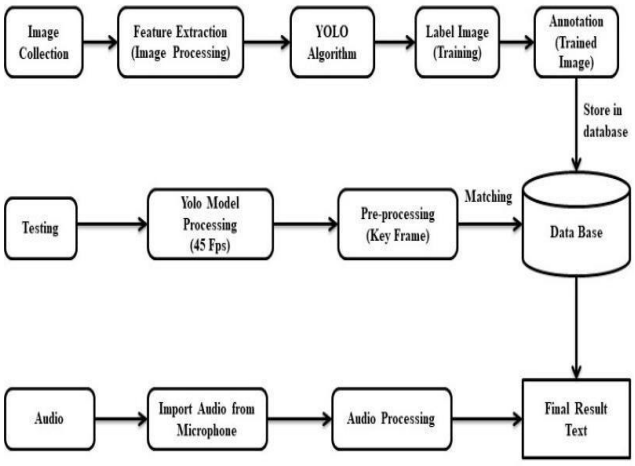


Fig.1 System Architecture

Initially, we need to collect different types of hand gesture images from camera which will be used for training the model**.** Further, an image labeling or annotation tool is used to label the images for bounding box object detection and segmentation. Labeling is a graphical image annotation tool. After labeling images then those labeled images should be imported to train and test model. Finally we will build and train the model using training data set. The generated model is validated by matching with already clustered and featured samples. The result is displayed in the form of text and audio. The audio of the final result will be generated from the system microphone.

III. IMPLEMENTTION MODULES

The implementation phase involves putting the project plan into action. Project implementation is the phase where vision and plans become reality. Project implementation consists of carrying out the activities

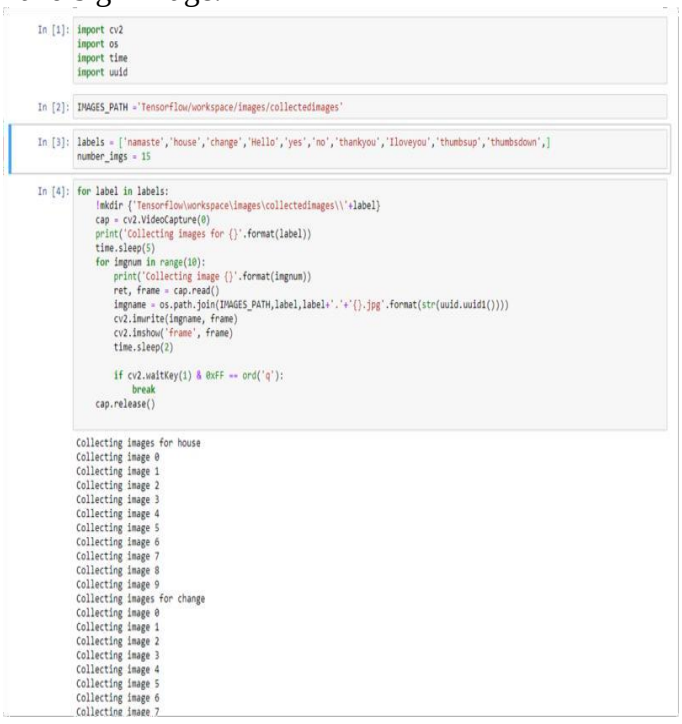
with the aim of delivering the outputs and monitoring progress compared to the work plan.

- Image collection
- Labeling Images
- Importing Images
- Create Label Map
- Create TF Records
- Training data

3.1 Image Collection

In this module, the camera captures the images of hand gesture. Further we will convert to frames and resize and arrange the frames for future use. Image resizing or image scaling, is a geometric image transformation which modifies the image size based on image interpolation. This image scaling process can increase or decrease the resolution of a target image so that the absolute size of image data is adjusted.

Computers are able to perform computations on numbers and are unable to interrupt images in the way that we do. We have to somehow convert the images to numbers for the computer to understand. The computer will assign each pixel a value based on how dark it is. All the numbers are put into an array and the computer does computations on that array. We then feed the resulting array for next step**.** The image undergoes various preprocessing and processing stages to extract features and finally identification, the various processing stage of hand sign image.



```
In [1]: import cv2
        import os
        import time
        import uuid

In [2]: IMAGES_PATH ='Tensorflow/workspace/images/collectedimages'

In [3]: labels = ['namaste','house','change','Hello','yes','no','thankyou','Iloveyou','thumbsup','thumbsdown',]
        number_imgs = 15

In [4]: for label in labels:
            !mkdir {'Tensorflow\workspace\images\collectedimages\\'+label}
            cap = cv2.VideoCapture(0)
            print('Collecting images for {}'.format(label))
            time.sleep(5)
            for imgnum in range(10):
                print('Collecting image {}'.format(imgnum))
                ret, frame = cap.read()
                imgname = os.path.join(IMAGES_PATH,label,label+'.'+'{}.jpg'.format(str(uuid.uuid1())))
                cv2.imwrite(imgname, frame)
                cv2.imshow('frame', frame)
                time.sleep(2)

                if cv2.waitKey(1) & 0xFF == ord('q'):
                    break
            cap.release()

        Collecting images for house
        Collecting image 0
        Collecting image 1
        Collecting image 2
        Collecting image 3
        Collecting image 4
        Collecting image 5
        Collecting image 6
        Collecting image 7
        Collecting image 8
        Collecting image 9
        Collecting images for change
        Collecting image 0
        Collecting image 1
        Collecting image 2
        Collecting image 3
        Collecting image 4
        Collecting image 5
        Collecting image 6
        Collecting image 7
```

Fig.2.Image Collection

ISSN [Online]: 2583-2654

From the physical environment image is acquired using a camera. Subsequent step is pre-processing that include skin color extraction, reducing dimension using blurring, RGB to binary conversion, edge detection. After preprocessing image feature is extracted by different steps like obtaining the contour, calculating the centroid, determining the hull and defect points etc. This gives the corresponding number values by identifying the number of fingers. Finger alphabets can be recognized using template matching algorithms from the database constituting template images created during the training phase.

3.2 Labeling Images

An image labeling or annotation tool is used to label the images for bounding box object detection and segmentation. Labeling is a graphical image annotation tool, annotations are saved as xml files in PASCAL VOC format, the format used by Image net. It also supports YOLO and creates ML formats.

Bounding boxes: Bounding boxes are the most commonly used type of annotation in computer vision. Bounding boxes are rectangular boxes used to define the location of the target object. They can be
determined by the x and y axis coordinates in the upper-left corner and the x and y axis coordinates in the lower right corner of the rectangle. Bounding boxes are generally used in object detection and localization tasks. Bounding boxes are usually represented by either two coordinates (x1,y1) and (x2,y2) or by one coordinates (x1,y1) and width (w) and height (h) of the boundingbox. Steps for selecting image (Pascal VOC)

1. Build and launch using the instructions above.
2. Click „change default saved annotation folder" in Menu/File.
3. Click „Open Dir".
4. Click „create Rectangular BOX".
5. Click and release left mouse to select a region to annotate the rectangular box.
6. You can use right mouse to drag the rectangular box to copy or move it.

When pressing space, the user can flag the image as verified, a green background will appear. This is used when creating a dataset automatically.

Fig.3.Image labeling command

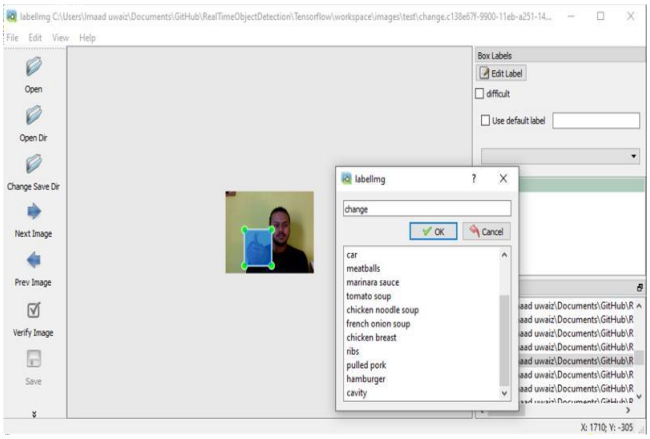


Fig.4.Labelling image

3.3 Importing images

After labeling images then those labeled images should be imported to train and test model. 80% of labeled images with annotation in train folder and 20% of labeled mages with annotation in test folder. To import hand gesture images to jupyter notebook.

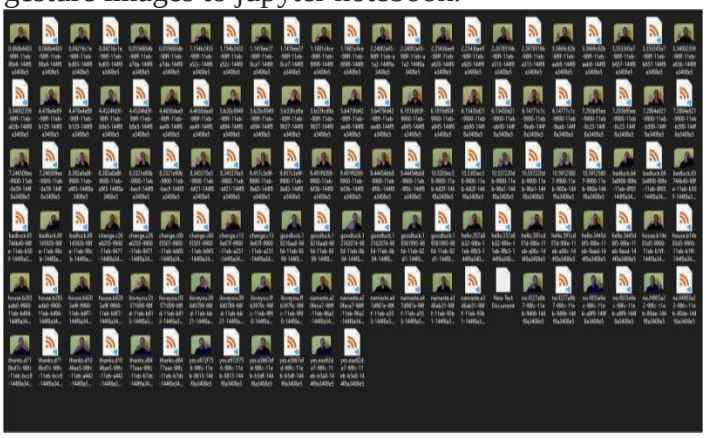


Fig.5. Importing images

ISSN [Online]: 2583-2654

3.4 Create Label Map

Creating a label map that contains gesture name and id. Tensor flow requires a label map, which namely maps each of the used labels to an integer values. This label map is used both by training and detection processes.

The label map maps and id to a image. We will put it in a folder called training, which is located in the object detection library. The id number of each item should match with the id of specified in the generate tfrecord.py file. The label map in Yolo maps integers to a class list specified in the label map, each integer maps to a position in the list and expresses class labels in the annotations. It is important to note that different label maps function slightly differently from format to format.

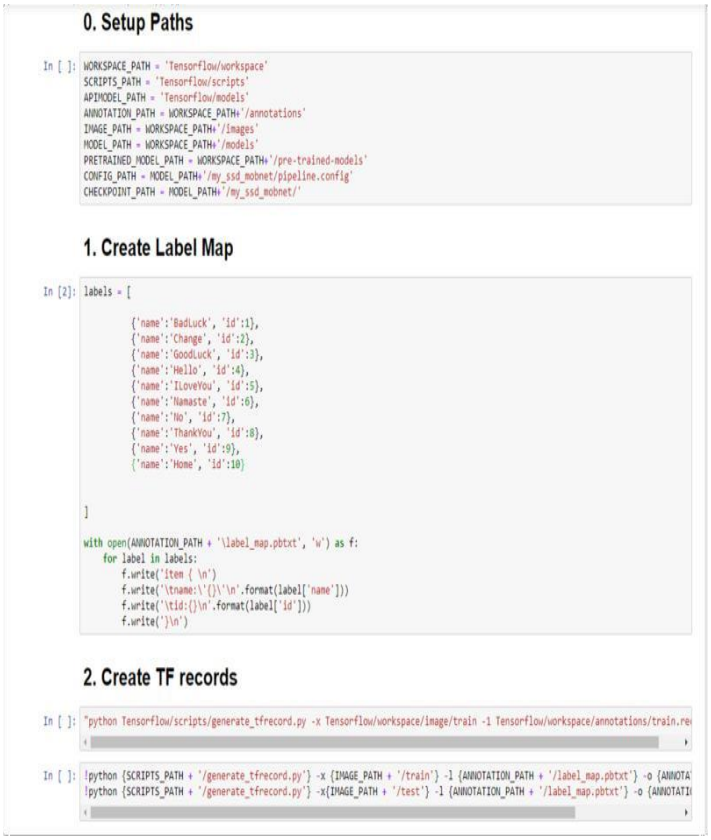


Fig.6.Creating Label Map

3.5 Create TF record

Creating TF record for both training and testing images, while TF records are "incompatible" with other file readers

1. As binary file formats, a TF record file takes up less disk space, which means every read/write operation that needs to be performed is faster.
2. TF records files are optimized to handle component parts of larger dataset. So, for example, if a given dataset exceeds the size of a given machines memory, streaming a subset of the dataset is easily done. This is exactly what happens when training on a single batch of data: the machine is using a subset of the overalldata.
3. TF record files are also optimized for stored sequenced data.

This is particularly important for word or time sequences,similar to how data can be easily broken into component pieces and streamed.

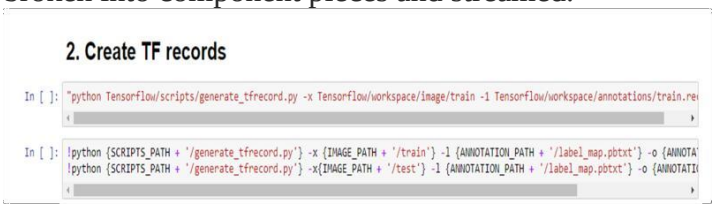


Fig.7.Creating TF record

A TF record file contains our training data. In the context of deep learning, that often includes having both an annotation and an image. Images are encoded to integer representations. Annotations are encoded to describe where in an image a given bounding box is, and an integer representation of that bounding box"s class. Critically because we are converting our annotations into integer representations as well, we need a dictionary that maps our integers back to our string value object names.

YOLO ALGORITHM

Object detection is one of the classical problems in computer vision. Since the problem of object detection is more complex than classification, so Yolo algorithm uses a complete different approach.

Yolo (you only looks once) algorithm is implemented

by Joseph Redman in 2015. Yolo algorithm employs Convolution neural network to detect objects in real time. As the name suggests, the

ISSN [Online]: 2583-2654

algorithm requires only a single forward propagation through a neural network to detect objects. This means that prediction in the entire image is done in a single run.

Yolo algorithm is based on regression instead of selecting the interesting part of an image it predicts classes and bounding boxes for the whole image in one run of the algorithm. Our base Yolo model processes images in real time at 45 frames per second, fast Yolo processes an astounding 155 frames per second.

Yolo divides up the images in to grid of 13 by 13 cells. Each of these cells is responsible for predicting 5 bounding boxes, it applies a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities of each region.

Yolo algorithm works using the following three techniques

1. Residual blocks: First, the images are divided into various grids. Each grid has dimension of S*S. Every grid will detect objects that appear within them.
2. .Bounding box regression: A bounding box is an outline that highlights an object in an image. Every bounding box in the image consists of the following attributes
- Width (bw)
- Height (bh)
- Class (for example, person, car, etc) this is represented by the letter c
- Bounding box center (bx, by)
3. Intersection over union (IOU): It is a phenomenon in object detection that describes how boxes overlap. Yolo uses IOU to provide an output box that surrounds the objects perfectly.

Working of YOLO algorithm

First, the image divides into grid cells. Each grid will forcasts B bounding boxes and provides their confidence scores. The cells predict the class probabilities to establish the class of each object. All the predictions are made simultaneously using a single convolution network.Intersection over union ensures that the predicted bounding boxes are equal to the red boxes of the objects. This phenomenon eliminates unnecessary bounding boxes that do not meet the characteristics of object (like height and width). The final detection will consist of unique bounding boxes that fit the objects perfectly.We obtain the class-specific confidence score as: Pr(class i/object)*Pr(object)*IOU=Pr(class i)*IOU

Advantages:

1. Speed: This algorithm improves the speed of detection because it can predict objects in real time.
2. High accuracy: Yolo is a predictive technique that provides accurate results with minimal background errors.
3. Learning capabilities: The algorithm has excellent learning capabilities that enable it to learn the representations of the objects and apply them in object detection.

3.6 Training Model

After model construction it is time for model training**.** We were able to build an artificial convolution neural network that can recognize images. Split the dataset into train and test dataset.

Finally we will build and train the model using training dataset. The proposed framework system can be visualized in two modules: Training and testing i.e. gesture recognition. For this training, 20 different classes of images corresponding to 20 different hand signs are used. Each class of images contain data corresponding to same sign but in all possible poses. This helps in designing a pose invariant model for hand sign detection. Transfer learning is implemented due to availability of smaller dataset. Region of interest in the test image is detected using the YOLO architecture.

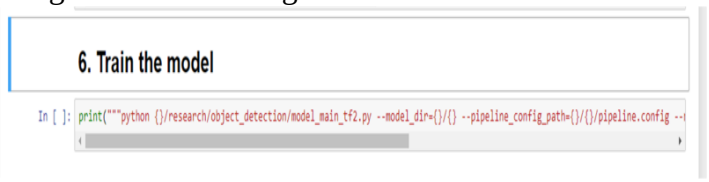


Fig.8.Training Model

ISSN [Online]: 2583-2654

3.7 Output

After successfully training the model and storing them as checkpoints. We then load those trained checkpoints for hand gesture detection. The output of the hand gesture is displayed with the accuracy of the sign language in the detection box

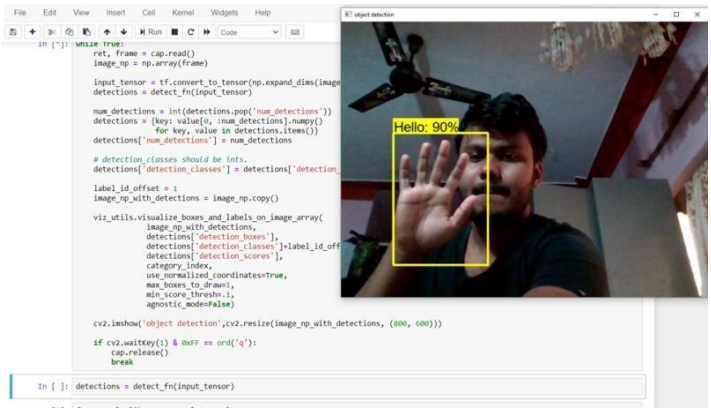


Fig.9.Output

IV CONCLUSION

Sign language is a very useful to ease useful the communication between the deaf community and the normal people. Yet there is a communication barrier between these deaf and dumb, with the normal people. This project aims to overcome the communication gap between these communities. In this project a system for hand gesture identification is developed. The main objective of the study is to develop a system that can recognize the human generated gestures and make use of this data and converts it to text and audio. This system is divided into two basic parts. The first part is the static sign recognition that is based on posture recognition such as alphabet recognition. The second part is the voice recognition is either isolated or continuous. This speech is further converted to text and image.

V FUTURE WORK

At present we have developed an desktop application which recognize hand-gestures which is used in sign-language communication. Also we can extend this work for mobile devices, where mobile camera takes the input recognize the action and displays the result . In future using gestures like eye blinks , head motion and actions would help them to communicate effectively.

REFERENCES

[1]. Shoaib Ibrahim Shaikh , Ismail Mehmood Memon "Communication system to help deaf and dumb communicate with normal people"International Research Journal of Engineering and Technology (IRJET) ,Volume: 03 Issue: 04 , pp. 1793-1799,April-2016.

[2]. Shraddha R. Ghorpade, Prof. Surendra K. Waghmare"A Communication System for Deaf and Dumb People" International Journal of Science and Research (IJSR), Volume:04 Issue: 9, pp.1914-1917, September 2015.

[3]. Bindhushree C, Chaithra M, Monisha G "Communication System For Blind, Deafand Dumb People Using Internet of Things" International Research Journal of Engineering and Technology (IRJET) , Volume: 06 Issue: 04 ,pp.4381- 4384,Apr 2019.

[4]. Sruthi C. J and Lijiya"A Deep Learning based Indian Sign Language Recognition System" International Conference on Communication and Signal Processing, pp. 596-600, April 4-6, 2019.

[5]. Roger Achkar, Gaby Abou Haidar, Dian Salhab, "Sign Language Translator using the BackPropagation Algorithm of an MLP" 7th International Conference on Future Internet of Things and Cloud Workshops, pp. 31-35, 2019.

[6]. Surbhi Rathi, Ujwalla Gawande "Development of full duplex intelligent communication system for deaf and dumb people" 7th International Conference on Cloud Computing,Data Science & Engineering – Confluence , pp. 733-738, 2017.

[7]. Geethu G Nath, Arun C.S "Real Time Sign Language Interpreter" International Conference on Electrical, Instrumentation and Communication Engineering, pp.103- 107, 2017.

[8]. Soeb Hussain and Rupal Saxena "Hand Gesture Recognition Using Deep Learning" International Conference of Learning Representation, pp. 48-49, 2017.

[9]. Sampada S. Wazalwar,Urmila Shrawankar"Multimodal Interface for Deaf and Dumb Communication" 6th International Conference on Computing for Sustainable Global Development,pp. 418-422, 2019.

[10]. Jayatilake,C.Darshana,G.A.Madhuwantha"Communication between Deaf-Dumb People and Normal People: Chat Assist" International Journal of Scientific and Research Publications, Volume 7, Issue 12, pp. 90-95, December 2017.

[11]. Saransh Sharma, Samyak Jain, Khushboo "A Static Hand Gesture and Face Recognition System For BlindPeople"6th International Conference on Signal Processing and Integrated Networks (SPIN),pp. 534-539, 2019.

[12]. AnchalSood , Anju Mishra "AAWAAZ: A Communication

ISSN [Online]: 2583-2654

System for Deaf and Dumb" 5th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) , pp. 620-624, 2016.

[13]. Prof. Prashant G. Ahire,Tejaswini , Pramod B. Warale"Two Way Communicator Between Deaf and Dumb People And Normal People" International Conference on Computing Communication Control and Automation, pp. 641-644, 2015.

[14]. Jaya Shukla, Ashutosh Dwivedi, "A Method for Hand Gesture Recognition"Fourth International Conference on Communication Systems and Network Technologies , pp. 919- 923, 2014.

[15]. Shweta S. Shinde , Rajesh, Vitthal K. Bhosale "Real Time Two Way Communication Approach for Hearing Impaired and Dumb Person Based on Image processing" International Conference on Computational Intelligence and Computing Research,pp. 749- 752, 2016.

[16]. B.Lakshmi,RasheedAhamed,Ravi Kishore Kodali "Assistive Sign Language Converter For Deaf And Dumb"International Conference on Internet of Things,Physical and Social Computing, pp. 302-307, 2019.

[17]. Archana S. Ghotkar, RuchaKhatal , Sanjana Khupase, Surbhi Asati& Mithila Hadap "Hand Gesture Recognition for Indian Sign Language" International Conference on Computer Communication and Informatics (ICCCI),pp. 877- 881, Jan-10-2012.

[18]. Krishna Sai, Sasikala T "Object Detection and count of object in images using Tensor flow Object Detection APT" Second International Conference on smart System and Inventive Technology (ICSSIT), PP. 542-546,2019