Sign Language Recognition using Deep Learning

Brinzel Rodrigues[1], Ankita Dodamani[2], Pranali Wadile[3], Amit Kuveskar[4]

[1] Assistant Professor, [2,3,4] Student, Bachelor of Engineering in Information Technology
Department of Information Technology , St. John College of Engineering and Management, Palghar
Maharashtra, India

--***---

Abstract: *According to the 2011 Census, In India, out of the total population of 121 crores, approximately 2.68 Crore humans are 'Disabled' (2.21% of the whole population)). Sign Language serves as a means for these people with special needs to communicate with others, but it is not a simple task. This barrier to communication has been addressed by researchers for years. The goal of this study is to demonstrate the MobileNets model's experimental performance on the TensorFlow platform when training the Sign language. Language Recognition Model, which can drastically reduce the amount of time it takes to learn a new language. Classification of Sign Language motions in terms of time and space Developing a portable solution for a real-time application. The Mobilenet V2 Model was trained for this purpose and an Accuracy of 70% was obtained.*

Keywords– Gesture Recognition, Deep Learning(DL), Sign Language Recognition(SLR), TensorFlow, Mobilenet V2.

1. INTRODUCTION

1.1 BACKGROUND

Since the beginning of Evolution, Humans have kept evolving and adapting to their available surroundings. Senses have developed to a major extent. But unfortunately, some people are born special. They are called special because they lack the ability to use all their five senses simultaneously. According to WHO, About 6.3% i.e. About 63 million people suffer from an auditory loss in India. Research is still going on in this context. According to Census 2011 statistics, India has a population of 26.8 million people who are differently-abled. This is roughly 2.21 percent in percentage terms. Out of the total disabled person, 69% reside in rural areas whereas 31% in urban areas.[1] There are various challenges faced by the specially-abled people for Health Facilities, Access to Education, Employment Facilities and the Discrimination/ Social Exclusions top it all. Sign Language is commonly used to communicate with deaf people.

1.2 SIGN LANGUAGE & ITS CONTRIBUTION.

Sign Language was discovered to be a helpful way of communication since it used hand gestures, facial emotions, and mild bodily movements to transmit the message. It is extremely important to understand and interpret the sign language and frame the meaningful sentence to convey the correct message which is extremely important and challenging at the same time. The purpose of this work is to contribute to the field of sign language recognition. Humans have been trying hard to adapt to these sign languages to communicate for a long time. Hand gestures are used to express any word or alphabet or some feeling while communicating.

Sign Language Recognition is a multidisciplinary subject on which research has been ongoing for the past two decades, utilising vision-based and sensor-based approaches. Although sensor-based systems provide data that is immediately usable, it is impossible to wear dedicated hardware devices all of the time. The input for vision-based hand gesture recognition could be a static or dynamic image, with the processed output being either a text description for speech impaired people or an audio response for vision-impaired people. In recent years, we have seen the involvement of machine learning techniques with the advent of Deep Learning techniques contributing as well. A dataset is an essential component of every machine learning program. We can't train a machine to produce accurate results without a good dataset. We created a dataset of Sign language images for our project. The photos were taken with a variety of backgrounds. After collecting all of the photos, they were cropped, converted to RGB channels, and labelled.

The benefit of this is that the image size and other supplementary data are minimised allowing us to process it with the fewest resources possible.

1.3 CONVOLUTION NEURAL NETWORK

The Convolution Neural Network (CNN) is a deep learning method inspired by human neurons. A neural network is a collection of artificial neurons known as nodes in technical terms. A neuron in simple terms is a

graphical representation of a numeric value. These neurons are connected using weights(numerical values).

Training refers to the process where a neural network learns the pattern required for performing the task such as classification, recognition, etc. When a neural network learns, the weight between neurons changes which results in a change in the strength of the connection as well. A typical neural network is made up of various levels. The first layer is called the input layer, while the output layer is the last. In our case of recognizing the image, this last layer consists of nodes that represent a different class. We have trained the model to recognize Alphabets A to Z & Numerals 0 to 9. The likelihood of the image being mapped to the class represented by the node is given by the output neuron's value. Generally, there are 4 layers in CNN Architecture: the convolutional layer, the pooling layer, the ReLU correction layer, and the fully-connected layer. The Convolutional Layer is CNN's first layer, and it works to detect a variety of features. Images are fed into the convolutional layer, which calculates the convolution of each image with each filter. The filters match the features we're looking for in the photographs to a match. A feature map is created for each pair (picture, filter). The pooling layer is the following tier. It takes a variety of feature maps as inputs and applies the pooling method to each of them individually. In simple terms, the pooling technique aids in image size reduction while maintaining critical attributes. The output has the same number of feature maps as the input, but they are smaller. It aids in increasing efficiency. and prevents over-learning. The ReLU correction layer is responsible for replacing any negative input values with zero.It serves as a mode of activation. The fully connected layer acts as the final layer. It returns a vector with the same size as the number of classes the image must be identified from.

Mobilenet is a CNN Architecture that is faster as well as a smaller model. It makes use of a Convolutional layer called depth-wise separable convolution.

2. LITERATURE REVIEW

In this section, we examine a few similar systems that have been explored and implemented by other researchers in order to have a better understanding of their methods and strategies.

Smart Glove For Deaf And Dumb Patient[3], The author's objective in this paper is to facilitate human beings by way of a glove-based communication interpreter system. Internally, The glove is attached to five flex sensors and is fastened. For each precise movement, the flex sensor generates a proportionate change in resistance. The Arduino uno Board is used to process these hand motions. It's a combination of a microcontroller and the LABVIEW software that's been improved. It compares the input signal to memory-stored specified voltage values.

According to this, a speaker is used to provide the appropriate sound.

Digital Text and Speech Synthesizer using Smart Glove for Deaf and Dumb[4], In the year 2017, the authors presented a system to increase the accuracy. an accelerometer was also incorporated which measured the orientation of the hand. It was pasted on the palm of the glove to determine the glove's orientation. The output voltage of the accelerometer altered with regard to the earth's orientation. Unlike the previous paper, this model had 5 outputs from flex sensors and 3 from the Accelerometer (the value of X, Y, Z-axis). Arduino is the controller used in this project. All of the flex sensor and accelerometer values are converted to gestures, and then code is built for all of them. In the hardware part, there is also a Bluetooth device. The received data is transmitted by the Bluetooth module across a wireless channel and received by a Bluetooth receiver in the smartphone. Using MIT App Inventor software, according to the user's needs, the author designed a text to speech program that receives all data and converts it to text or corresponding speech. Android applications proved to be efficient in use. but the weight of the gloves was still there.

Sign Language Recognition[5], In 2016, proposed a unique approach to assist persons with vocal and hearing difficulties in communicating. This study's authors discuss a new method for recognizing sign language and translating speech into signs. Using skin colour segmentation, the system developed was capable of retrieving sign images from video sequences with less crowded and dynamic histories. It can tell the difference between static and dynamic gestures and extract the appropriate feature vector. Support Vector Machines are used to categorise them. Experiments revealed satisfactory sign segmentation in a variety of backdrops, as well as fairly good accuracy in gesture and speech recognition.

Real-Time Recognition of Indian Sign Language[6], The authors have designed a system for identifying Indian sign language motions in this work. (ISL). The suggested method uses OpenCV's skin segmentation function to locate and monitor the Region of Interest (ROI). To train and predict hand gestures, fuzzy c-means clustering machine learning methods are utilised. The proposed system, according to the authors, can recognize real-time signs, making it particularly useful for hearing and speech-challenged individuals to communicate with normal people.

MobileNets for Flower Classification using TensorFlow[7], In this paper, the authors have experimented with the flower classification problem statement with Google's Mobilenet model Architecture. The authors have demonstrated a method for creating a

mobileNets application that is smaller and faster. The experimental results show that using the Mobilenets model on the Tensorflow platform to retrain the flower category datasets reduced the time and space required for flower classification significantly. but it compromised marginally with the accuracy when compared to Google's Inception V3 model.

Deep Learning for Sign Language Recognition on Custom Processed Static Gesture Images[8], The outcomes of retraining and testing this sign language are presented in this research. Using a convolutional neural network model to analyse the gestures dataset Inception v3 was used. The model is made up of several parts. Convolution filter inputs are processed on the same input. The accuracy of validation attained was better than 90%. This is a paper describing the multiple attempts at detecting sign language images using machine learning and depth data.

Gesture Recognition in Indian Sign Language Using Image Processing and Deep Learning[9], Microsoft Kinect RGBD camera was used to obtain the dataset. In this study, the authors proposed a real-time hand gesture recognition system based on the data acquired. They used computer vision techniques like 3D construction to map between depth and RGB pixels. The hand gestures were split from the noise when one-to-one mapping was achieved. The 36 static motions related to Indian Sign Language (ISL) alphabets and digits were trained using Convolutional Neural Networks (CNNs). Using 45,000 RGB photos and 45,000 depth images, the model obtained a training accuracy of 98.81 percent. The training of 1080 films resulted in a 99.08 percent accuracy. The model confirmed that the data was accurate in real-time.

Indian Sign Language Recognition[10]. This study outlines a framework for a human-computer interface that can recognize Indian sign language motions. This paper also suggests using neural networks for recognition. Furthermore, it is advocated that the number of fingertips and their distance from the hand's centroid be employed in conjunction with PCA for more robust and efficient results.

Signet: Indian Sign Language Recognition System based on Deep Learning [11], In this paper, The authors proposed a deep learning-based, signer independent model. The purpose behind this was to develop an Indian static Alphabet recognition system. It also reviewed the current sign language recognition techniques and implemented a CNN architecture from the binary silhouette of the signer hand region. They also go over the dataset in great depth, covering the CNN training and testing phases. The proposed method had a likelihood of success of 98.64 percent, which was higher than the majority of already available methods.

Deep Learning for Static Sign Language Recognition[12], Explicit skin-colour space thresholding, a skin-colour modelling technique, is used in this system. The skin-colour range that will be extracted is predefined (hand) made up of non–pixels (background). The photographs were fed into the system into the Convolutional Neural Network (CNN) model.CNN for image categorization. Keras was used for a variety of purposes. Images are being trained If you have the right lighting, you can do a lot of things. Provided a consistent background, the system was able to obtain an average. Testing accuracy was 93.67 percent, with 90.04 percent ascribed to human error. ASL alphabet recognition is 93.44 percent, while number recognition is 93.44 percent and 97.52 percent for static word recognition, outperforming the previous record for a number of other relevant research.

Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation[13], A Deep Learning-Based Approach to Recognizing Sign Language Gestures with Efficient Hand Gesture Representation. The authors' proposed approach combines local hand shape attributes with global body configuration variables to represent the hand gesture, which could be especially useful for complex organised sign language hand motions. In this study, the open pose framework was employed to recognize and estimate hand regions. A robust face identification method and the body parts ratios theory were utilised to estimate and normalise gesture space. Two 3DCNN instances were used to learn the fine-grained properties of the hand shape and the coarse-grained features of the overall body configuration. To aggregate and globalise the extracted local features, MLP and autoencoders were used, as well as the SoftMax function. Common Garbage Classification Using MobileNet[14]

Sign Language Recognition System using TensorFlow Object Detection API[15], The authors of this paper investigated a real-time method for detecting sign language. Images were captured using a webcam (Python and OpenCV were used) for data acquisition, lowering the cost. The evolved system has a confidence percentage of 85.45 percent on average.

Sign Language Recognition system[16], Here, the system consists of a webcam that captures a real-time image of the hand, a system that processes and recognizes the sign, and a speaker that outputs sounds.

3. METHODOLOGY

MobileNets for TensorFlow are a series of mobile-first computer vision models that are designed to maximise accuracy while taking into account the limited resources available for an on-device or embedded application[19].

Model	ImageNet Accuracy	Million Mult-Adds	Million Parameters
1.0 MobileNet-224	70.6%	569	4.2
GoogleNet	69.8%	1550	6.8
VGG 16	71.5%	15300	138

Figure 1. MobileNet parameter and accuracy comparison against GoogleNet and VGG 16 [2]

As seen in the above table, it can be concluded that Mobile net gives fairly similar results as compared with Google Net Model and VGG 16, but the number of Parameters required for the purpose is significantly less, which makes it ideal to use. The main difference between the 2D convolutions in CNN and Depthwise convolutions is that the 2D Convolutions are performed over multiple channels, whereas in Depthwise convolutions each channel is kept separate.[2]

The first layer of the MobileNet is a full convolution, while all following layers are Depthwise Separable Convolutional layers. All the layers are followed by batch normalisation and ReLU activations. The final classification layer has a softmax activation. In terms of our project's scope, our major goal is to create a model that can recognize the numerous signs that define Letters, Numerals, and Gestures using mobilenets. Using the Object Detection Technique, the Trained model can recognize the indicators in real-time. The idea behind this project is to develop an application that is handy and can detect the hand gestures (signs) and recognize what the specially-abled person is trying to speak with a motive to help to ease the efforts required by the specially-abled people to communicate with other Normal People.
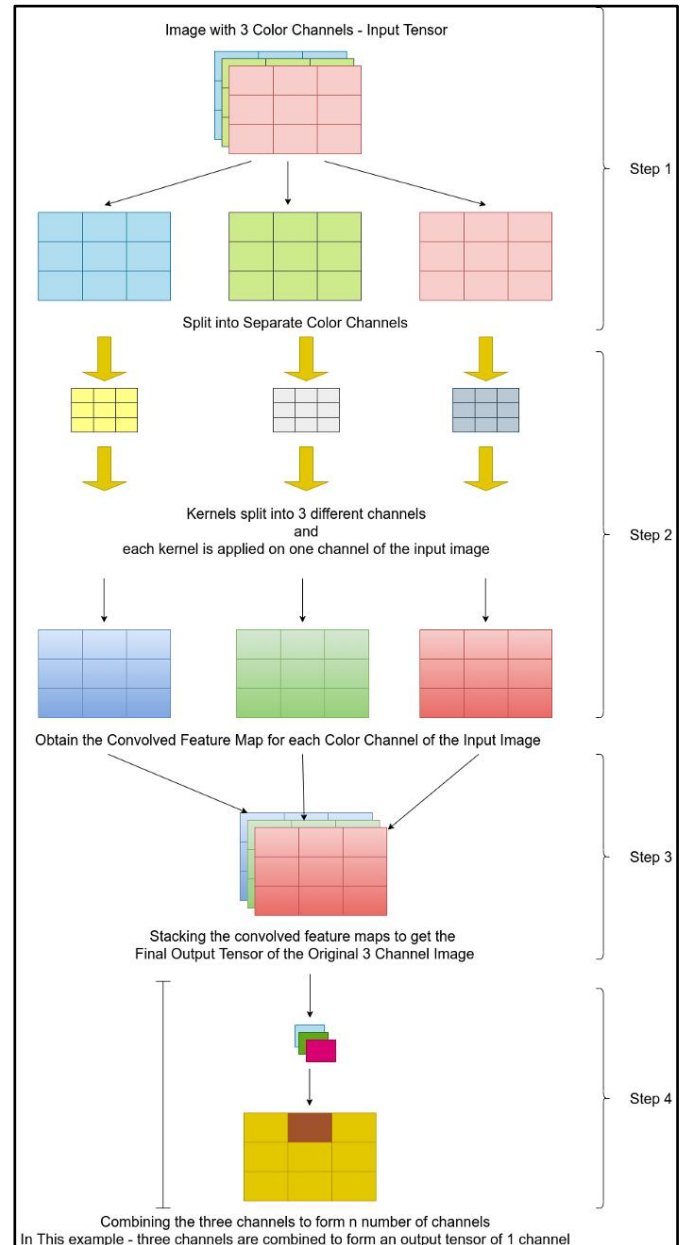
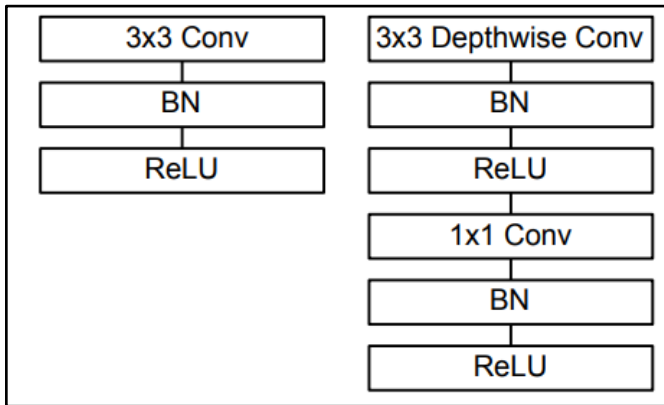**Figure 2.** Diagramatic explanation of Depth Wise Separable Convolutions[2]

Figure 3. Left: Standard Convolutional layer, Right: Depthwise Separable Convolutional layers in MobileNet[2]
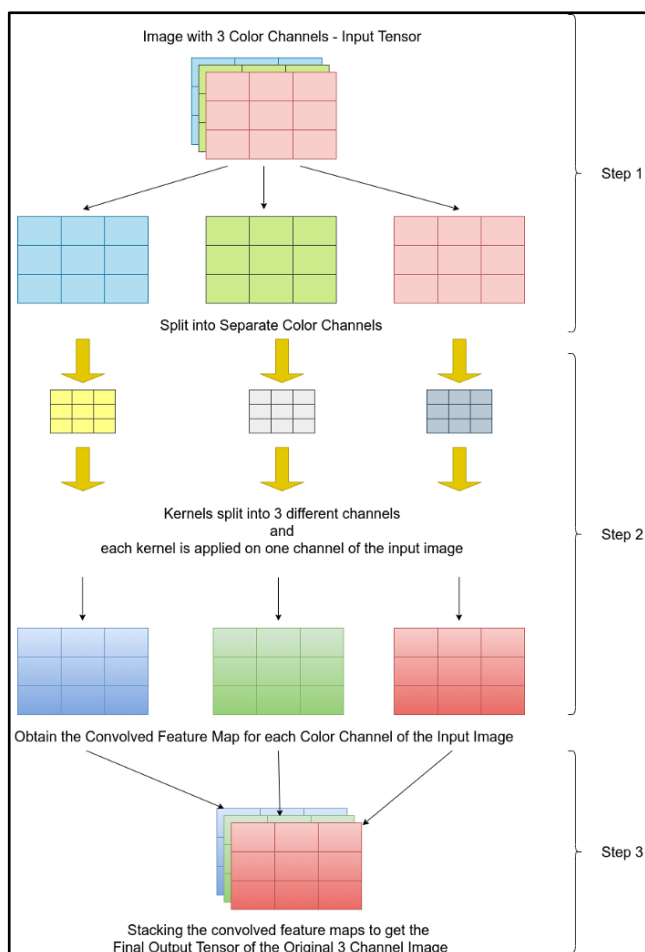

Figure 4. Diagramatic explanation of Depthwise Convolutions[2]

4. CONSTRUCTION OF MODEL

The experimental setup for the Sign Language Recognition model utilising MobileNet on the TensorFlow framework is covered in this section. The categorization model is divided into the four stages below: Phases include image preprocessing, training, verification, and testing.

For our project we have made our data set of around 15,000 images which consist of 26 Alphabets(A to Z), Numerals (0-9).After collecting all the images we labelled them using LabelImg[18] The Images collected are then divided into two categories: Train and Test. Train dataset is used to train the model and the Verification Phase uses the Test dataset to verify the accuracy. Then the Model is used to test the model in Real-Time

5. EXPERIMENTAL EVALUATION

5.1. DATASET AND EXPERIMENTAL SETUP

The dataset is generated for Indian Sign Language, whose signs are English alphabets and integers. A dataset of approximately 15000 photos has been developed.

A Windows 10 PC with an Intel i5 7th generation 2.70 GHz processor, 8 GB of RAM, and a webcam was used for the test (HP TrueVision HD camera with 0.31 MP and 640x480 resolution). Python (version 3.8.9), Jupyter Notebook, OpenCV, and TensorFlow Object Detection API are all part of the development environment.

5.2. RESULTS AND DISCUSSION

In real-time, the created system can detect Indian Sign Language alphabets and digits. TensorFlow object detection API was used to build the system. The TensorFlow model that has already been pre-trained SSD MobileNet v2 640x640 is the model zoo. It has undergone training. Using transfer learning on the newly created dataset. There are 15000 photos in all, one for each letter of the alphabet.
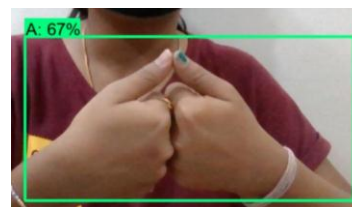

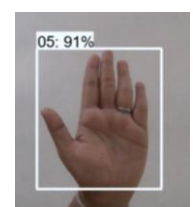*Figure 5:Recognizing Alphabet A with 67% 05 Accuracy*

Figure   6:Recognizing   Number with 67% Accuracy

Figure 7: *Recognizing Alphabet G with 79% Accuracy*

Figure 8: *Recognizing Number 05 with 72% Accuracy*

The Overall Accuracy of the Model has turned out to be 70%.

6. CONCLUSION

A technique for recognizing Indian Sign Language is presented in this work. The Tensorflow Mobilenet V2 Model was used to recognize static indicators successfully. The Model can further be improved by adding more numbers of signs and increasing the dynamicity of the Images.

REFERENCES

1. Persons with Disabilities (Divyangjan) in India. New Delhi, India: Ministry of Statistics and Programme Implementation, Government of India, 2021.

2. Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications ",Computer Vision and Pattern Recognition, Cornell University.

3. P.B.Patel, Suchita Dhuppe, Vaishnavi Dhaye "Smart Glove For Deaf And Dumb Patient " International Journal of Advance Research in Science and Engineering, Volume No.07, Special Issue No.03, April 2018

4. Khushboo Kashyap, Amit Saxena, Harmeet Kaur, Abhishek Tandon, Keshav Mehrotra "Digital Text and Speech Synthesizer using Smart Glove for Deaf and Dumb" International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 6, Issue 5, May 2017

5. Anup Kumar, Karun Thankachan and Mevin M. Dominic "Sign Language Recognition" 3rd InCI Conf. on Recent Advances in Information Technology I RAIT-20161

6. Muthu Mariappan H, Dr Gomathi V "Real-Time Recognition of Indian Sign Language" Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019)

7. Nitin R. Gavai,Yashashree A. Jakhade,Seema A. Tribhuvan, Rashmi Bhattad "MobileNets for Flower Classification using TensorFlow" 2017 International Conference on Big Data, IoT and Data Science (BID) Vishwakarma Institute of Technology, Pune, Dec 20-22, 2017

8. Aditya Das, Shantanu Gawde, Khyati Suratwala and Dr. Dhananjay Kalbande "Sign Language Recognition Using Deep Learning on Custom Processed Static Gesture Images" Department of Computer Engineering Sardar Patel Institute of Technology Mumbai, India

9. Neel Kamal Bhagat, Vishnusai Y,Rathna G N "Indian Sign Language Gesture Recognition using Image Processing and Deep Learning" Department of Electrical Engineering Indian Institute of Science Bengaluru, Karnataka ©2019 IEEE

10. Divya Deora, Nikesh Bajaj "Indian Sign Language Recognition" 2012 1st International Conference on Emerging Technology Trends in Electronics, Communication and Networking ©2012 IEEE

11. Sruthi C. J and Lijiya A "Signet: A Deep Learning based Indian Sign Language Recognition System" International Conference on Communication and Signal Processing, April 4-6, 2019, India

12. Lean Karlo S. Tolentino, Ronnie O. Serfa Juan, August C. Thio-ac, Maria Abigail B. Pamahoy, Joni Rose R. Forteza, and Xavier Jet O. Garcia "Static Sign Language Recognition Using Deep Learning" International Journal of Machine Learning and Computing, Vol. 9, No. 6, December 2019

13. Muneer Al-Hammadi (Member, Ieee), Ghulam Muhammad (Senior Member, Ieee), Wadood Abdul(Member, Ieee), Mansour Alsulaiman Mohammed A. Bencheriftareq S. Alrayes, Hassan Mathkourand Mohamed Amine Mekhtiche, "Deep Learning-Based Approach for Sign Language Gesture Recognition With Efficient Hand Gesture Representation", IEEE Access Version-November 2, 2020.

14. Stephenn L. Rabano, Melvin K. Cabatuan, Edwin Sybingco, Elmer P. Dadios, Edwin J. Calilung, "Common Garbage Classification Using MobileNet" IEEE Xplore: 14 March 2019

15. Sharvani Srivastava, Amisha Gangwar, Richa Mishra, Sudhakar Singh, "Sign Language Recognition System using TensorFlow Object Detection API" International Conference on Advanced Network Technologies and Intelligent Computing (ANTIC-2021), part of the book series 'Communications in Computer and Information Science (CCIS)', Springer.

16. Priyanka C Pankajakshan,Thilagavati B, "Sign Language Recognition system" IEEE Sponsored 2nd International Conference on Innovations in Information Embedded and Communication Systems, ICIIECS'15

17. Mohammed Safeel, Tejas Sukumar,Shashank K S, Arman M D, Shashidhar R, Puneeth S B, "Sign Language Recognition Techniques- A Review", 2020 IEEE International Conference for Innovation in Technology (INOCON) Bengaluru, India. Nov 6-8, 2020

18. LabelImg - A tool used for Image Annotation of dataset - https://github.com/tzutalin/labelImg.git

19. Google, "MobileNets: Open-Source Models for Efficient On-Device Vision," Research Blog. [Online]. Available:https://research.googleblog.com/2017/06/mobilenets-open-source-models-for.html.

20. https://innovate.mygov.in/

21. https://towardsdatascience.com/sign-language-recognition-using-deep-learning-6549268c60bd

22. https://www.tensorflow.org/api_docs/python/tf/keras/applications/mobilenet_v2/MobileNetV2