Real Time Sign Language Detection

Sneha Santoshkumar[1], Riya Divakaran[2], Shruti Krishnakumar[3], Sukrishna Nair[4], Prof. Shweta Patil[5]

[1,2,3,4]UG Student, Dept. of Computer Engineering, Pillai College of Engineering, New Panvel, India.
[5]Assistant Professor, Dept. of Information Technology, Pillai College of Engineering, New Panvel, India.

Abstract - *The ML based Sign Language Detection system aims at communicating with differently abled people without the help of any expensive human interpreter. This model translates the signs gestures captured into text so that the user can simply read and know what the person is trying to convey irrespective of whether the user has knowledge about sign language or not. In this project, a real-time ML-based system is built using images captured with the help of webcam for sign language detection. The main purpose of this project is to design a system for people with different abilities so that they can easily communicate with other people. The existing digital translator is very slow, because each letter must be gestured with, and it takes a long time to form a simple sentence. Model will be built with the help of Label-Img software and TensorFlow Object Detection API, using real colouring images. And detect sign language in real time using OpenCV.*

Key Words: **Deep Learning, SSD ML algorithm, LabelImg software, real time, TensorFlow object detection module.**

1.INTRODUCTION

Very few people know how to communicate using sign language as it is not a mandatory language, making it difficult for people with disabilities to communicate with other people.

The most common means of communicating with them is with the help of human interpreters, which is again very expensive and not many can afford it. There are many different sign languages in the world. There are approximately 200 sign languages in the world including Chinese, Spanish, Irish, American Sign Language, and Indian Sign Language, which are the most commonly used sign languages.

The ML-based sign language recognition system is designed to communicate with people with disabilities without the help of expensive human interpreters. This model translates the captured characters or gestures into text so that the user can easily read and know what the person is trying to convey, whether or not the user has knowledge of sign language.

2. LITERATURE SURVEY

A. Principal component analysis:

Cristian Amaya and Victor Murray [1] use PCA for feature extraction in hand regions and classification using SVM. Hand segmentation is performed using a skin probability model. Next, morphological operations and filters are used to enhance the segmented hand.

B. Fuzzy c-means clustering machine learning algorithm:

Dr. Gomathi V [2] trained and predicted hand gestures by applying fuzzy c-means clustering machine learning algorithms. In fuzzy clustering, the things may belong to over one cluster. Among several fuzzy clustering algorithms, fuzzy c-means clustering (FCM) algorithm is employed most generally, and this may be used for both supervised learning and unsupervised learning, depending upon the requirements. The proposed system is employed to acknowledge the real-time signs.

C. Neural Networks:

Balbin et al. [3],developed a system that recognized five Filipino words and used colored gloves for hand position recognition. The system was developed using a neural network toolbox and graphical user interface in MATLAB. Networks are divided into two types: Supervised and Unsupervised. Self-organising Map or SOM is a field of Neural Networks which learns to detect regularities and correlations in the input.

D. Wrist-worn Motion and Surface EMG Sensors:

Jian Wu, Zhongjun Tian, Lu Sun, Leonardo Estevez and Roozbeh Jafari [4] developed a real-time American SLR system leveraging fusion of surface electromyography (sEMG) and a wrist-worn inertial sensor at the feature level. A feature selection is provided for 40 most typically used words and for four subjects.sEMG could be a non-invasive technique to measure the electrical potential of muscle activities.Results show that after feature selection and conditioning achieves decent recognition rate

2.1 SUMMARY OF RELATED WORK

The summary of methods used in literature is given in Table 1.

Table -1: Summary of literature survey

Literature	**Advantages**	**Disadvantages**
Cristian Amaya et al. 2020 [1]	The algorithm shows correct predictions in addition to 82% of evaluated images.	It is applicable to just some letters The accuracy achieved isn't up to 90%.
Dr. Gomathi V et al. 2019 [2]	This FCM based real-time language recognition system, has produced 75 % accuracy ingesture labelling.	It requires more computation time than the others.
Balbin et al. 2016 [3]	The results show that the system can do 97.6% of the recognition rate for five persons.	The system only recognized five Filipino words and used colored gloves for hand position recognition.
Jian Wu et al. 2015 [4]	The results show that after feature selection and adjustment, the system achieves a recognition rate of 95.94%.	The images are captured using electromyograp hy which is costly, because it requires large-size datasets with diverse sign motion.

3. Proposed Work

To extract features from the required image Deep Learning SSD ML algorithm is used. For the detection, TensorFlow Object Detection API is used where the extracted features from the pictures taken are passed onto the TensorFlow module which goes to create comparisons with the real time video present within the frame. On detection of any of those features it's visiting, generate a bounding box round the gesture and make the prediction. The prediction goes to be identical because of the label of the image. we are going to be ready to detect linguistic communication in real time using OpenCV

3.1 SYSTEM ARCHITECTURE

The system architecture is given in Figure 1. Each block is described in this Section.

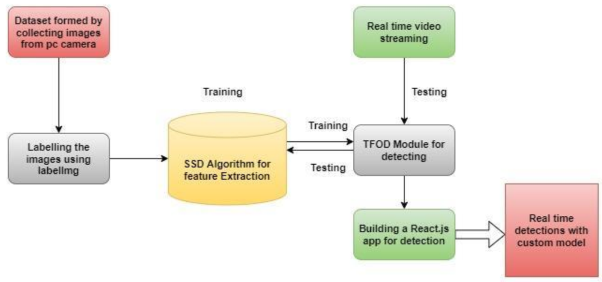

Fig.-1: Proposed system architecture

A. Dataset Creation:

The LabelImg software is used for graphically labelling the images that is further used when recognizing the images. We have to keep in mind that labelling has to be done correctly i.e, the gesture should be labelled with a right label so that we get the gestures recognized correctly later with the right label. Once the images are labelled and saved an XML file is created for that image. This XML file contains the information about where the model should be looking in the image during the training process. This model is trained for 5 different gestures hence 5 different labels were used for labelling them. For each gesture, 15 images were used and clicked from different angles. Code used to automatically take pictures and save them to a specific folder. Labelling is done by drawing a frame around the gesture being performed.. XML file associated with a tagged image indicating where the model should look for the gesture when training the ML model.

B. Training and Testing:

Out of 15 images collected with generated image's XML files for each image, 2 files are used for testing and The remaining 13 are used to train models. Model ML was trained using Deep Learning SSD ML algorithm and tested using TensorFlow Object Detection API.

C. SSD Algorithm:

A Deep Learning SSD ML algorithm used to extract features from the specified image. SSD (Single Shot Detection) algorithm is intended for object detection in real-time. Faster R-CNN uses an object proposal network to form boundary boxes and utilises those boxes to classify objects. The SSD architecture model is a single convolution network that learns to predict bounding box locations and classify these locations in one pass. Hence, SSD are often trained end-to-end.

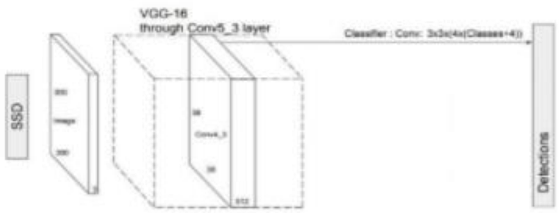

Fig.-2: SSD Architecture

D. Tensorflow Object Detection API:

TensorFlow is an open source library for large-scale numerical computation and machine learning, supporting Google Brain TensorFlow, data collection strategies, training models, serving prediction and refining future results. The TensorFlow Object Detection API is an open source framework based on TensorFlow that makes it easy to build, train, and deploy object detection models. There are already pre-trained models in their framework called Model Zoo. It consists of a collection of pre-trained models trained on different data sets like the Common Objects in Context (COCO) dataset, the KITTI dataset and also the Open Images dataset. The TensorFlow Object Detection API is a framework for building deep learning networks that solve object detection problems.

E. Real Time Sign Detection Application:

Converting a Tensorflow Object Detection API model to Tensorflow.JS Graph Model format Hosting a trained Tensorflow deep learning model for applications. Downloading the React and Tensorflow.JS Computer Vision Template. Making real time detections using a deployed Tensorflow.JS model. Visualising detections within the HTML canvas.

3.2 REQUIREMENT ANALYSIS

The implementation detail is given in this section.

3.2.1 Software

Table- 2 :Software details

Operating System	Windows 10
Programming Language	Python

3.2.2 Hardware

Table-3: Hardware details

Processor	2 GHz Intel
HDD	180 GB
RAM	2 GB

3.3 DATASET

The LabelImg software is used for graphically labelling the pictures that's further used when recognizing the photographs. The gesture should be labelled with a right label in order that we get the gestures recognized correctly later with the proper label. We have five different labels which are going to represent the different sign language poses. We have collected 15 different images of 5 different hand signs. Once the pictures are labelled and saved an XML file is formed for that image. This XML file contains the knowledge about where the model should be looking within the image during the training process. Among which 13 are for training and 2 of those per class for testing.

4. CONCLUSION

Real-time sign language detection with SSD algorithms using true colour images from PC cameras has been introduced. In this article, signs are converted into text sentences to help people with disabilities communicate easily with others. This system has shown good results in taking advantage of depth learning techniques.

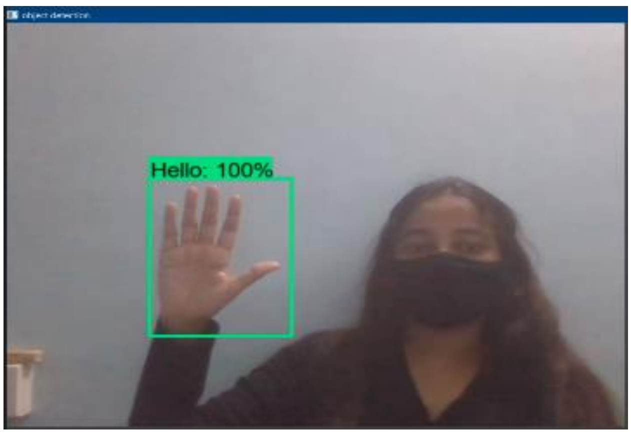

Fig. - 3: Gesture recognition for Hello

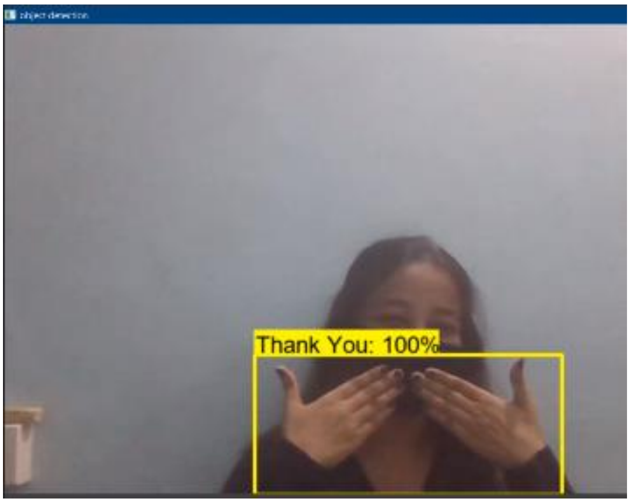

Fig.- 4: Gesture recognition for Thank You

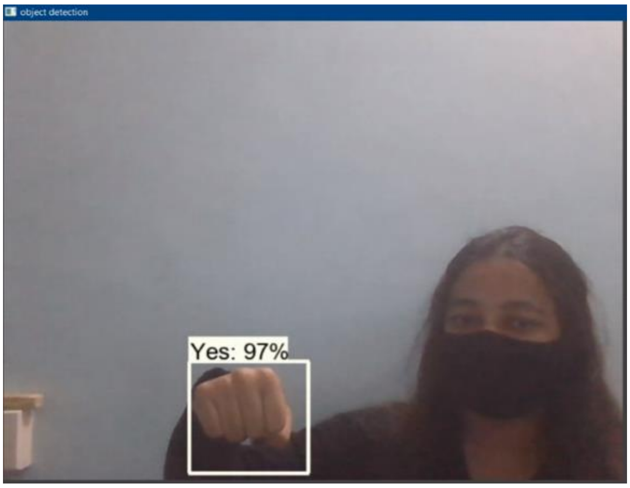

Fig. - 5: Gesture recognition for Yes

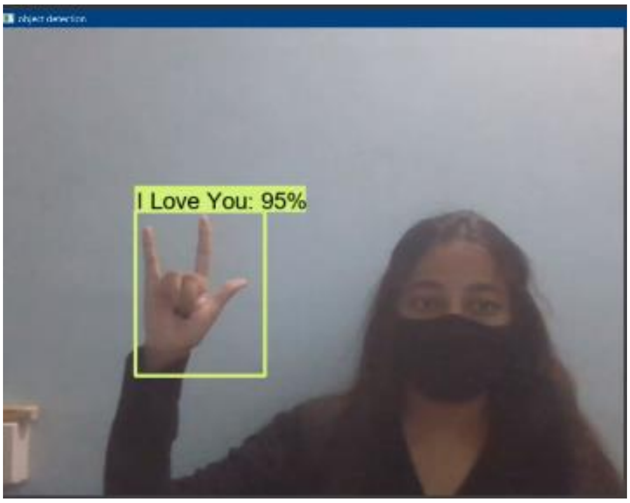

Fig. - 6: Gesture recognition for I Love You

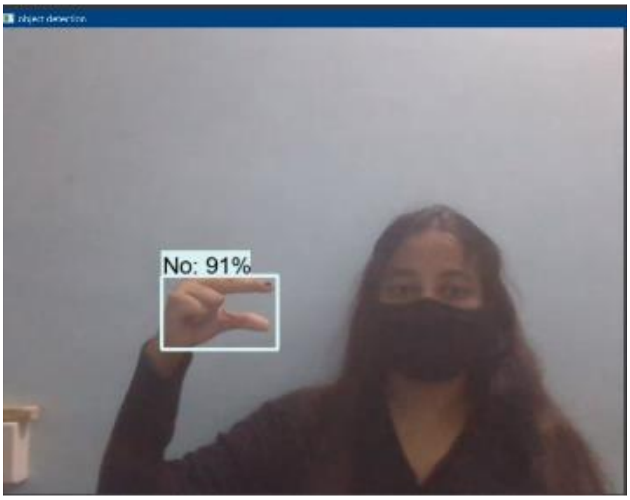

Fig. -7 : Gesture recognition for No

ACKNOWLEDGEMENT

It is our privilege to express our sincerest regards to our supervisor Professor Shweta Patil for the valuable inputs, able guidance, encouragement, whole-hearted cooperation and constructive criticism throughout the duration of this work. We deeply express our sincere thanks to our Head of the Department Dr. Sharvari Govilkar and our Principal Dr. Sandeep M. Joshi for encouraging and allowing us to present this work.

REFERENCES

[1] Cristian Amaya and Victor Murray, “Real-Time Sign Language Recognition” , 2020 IEEE

[2] Dr. Gomathi V, “Real-Time Recognition of Indian Sign Language”, 2019 IEEE Second International Conference on Computational Intelligence in Data Science (ICCIDS-2019)

[3] Jessie R, Balbin, Dionis A. Padilla, Felicito S. Caluyo, Janette C. Fausto, Carlos C. Hortinela IV, Cyrel O. Manlises, Christine Kate S. Bernardino, Ezra G. Finones, Lanuelle T. Ventura, “Sign Language Word Translator Using Neural Networks for the Aurally Impaired as a Tool for Communication”, 6th IEEE International Conference on Control System, Computing and Engineering(ICCSCE), 2016

[4] Jian Wu, Zhongjun Tian, Lu Sun, Leonardo Estevez and Roozbeh Jafari, “Real-time American Sign Language Recognition Using Wrist-worn Motion and Surface EMG Sensors”, 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks

[5] Michael Van den Bergh ETH Zurich, Luc Van Gool, KU Leuven, “Combining RGB and ToF Cameras for Real-time 3D Hand Gesture Interaction”, 2011 IEEE Workshop on Applications of Computer Vision

BIOGRAPHIES

Sneha Santoshkumar

Riya Divakaran

Shruti Krishnakumar

Sukrishna Nair