*As per UGC guidelines an electronic bar code is provided to seure your paper*

*International Journal for Modern Trends in Science and Technology*, 8(01): 32-37, 2022

ISSN: 2455-3778 online
DOI: https://doi.org/10.46501/IJMTST0801006
Available online at: http://www.ijmtst.com/vol8issue01.html

# Real Time Sign Language Detection

**Aman Pathak | Avinash Kumar | Priyam | Priyanshu Gupta | Gunjan Chugh**

Department of Information Technology, Dr. Akhilesh Das Gupta Institute of Technology and Management, New Delhi, India.

**To Cite this Article**
Aman Pathak, Avinash Kumar, Priyam, Priyanshu Gupta and Gunjan Chugh. Real Time Sign Language Detection. *International Journal for Modern Trends in Science and Technology* 2022, *8* pp. 32-37.
https://doi.org/10.46501/IJMTST0801006

**Article Info**
Received: 24 November 2021; Accepted: 21 December 2021; Published: 31 December 2021

## ABSTRACT

***A real time sign language detector is a significant step forward in improving communication between the deaf and the general population. We are pleased to showcase the creation and implementation of sign language recognition model based on a Convolutional Neural Network(CNN).We utilized a Pre-Trained SSD Mobile net V2 architecture trained on our own dataset in order to apply Transfer learning to the task. We developed a robust model that consistently classifies Sign language in majority of cases. Additionally, this strategy will be extremely beneficial to sign language learners in terms of practising sign language. Various human-computer interface methodologies for posture recognition were explored and assessed during the project. A series of image processing techniques with Human movement classification was identified as the best approach. The system is able to recognize selected Sign Language signs with the accuracy of 70-80% without a controlled background with small light.***

***KEYWORDS: CNN, Pre-Trained SSD Mobile net V2, Sign Language***

## 1.INTRODUCTION

Sign language is largely used by the disabled, and there are few others who understand it, such as relatives, activists, and teachers at SekolahLuarBiasa (SLB). Natural gestures and formal cues are the two types of sign language[1]. The natural cue is a manual (hand-handed) expression agreed upon by the user (conventionally), recognised to be limited in a particular group (esoteric), and a substitute for words used by a deaf person (as opposed to body language). A formal gesture is a cue that is established deliberately and has the same
language structure as the community's spoken language.[2]

More than 360 million of world population suffers from hearing and speech impairments [3]. Sign language detection is a project implementation for designing a model in which web camera is used for capturing images of hand gestures which is done by open cv. After capturing images, labelling of images are required and then pre trained model SSD Mobile net v2 is used for sign recognition. Thus, an effective path of communication can be developed between deaf and normal audience. Three steps must be completed in real time to solve our problem:

1. Obtaining footage of the user signing is step one (input).
2. Classifying each frame in the video to a sign.
3. Reconstructing and displaying the most likely Sign from classification scores (output).

**This topic poses a big difficulty in terms of computer vision because of a variety of factors, including:**

- Environmental disturbance (e.g., lighting sensitivity, background, and camera position)
- Closure (e.g., some fingers, or an entire hand can be out of the field of view)
- Sign boundary detection (when a sign ends and the next begins).

This model uses pipeline that takes input through a web camera from a user who is signing a gesture and then by extracting different frames of video, it generates sign language possibility for each gesture.

## 2. RELATED WORK

With the continuous development in Information technology the ways of interaction between computers and Humans have also evolved.

There has been a lot of work done in this field to help deaf and able-bodied people communicate more effectively.

Because sign language is a collection of gestures and postures, any effort to recognise sign language falls under the purview of human computer interaction.[4] Sign Language Detection is categorised in two parts.

The first category is the Data Glove approach, in which the user wears a glove with electromechanical devices attached to digitise hand and finger motion into processable data.

The disadvantage of this method is that you must always wear extra gear and the results are less accurate.

In contrast, the second category, computer-vision-based approaches, require only a camera, allowing for natural interaction between humans and computers without the use of any additional devices.

Apart from various developments in ASL field, Indian people started putting work in ISL. Like Image key pointdetection using SIFT, and then comparing the key point of a new image to the key points of standard images per alphabet in a database to classify the new image with the label of the closest match.[5]

Similarly various work has been put into recognising the edges efficiently one[6] of the idea was to use a combination of the colour data with bilateral filtering in the depth images to rectify edges.

With Advancement in Deep Learning and neural networks people also implementing them in improving the detection system. In reference [7], the ASL is recognised using a variety of feature extraction and machine learning techniques, including the Histogram technique, the Hough transform, OTSU's segmentation algorithm, and a neural network.

Image processing is concerned with computer processing the images which include collecting, processing, analysing and understanding the results obtained. Computer vision necessitates a combination of low-level image processing to improve image quality (e.g., removing noise and increasing contrast) and higher-level pattern recognition and image understanding to recognise features in the image.

## 3. REVIEW OF HAND GESTURE AND SIGN LANGUAGE RECOGNITION TECHNIQUES:

Methods like identifying hand motion trajectories for distinct signs and segmenting hands from the background to forecast and string them into sentences that are both semantically correct and meaningful are used in sign language recognition. Furthermore, motion modelling, motion analysis, pattern identification, and machine learning are all issues in gesture recognition. Handcrafted parameters or parameters that are not manually set are used in SLR models. The model's ability to do the categorization is influenced by the model's background and environment, such as the illumination in the room and the pace of the motions. Due to changes in views, the gesture seems distinct in 2D space.

There are several ways for recognising gestures which includes sensor-based and vision-based systems. Sensor-equipped devices capture numerous parameters such as the trajectory, location, and velocity of the hand in the sensor-based approach. On the other hand,vision-based approaches are those in which images of video footages of the hand gestures are used.[8] The steps followed for achieving the sign language recognition are:

- The Camera used in the sign language recognition system:The proposed sign language recognition system is based on frame captured by a web camera

on a laptop or PC.Using the OpenCV Python computer library image processing is done.

- Capturing Images: Multiple images of different sign language symbols were taken from various angles and varying light conditions in order to achieve better accuracy through a large dataset.
- Segmentation: As the capturing part is done, further a particular region is selected from the entire image which has the sign language symbol that is to be predicted. Bounding boxes are enclosed for the sign to be detected. These boxes should be tight around the region which is to be detected from the image. Specific names were given to the hand gestures which were labelled.LabelImg tool was used for the labelling part.
- Selection of images for the training and testing purpose
- Creating TF Records: Record files were created from the multiple training and testing images.
- Classification:Machine learning approaches can be classified as supervised or unsupervised. Supervised machine learning is a technique for teaching a system to detect patterns in incoming data so that it can predict future data. Supervised machine learning uses a collection of known training data and applies it to labelled training data to infer a function.[8]

## 4.DESIGN AND IMPLEMENTATION:

- **Dataset:** For this project, a user defined dataset is used. It is a collection of over 2000 images, around 400 for each of its classes.This dataset contains a total of 5 symbols i.e.,*Hello, Yes, No, I Love You* and *Thank You*,which is quite useful while dealing with the real time application.

Fig. 1 Sign Symbols

## 5. ALGORITHM USED:

- Convolutional Neural Network:

A Convolutional Neural Network (ConvNet/CNN) is a Deep Learning system that can take an input picture and assign importance (learnable weights and biases) to various aspects/objects in the image, as well as differentiate between them. The amount of pre-processing required by a ConvNet is much less than that required by other classification techniques. ConvNets can learn these filters/characteristics with adequate training, whereas simple techniques need hand-engineering of filters.[9]

ConvNets are multilayer artificial neural networks designed to handle 2D or 3D data as input. Every layer in the network is made up of several planes that may be 2D or 3D, and each plane is made up of numerous independent neurons composition, where nearby layer neurons are linked but same layer neurons are not.[9]

A ConvNet can capture the Spatial and Temporal aspects of an image by applying appropriate filters. Furthermore, reducing the number of parameters involved and reusing weights resulted in the architecture performing better fitting to the picture collection. ConvNet's major goal is to make image processing easier by extracting relevant characteristics from images while preserving crucial information that is must for making accurate predictions.This is highly useful for developing an architecture that is not just capable of collecting and learning characteristics but also capable of handling massive volumes of data.[9]

- Overall Architecture:

CNNs are made up of three different sorts of layers. There are three types of layers: convolutional layers, pooling layers, and fully-connected layers. A CNN architecture is generated when these layers are layered. Figure 1 depicts a simple CNN architecture for MNIST classification.[10]

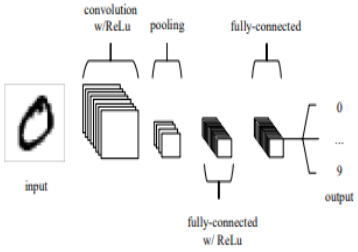



Fig. 2 Simple CNN architecture

## 6.TOOLS USED:

- **TensorFlow:**It is an open-sourceartificial intelligence package that builds models using data flow graphs. It enables developers to build large-scale neural networks with several layers. TensorFlow is mostly used for classification, perception, comprehension, discovery, prediction, and creation.[11]
- **Object Detection API:**It is an open source TensorFlow API to locate objects in an image and identify it.
- **Open CV:**OpenCV is an open-source, highly optimised Python library targeted at tackling computer vision issues. It is primarily focused on real-time applications that provide computational efficiency for managing massive volumes of data. [12] It processes photos and movies to recognise items, people, and even human handwriting
- **LabelImg:**LabelImg is a graphical image annotation tool that labels the bounding boxes of objects in pictures.[13]


Fig. 3 Label Image[13]

## 7.MODEL ANALYSIS AND RESULT:

The model was trained using the technique of transfer learning and a pre-trained model SSDmobile net v2 was used.

- Transfer Learning:

Transfer learning is a concept that describes a process in which a model that has been trained on one problem is applied in some way to a second, related problem. Transfer learning is a deep learning technique that includes training a neural network model on an issue that is similar to the one being addressed before applying it to the problem at hand. Using one or more layers from the learnt model, a new model is then trained on the problem of interest.14]

- SSD Mobile net V2:

The Mobile Net SSD model is a single-shot multibox detection (SSD) network that scans the pixels of an image that are inside the bounding box coordinates and class probabilities to conduct object detection. In contrast to standard residual models, the model's architecture is built on the notion of inverted residual structure, in which the residual block's input and output are narrow bottleneck layers. In addition, nonlinearities in intermediate layers are reduced, and lightweight depthwise convolution is applied. The TensorFlow object detection API includes this model.[15]

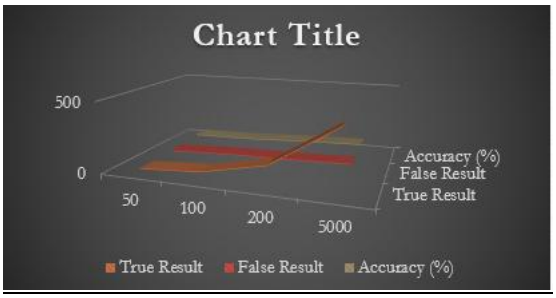
- Result:

Table-1 analysis

| Images used to train | True Result | False Result | Accuracy (%) |
|---|---|---|---|
| 50 | 23 | 27 | 46 |
| 100 | 52 | 48 | 52 |
| 200 | 145 | 55 | 72.5 |
| 500 | 432 | 68 | 86.4 |

Table-2 Sign Recognition

| Gesture Name | Accuracy (%) |
|---|---|
| Yes | 88.7 |
| No | 88.6 |
| Thank You | 84.1 |
| Hello | 91.0 |
| I Love You | 82.4 |




Graph-1 Accuracy/Images


Fig-4 Real time Sign Detection

## 8.APPLICATION AND FUTURE SCOPE:

- Application:

- The dataset can easily be extended and customized according to the need of the user and can prove to be an important step towards reducing the gap of communication for dumb and deaf people.
- Using the sign detection model, meetings held at a global level can become easy for the disabled people to understand and the value of their hard work can be given.
- The model can be used by any person with a basic knowledge of tech and thus available for everyone.
- This model can be implemented at elementary school level so that kids at a very young age can get to know about the sign language.

- Future scope:
  - The implementation of our model for other sign languages such as Indian sign language or American sign language.
  - Further training the neural network to efficiently recognise symbols.
  - Enhancement of model to recognise expressions.

## 9.CONCLUSION:

The main purpose of sign language detection system is providing a feasible way of communication between a normal and dumb people by using hand gesture. The proposed system can be accessed by using webcam or any in-built camera that detects the signs and processes them for recognition.

From the result of the model, we can conclude that the proposed system can give accurate results under controlled light and intensity. Furthermore, custom gestures can easily be added and more the images taken at different angle and frame will provide more accuracy to the model. Thus, the model can easily be extended on a large scale by increasing the dataset.

The model has some limitation such as environmental factors like low light intensity and uncontrolled background which cause decrease in the accuracy of the detection. Therefore, we'll work next to overcome these flaws and also increase the dataset for more accurate results.

## REFERENCES

[1] Martin D S 2003 Cognition, Education, and Deafness: Directions for Research and Instruction (Washington: Gallaudet University Press)

[2] McInnes J M and Treffry J A 1993 Deaf-blind Infants and Children: A Developmental Guid (Toronto : University of Toronto Press)

[3] http://www.who.int/mediacentre/factsheets/fs300/en/

[4] Harshith.C, Karthik.R.Shastry, Manoj Ravindran, M.V.V.N.S Srikanth, Naveen Lakshmikhanth, "Survey on various gesture recognition Techniques for interfacing machines based on ambient intelligence", International Journal of Computer Science & Engineering Survey (IJCSES) Vol.1, No.2, (November 2010)

[5] SAKSHI GOYAL, ISHITA SHARMA, S. S. Sign language recognition system for deaf and dumb people. International Journal of Engineering Research Technology 2, 4 (April 2013).

[6] Chen L, Lin H, Li S (2012) Depth image enhancement for Kinect using region growing and bilateral filter. In: Proceedings of the 21st international conference on pattern recognition (ICPR2012). IEEE, pp 3070–3073

[7] Vaishali.S.Kulkarni et al., "Appearance Based Recognition of American Sign Language Using Gesture Segmentation", International Journal on Computer Science and Engineering (IJCSE), 2010

[8] Cheok, M. J., Omar, Z., &Jaward, M. H. (2019). A review of hand gesture and sign language recognition techniques. International Journal of Machine Learning and Cybernetics, 10(1), 131-153

[9] Al-Saffar, A. A. M., Tao, H., & Talab, M. A. (2017, October). Review of deep convolution neural network in image classification. In 2017 International Conference on Radar,

Antenna, Microwave, Electronics, and Telecommunications (ICRAMET) (pp. 26-31). IEEE.
[10] Kiron Tello O'Shea, An Introduction to Convolutional Neural Networks, (2015 November). Research GATE
[11] https://www.exastax.com/deep-learning/top-five-use-cases-of-tensorflow/
[12] https://en.m.wikipedia.org/wiki/OpenCV
[13] https://github.com/tzutalin/labelImg
[14] [Page 538 https://www.amazon.com/Deep-Learning-Adaptive-Computation-, Deep Learning, 2016.]
[15] https://machinethink.net/blog/mobilenet-v2/ by Matthijs Hollemans [22 April 2018]