Indian Sign Language Recognition using Deep Learning.

Aman Desai-(19IT031)
Student, B.Tech-IT
CSPIT,CHARUSAT
Changa, Anand, Gujarat, 388421, India

Mansi Desai-(19IT032)
Student, B.Tech-IT
CSPIT,CHARUSAT
Changa, Anand, Gujarat, 388421, India

Hardik Gandhi-(19IT037)
Student, B.Tech-IT
CSPIT,CHARUSAT
Changa, Anand, Gujarat, 388421, India

Jalpesh Vasa
Assistant Professor
CSPIT,CHARUSAT
Changa, Anand, Gujarat, 388421, India

Abstract:

Sign languages are good way to help the deaf and dumb community to interact with normal people. A lot of research has been done in the field of American Sign Language (ASL), but unfortunately the same cannot be said for Indian Sign Language (ISL) as ISL has got less attention by the researchers around the world. Major barriers/problems we got are Lack of standard dataset, variation in The language with locality. Experiments are performed on our own continuous ISL dataset which is created using laptop webcam at home. In this paper we proposed an Indian Sign Language (ISL) gesture recognition system where only one hand or both the hands are used for performing any gesture. The proposed method doesn't require that signers wear gloves or any other marker devices to simplify the process of hand segmenting. Though existing systems identify Indian sign language with sufficient accuracy, this proposal incorporates the classification of images using Convolutional Neural Network(CNN).

Keywords: Sign Language, ISL, CNN, Deep Learning, Deaf and Dumb.

I. INTRODUCTION

Humans have many different ways to communicate to express themselves for e.g, by verbally speaking in different regional language or by expression and many more. Talking to a person in his own language makes the conversation more fruitful. Sign Language is specifically focused towards Deaf and Dumb People. It's the language deaf and dumb people use to communicate with each other and other people who know SL(Sign Language). Sign Languages are gestural languages which contain symbolic encoded message for communication without speech channel. They are unique in some ways in that they cannot be written like spoken language. Although Sign Language is a medium of communication to deaf people, it still doesn't have any meaning when conveyed to a person who doesn't know sign language. Thus, it's broadening the communication gap. To prevent this from happening, we are putting forward a sign language recognition system.

Our system will help to reduce the communication gap by helping the hearing disable people to communicate their thoughts to other normal people. Sign language varies from country to country with its own vocabulary and grammar like American Sign

Language(ASL), French Sign Language(FSL), Indian Sign Language and many more depending upon the region. Our main focus is on Indian Sign Language (ISL) which is used by Indian deaf and dumb community. In India the count of hearing impaired people, is more compared to other countries. Not all of them use ISL but, more than one million deaf adults and around half million deaf children use ISL as a mode of communication. But as per our research, Among all the sign languages , American Sign Language (ASL) is the one in which most research has been done and there hasn't been significant amount of research done on ISL. So we have selected ISL. Our attempt is to overcome the communication barrier in such a way that it may help normal people and speech and hearing impaired people to communicate with each other effectively.

II. LITERATURE SURVEY

Literature review of the proposed system shows that there have been many explorations done in this field to tackle the sign recognition in videos and images using several methods and algorithms.

Deaf people, who live in villages usually, do not have access to sign language. However, in all large towns and cities across the Indian subcontinent, deaf people use sign language which is not standard sign language. Extensive work and awareness program are being done for implementation of ISL in education systems [1].

As per S.Saravana Kumar and Vedant L. Iyangar [2], they proposed a system which was trained under the machine learning algorithm which is Support Vector Machine (SVM). SVM trai0ning algorithm builds a model that assigns new examples to one category or the other, making it a non-probabilistic binary linear classifier. In there system some times many letters were mis-interpreted. Thus there model accuracy was not enough.

Sanil Jain and KV Sameer Raja [3] worked on Indian Sign Language Recognition, using coloured images. They used feature extraction methods like bag of visual words, Gaussian random and the Histogram of Gradients (HoG). Three subjects were used to train SVM, and they achieved an accuracy of 54.63% when tested on a totally different user.

Also, there are some common wrong beliefs about sign language which is reported in ISL literature [4]:

i) "Sign language is same all over the world"
ii) "Sign language is not a complete language. It is just a sort of pantomime or gesturing, and it has no grammar"
iii) "Sign language is dependent on spoken language. It is a representation of the spoken language of the hands"
iv) "Sign language is the language of the hands only"
v) "Sign language has been invented by other people to help deaf people"
vi) "Signed Hindi or signed English is better than Indian sign language"

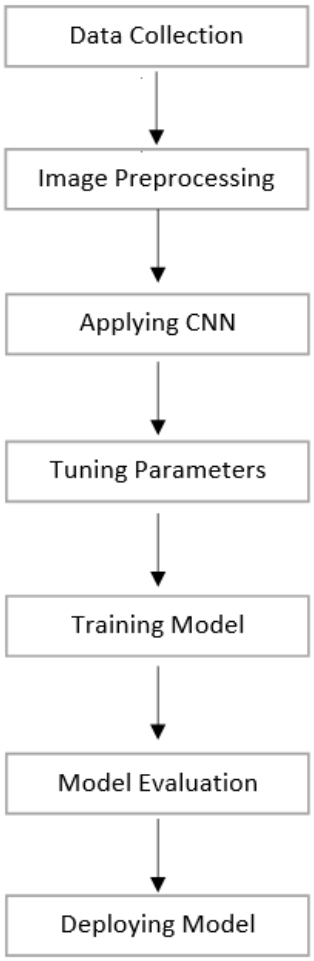
So overcoming these wrong beliefs, there is a need of developing ISL interpretation system to aid Indian hearing impaired people and making them literate and self dependent.

III. METHODOLOGY

First step was to find a good dataset. But in market there was no such standard/reliable datasets of Indian Sign Language gestures. So we decided to make custom dataset. With the help of OpenCV library and laptop webcam we collected train and test separate dataset of total 36 classes i.e. 26 Alphabets + 10 Digits. Our custom dataset contains approx. 3000 images.

The train test ratio is 70:30. After the data-collection the next step was to apply filters on the images i.e. image processing.

To increase the size of dataset artificially we have used data augmentation method. Through data augmentation we can artificially increase number of train dataset images by applying random transformation filters to the images. The filters may be like randomly cropping images, flipping them vertically or horizontally, and many more. Because of data augmentation we can make our model robust in such a way that the objects in the image should be recognized regardless of the orientation. It also makes the model invariant to transformation of the provided input data. Then Convolutional Neural Network is applied for training and for classifying the images. The last step was to evaluate from the real time video.



- **Gestures**

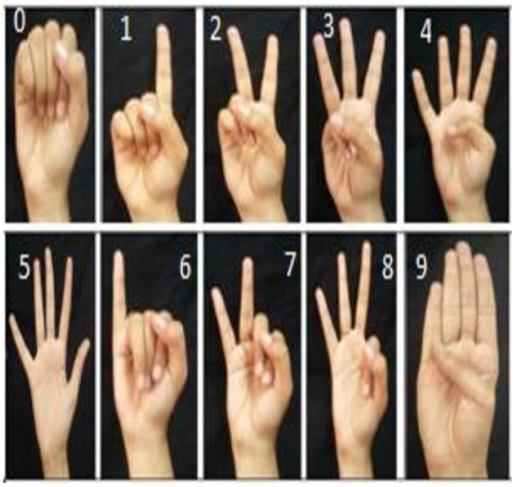
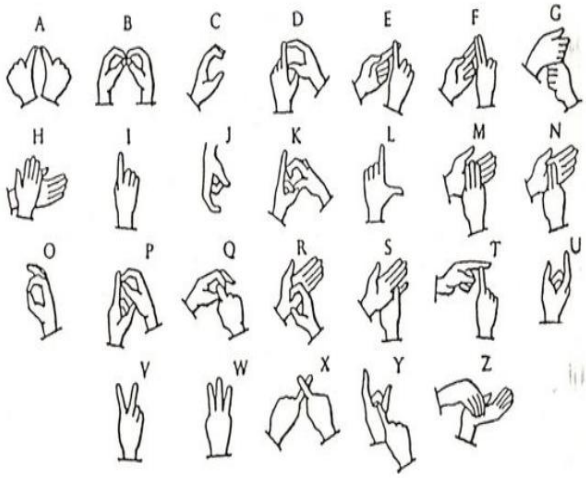


Figure 1. Indian Sign Language (ISL) numbering system



IV. SYSTEM ARCHITECTURE

Our system consist of a CNN model. It is used to extract features from the frames and to predict hand gestures. A CNN model consists of four main operations: Convolution, Relu(Non-Linearity) Activation Function, Pooling and Classification (Fully-connected layer)[5].

Convolution: Convolution is a matrix operation use to apply filter on the input image. Its purpose is for extracting features from the provided input image. It preserves the spatial relationship between the pixels by learning extracted features. It is usually followed by activation function (here Relu).

Pooling: In ConvNet, it's a common practice to add pooling layer after convolution layer. The input image resolution is lowered to such a limit that only required details are kept.

Fully-connected layer: Fully connected layer is a multi layer perceptron that makes use of softmax activation function in the output layer. It's main purpose is to use features from previous layers for classification of images into different classes based on the requirement.

The combination of these layers is used to create a CNN model. The last layer is a fully connected layer.

Number of Epochs: It basically is the number of times the dataset is passed to the neural network for training the model. There is no such formula or ideal number for the epochs. It depends on the dataset. Higher the number of epochs higher will be the training time. Also higher number of epochs may lead to better accuracy.

Activation Function: Its purpose is to decide whether the neuron is activated or not. It is applied to the value of sum of weights and bias. Here ReLU activation function is used , which is most commonly used activation function because of its computational advantages. This function works in such a way that it returns zero for inputs that are negative and for the positive inputs it returns the value itself. There are many other activation function like Sigmoid, Tanh, etc but for most of the modern neural network, Relu is used.

Softmax: This layer is used in multi-class classification right before the output layer. The parameter passed to softmax is the number of classes in classification. It gives the likelihood of the provided input image which is belonging to particular class/category.

Our Proposed system is capable enough of recognizing gestures from real time video. OpenCV library is used to access the laptop webcam and process the video frame by frame and after extracting the feature the image/gesture is recognized. Currently, the ROI must contain a gesture of hand with a proper blank/clear background.

Before using the actual system as mentioned above we first need to train the dataset to make model with CNN.

V. CONCLUSION

We can conclude that Convolutional Neural Networks(CNN) can be used as classification algorithms for sign language recognition systems as it provides better accuracy. However, pre-training has to be performed with a larger dataset in order to show increase in accuracy. We were able to achieve approximate accuracy of 90%. The system requires some constraints, like a clear background and the palm of the hand to face the camera.

Further, similar symbols were sometimes misinterpreted for one another. This could be due to the limited training provided to the system. The system might overcome these limitations if a more detailed dataset in different environmental conditions is provided for training. Also letters that needed movement of the hands could not be properly identified.

VI. FUTURE WORK

In our future work, we will also train model on some short videos of some commonly used phrases like I am coming , How are you, I am Hardik(name of the person), This is my house etc.

As this dataset is custom made, the number of images are not enough to gain more accuracy so we will try to increase number of images as well as accuracy at different angles.

Pre Training the model with detailed dataset:

The system can be trained with a more detailed dataset with thousands of samples for each letter of the alphabet, spanning different environmental conditions, lighting, different hand positions, and skin tones for optimizing.

VII. REFERENCES

[1] U. Zeshan, M. Vasishta, M. Sethna, “Implementation of Indian Sign Language in Educational Setting”, Asia pacific Disability Rehabilitation Journal,Vo.16, No.1, 2005, pp. 16-39.

[2] S.Saravana Kumar, Vedant L. Iyangar, "SIGN LANGUAGE RECOGNITION USING MACHINE LEARNING", International Journal of Pure and Applied Mathematics, Volume 119 No. 10 2018, 1687-1693.

[3] Sanil Jain , K.V.Sameer Raja, " Indian Sign Language Character Recognition".

[4] U. Zeshan, “ ‘A’ level Introductory course in INDIAN SIGN LANGUAGE”, Ali Yavar Jung National Institute for Hearing Handicapped, Mumbai, 2001, pp. 1-38.

[5] https://edu.authorcafe.com/academies/6813/sign-language-recognition

Indian Sign Language Recognition

ORIGINALITY REPORT

37%	33%	23%	%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	www.s2is.org Internet Source	14%
2	edu.authorcafe.com Internet Source	9%
3	www.ijert.org Internet Source	5%
4	Alpesh Vala, Amit Patel, Riddhi Goswami, Keyur Mahant. "Defected ground structure based wideband microstrip low-pass filter for wireless communication", Microwave and Optical Technology Letters, 2017 Publication	1%
5	www.mdpi.com Internet Source	1%
6	www.taibahu.edu.sa Internet Source	1%
7	Kumud Tripathi, Neha Baranwal, G.C. Nandi. "Continuous dynamic Indian Sign Language gesture recognition with invariant backgrounds", 2015 International Conference	1%

on Advances in Computing, Communications and Informatics (ICACCI), 2015
Publication

8	ijarcce.com Internet Source	1 %
9	Archana S. Ghotkar, Dr. Gajanan K. Kharate. "STUDY OF VISION BASED HAND GESTURE RECOGNITION USING INDIAN SIGN LANGUAGE", International Journal on Smart Sensing and Intelligent Systems, 2017 Publication	1 %
10	Kumud Tripathi, Neha Baranwal G.C. Nandi. "Continuous Indian Sign Language Gesture Recognition and Sentence Formation", Procedia Computer Science, 2015 Publication	1 %
11	www.ijcse.com Internet Source	1 %
12	libres.uncg.edu Internet Source	1 %
13	onlinelibrary.wiley.com Internet Source	<1 %

Exclude quotes	Off	Exclude matches	< 5 words
Exclude bibliography	On		