American Sign Language Recognition Using CNN Algorithm

[1] Shardul Rudraksha,[2]Satyadev Mishra,[3]Vipul Munot, [4] Swapnil Misal, [5] Yogita Kandhare

[6] Prof. Abhilasha Kulkarni

[1,2,3,4,5] UG Students, Department of Computer Engineering, MMCOE, Karvenagar , Pune, India,

[6]Assistant Professor, Department of Computer Engineering, MMCOE, Karvenagar , Pune, India.

Abstract: **According to W.H.O (world Health Organization) 446 million people i.e. 5% of word population facing disabling hearing loss. These people can only communicate through sign language. The drawback of this way of communication is that most of the normal people cannot understand sign language. This makes hard to communicate for both. With the help of our project we are trying to build a system which act as interpreter for both types of these people. With the help of AI and Deep learning we can build a model that can understand sign language and convert it into normal language. Our project aims to bridge the gap between the speech and hearing impaired people and the normal people.**

***Key Words*: Machine learning, Computer vision, Sign language, Convolution Neural Networks (CNN), Sign recognition.**

I. INTRODUCTION

To successfully recognition of hand sign a system has to extract the features such that shape and dimension of palm. CNN (Convolutional Neural Network) is deep learning algorithm. It is highly capable of recognition and classification for image as input with high accuracy. Following fig. shows the network connection of CNN algorithm.

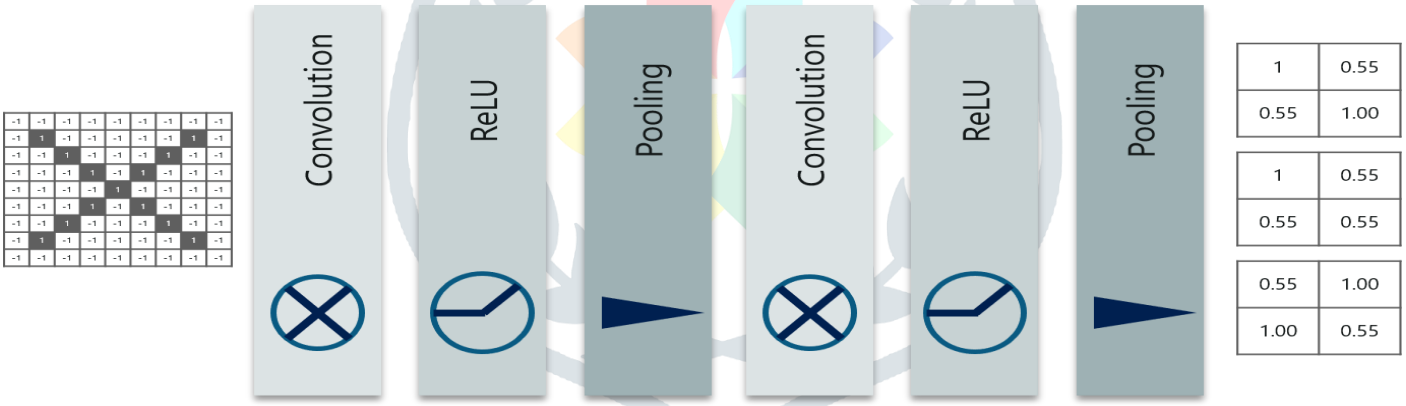


Figure 1: Convolutional Neural Network Architecture

The main layers are convolution layer, pooling layer and finally connected layer. By using multiple layers we can extract more features from images and finally we stack these layers to form fully-connected layer. CNN is based on tenser flow library which comes under deep learning. CNN Generally used for images for classification as it performs well against other deep learning network.

1.1 Convolution layer

In Convolutional layer input images is represented as matrix and a filter or kernel k is used to extract the features. Kernel scans the image from top left to right and moving downwards bit by bit. By moving kernel over input image we subsequently build our feature map.

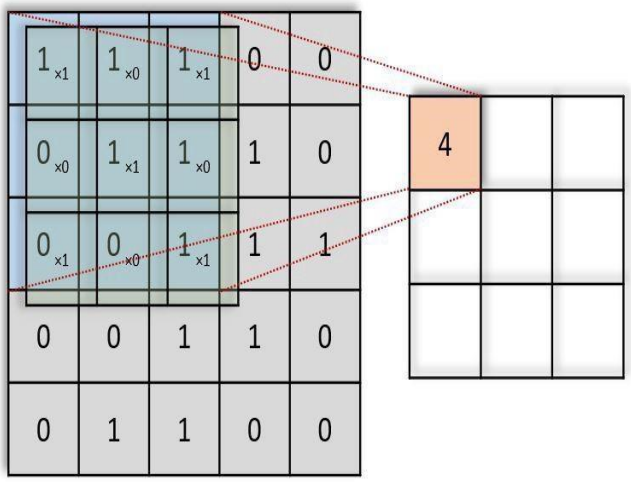


Figure 2: Convolutional

The feature matrix is formed by using following

$$G[m,n] = (f * h)[m,n] = \sum_j \sum_k h[j,k] f[m-j, n-k]$$

Where G [m, n] is feature matrix, f denotes input image and h is kernel. Index m, n represents the row and column of the matrix. With each successful filtering we multiply each value of kernel with corresponding value of the image and finally we sum them up and place them to the feature map.

Rectified Linear Unit used as activation function in CNN. It simply converts the negative values to '0' and keeps positive

1.2 ReLU

values as it is.

1.3 Pooling

After getting feature map pooling is applied to reduce the spatial size of feature map. Pooling reduce the complexity and computational power. By aping pooling we further extract the dominant features from image. It speeds up the process and avoid over fitting. There are two types of pooling first is Max pooling and other is Average pooling.

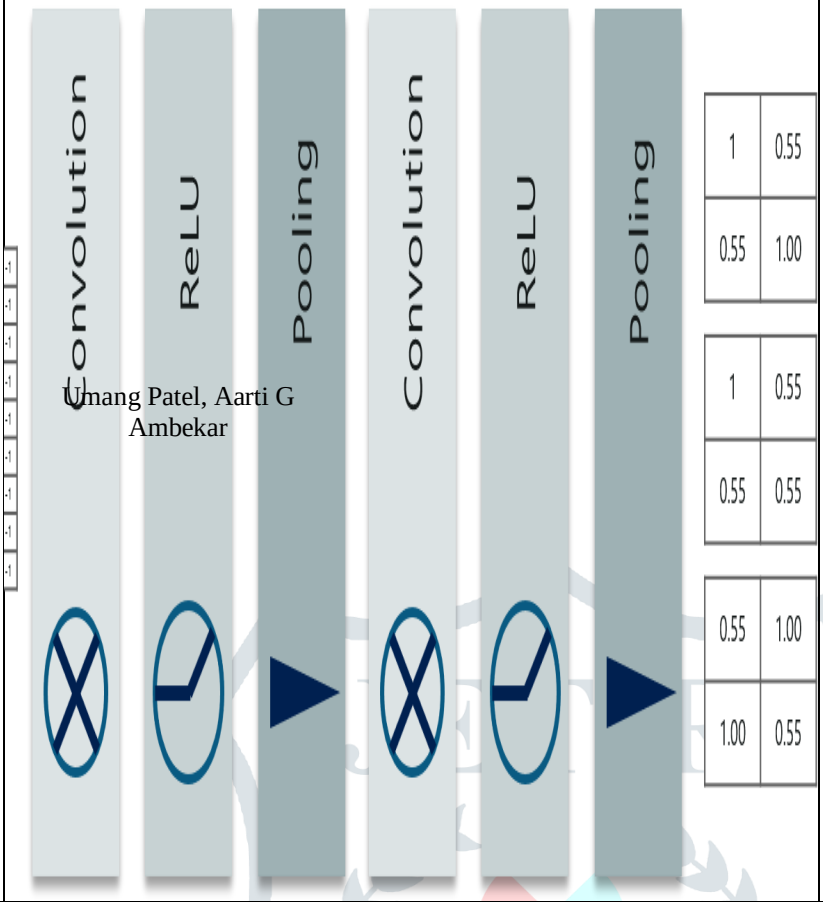
i) Max-pooling: it returns maximum vremotelyalue from the portion of feature map with windows size. ii) Average pooling: it returns average value from the portion of feature map with windows size.

1.4 Classification

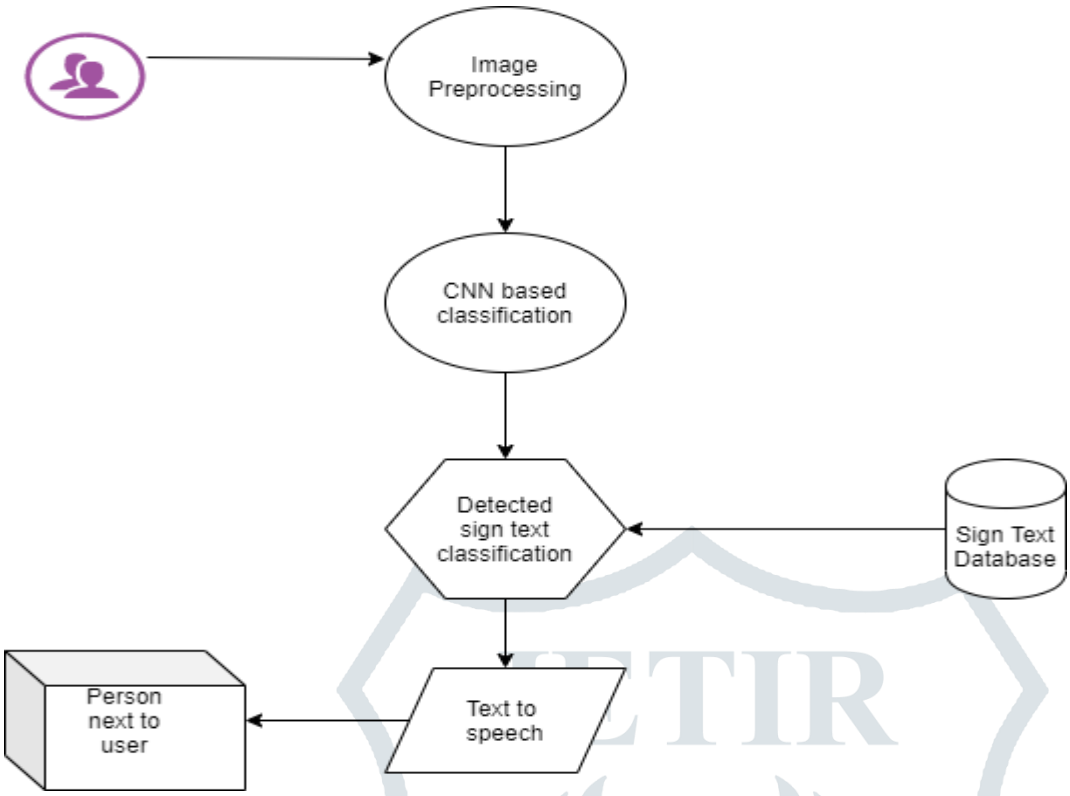
For classification a feature matrix gets flatten and converted into column vector. This vector is now fed to fully connected layer and back propagation is used to learn from this vector. With each iteration or epoch high level of features are learned by the networks.

II. LITERATURE REVIEW

Title	Author	Details of Publication	Description
American Sign Language recognition using deep learning and computer version	Kshitij Bantupalli , Ying Xie	2018 IEEE	In this paper author said that With recent advances in deep learning and computer vision there has been promising progress in the fields of motion and gesture recognition using deep learning and computer vision based techniques. The focus of this work is to create a vision-based application which offers sign language translation to text thus aiding communication between signers and non-signers.

Moment Based Sign Language Recognition for Indian Languages	 Umang Patel, Aarti G Ambekar	2018 IEEE	In this paper minimises the barrier of communication between them. In this paper, Hand gestures are captured, processed with the help of MATLAB and then converted into speech & text. For speech & text, two languages are chosen English & Hindi. In this paper features value of images is evaluated based on moment technique.
User Dependent Sign Lan guage Recognition Using Motion Detection	Mohamed Hassan, Khaled Assalhen	2018 IEEE	In this paper they provide proposed system which is tested on two datasets. The first was collected using DG5-VHand data gloves and the second was collected using Polhemus G4 tracker. Each dataset was collected by a different signer.

III. IMPLEMENTATION



There are 3 main components for this project. They are as follow:

For this project we have taken custom dataset. From A to Z alphabet we have taken 2000 images.

These images then passed through image processing unit. With the help of Image processing unit, on each image basic segmentation and morphological operations were performed to improve the features of image for better model training. After image pre-processing images are divided into 75:25 ratios for training and testing purpose. After splitting of the data, it is fed on CNN algorithm for classification.

CNN Image classifier

1. Input data

2.

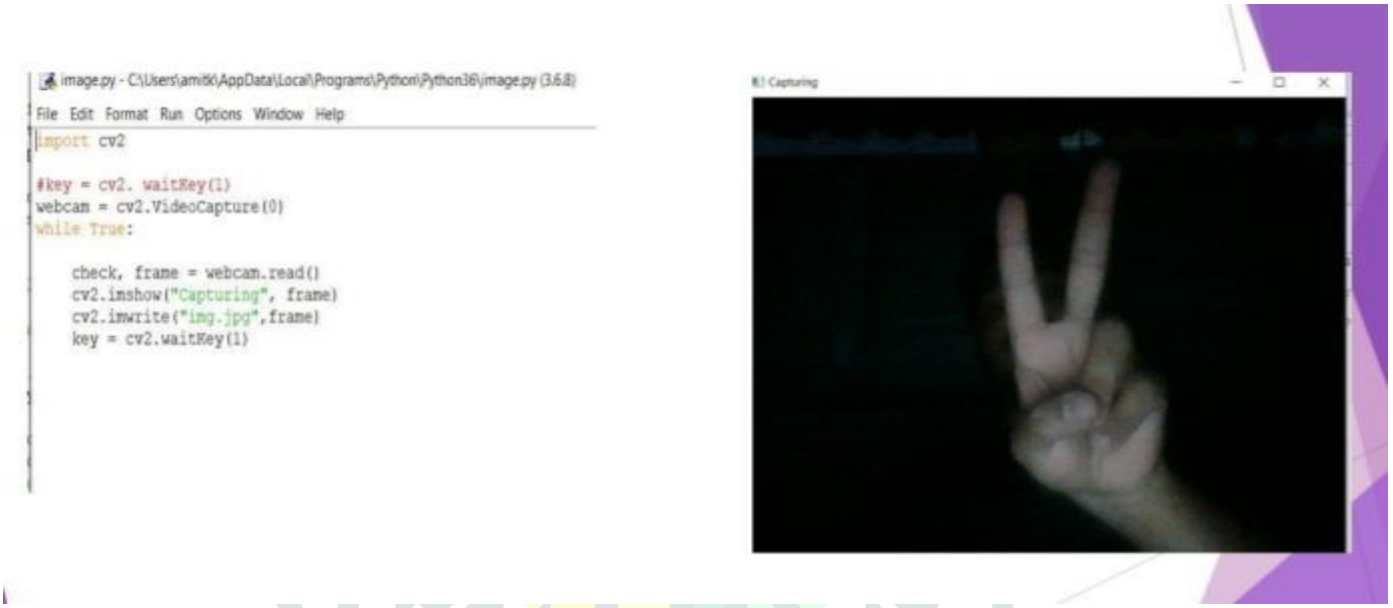
This module used to train the dataset and classify them according to their respective classes. When input images are fed in this algorithm via web cam or external camera, CNN classify that image against 26 classes. After classification we get Image classified with respect to its class.

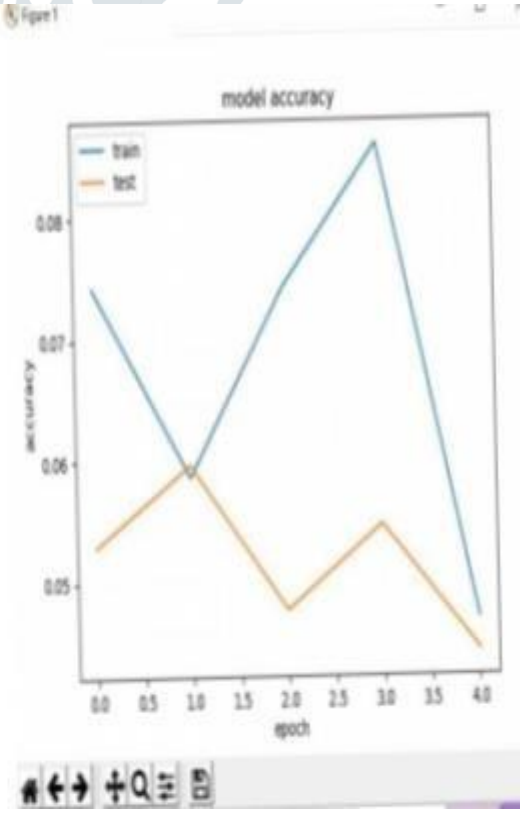
3. Text to Speech

After successful classification the output is converted into audio format so that both users can confirm the output and can agreed on same statement. For converting text to speech we are using python's text to speech library. Pyttx3 is library is compatible for both python 2 and python 3. Unlike other text-to-speech library Pyttx3 can work offline.

For front-end we have used flask web frame work.

IV. Result





V. Conclusion and Future work

We have successfully implemented project by achieving higher accuracy of image detection by using CNN algorithm .Our results are promising by comparing through various algorithm .To take project further we can implement for other sign language worldwide and can be deployed it on android / ios systems.

Further We can implement the project on hardware devices such as raspberry-pi that can be placed in various restaurants, offices ,malls .We can provide text-to-speech input for detected sign language. Project can be further extended to video calls and online meetings.

REFRENCES

1.Kshitij Bantupalli, Ying Xie "American Sign Language recognition using deep learning and computer version" IEEE 2018.

2. Umang Patel, Aarti G Ambekar "Moment Based Sign Language Recognition for Indian Languages" IEEE 2018.

3. Mohamed Hassan, Khaled Assalhen " UserDependent Sign Language Recognition Using Motion Detection" IEEE 2017.