International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 05 Issue: 04 | April - 2021 ISSN: 2582-3930

GESTURAL TALK USING DEEP LEARNING (CNN)

Kiran P, Lakshana R, Sujeetha S

[1]*Information Technology &Sri Sairam Institute Of Technology*
[21]*Information Technology &Sri Sairam Institute Of Technology*
[31]*Information Technology &Sri Sairam Institute Of Technology*

ABSTRACT - Speech and hearing impairment is a disability which affects one's ability to speak and hear. Such individuals use sign language to communicate with other people . Although it is an effective form of communication, there remains a challenge for people who do not understand sign language to communicate with speech impaired people. The aim of this paper is to develop an application which will translate sign language to English in the form of text and audio and vice versa, thus aiding communication with sign language. The application acquires image data using the webcam of the computer, then it is pre-processed under various layers of convolutional neural network and recognition is done using classification. The translation in the form of text is then converted to audio

Key Words: SignLanguage,CNN,Preprocessing, Image classification,Speech impaired.

1.INTRODUCTION:

Sign Languages (also known as signed languages) is a language that uses hand shape ,palm orientation, movement , location and facial to convey meaning . Like any other language sign language is also has its own grammar and lexicon .Though there is no specific or common Sign language that is used worldwide. But there are more similarities among all Sign language .According to linguists , sign language and spoken language are the two types of natural language that is, these two emerged over an abstract and evolved over time. Sign Language is different from body language .A group of people in the local use sign language as a means of communication and form the core of deaf culture .Sign language is primarily used by deaf and hard of hearing it is also used by people who have trouble with spoken language due to disability or people with who are with deaf family members like children or adults.

Every local people use their own sign language that is their native or local Sign language .It is not known how many sign language exits worldwide .Legal Recognition was given to some sign language .In Ethnologue 2020 edition reports that 144 Sign languages whereas SIGN-HUB Atlas of Sign Language Structures reports 200 of them and more where not yet discovered.

2.LITERATURE SURVEY:

Many sign language recognition systems have been proposed. The authors recognize and classify different hand signs using computer vision. The features differentiated based on texture, color, and shapes of different hand sign. They use Support Vector Machine(SVM) for classification. The average Accuracy rate was 83.84%.

In Ref[1] Omkar vedak,Prasad zavre,Abhijeet,Manoj uses SVM (Support Vector Machine)algorithm for classification. In this proposed system In this system is comprised of two main stages that is Hand Segmentation and Recognition of Hand Sign. Hand Gestures are captured in the form of Video and by using image processing and Edge detection segmentation of hand is done by using various methods such as Fourier Descriptor, Scale Invariant Feature Transform (SIFT) or Principle Component Analysis (PCA) . Histogram of Oriented Gradients (HOG).Then for the Recognition Image Classifier Support Vector Machine and template Matching is used. Output is given in the form of Text.

S Kumutha ,K Preethi Mane [2] In this system they design a model for smart gloves and design a flex sensor which is used for recognition of sign language .This model is totally based on detection of movements of fingers and gestures of hand to correctly identify the hand sign using LabView software and a acquisition device for a data .The sign is recognized from the sensors in the gloves and it is processed and gives an output as an audio .

J.Yashas,G Sivakumar [3] In this system they presented a literature survey on Hand Gesture recognition HGR , having reached all the best possible ways for data acquisition like cameras, wrist sensors, hand gloves.

VijayKumar,Ananthu Nair,Nishanth Tomar [4] The system mainly consists of an Arduino microcontroller and flex sensors. Flex sensors are used for sensing the gestures .The output of the flex Sensor is processed by the Arduino. The output from the microcontroller is then transmitted via Bluetooth module. LCD display and Speaker is connected to the Arduino Microcontroller .An android device which is connected to the device uses app to convert sign to speech .

Purva Badhe [5] uses Fourier Descriptor for feature extraction. The system translates Indian Sign language gestures to English

language. To represent the boundary points, the Fourier Series were calculated using Fast Fourier Transform (FFT) algorithm. The extracted data being too large is compressed using vector quantization. This data is then stored into a codebook. For testing purpose, the code vector generated from gestures is compared with existing codebook and gesture is recognized.

3.METHODOLOGY:

This proposed system consist of three main process

3.1.Static Hand Gesture Recognition

3.2.Dynamic Hand gesture Recognition

3.3.Text to Sign Language conversion

In this (Convolutional Neural Network) CNN used as classifier for finding the hand sign through web camera .We trained the (Convolutional Neural Network)CNN using around 300 images per sign.

3.1 STATIC HAND GESTURE RECOGNITION:

In Static Hand Gesture Recognition the input is in the form of image captured from web camera.This system consist 2 main stages [1]Dataset collection for hand gesture,[2]Training and testing the datasets for hand gesture .The block diagram shows the working of the system.

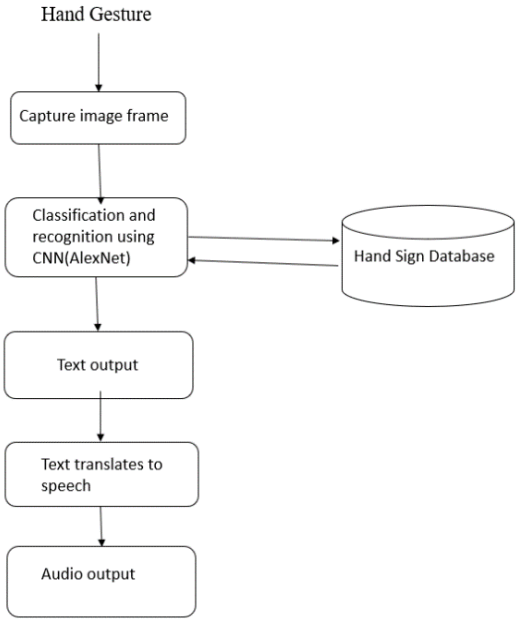

Fig 1.Static hand gesture recognition flowchart

3.2DYNAMIC HAND GESTURE RECOGNITION:

In dynamic hand gesture recognition the input is in the form of video captured from web camera. This system also contains 2 main stages [1]Dataset collection for hand gesture[2]Training and testing the hand dataset. The block diagram shows the working of the system.

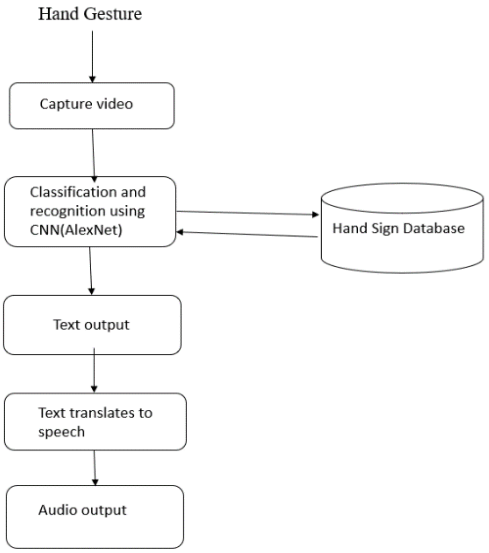

Fig 2.Dynamic hand gesture recognition flowchart

3.1.1,3.2.1DATASET COLLECTION FOR STATIC AND DYNAMIC HAND GESTURE RECOGNITION:

The dataset collection is common for both static and dynamic hand gesture recognition .We have created a data set for English words. Each sign gesture is collected from different individuals from different angles ,size and color of the hand.

We have collected around 300 images per hand sign and it is stored in the database.

3.1.2,3.2.2TRAINING AND TESTING FOR STATIC AND DYNAMIC HAND GESTURE RECOGNITION:

For training the dataset we use AlexNet in matlab which contains 25 layers .In the 1st layer hand gesture input(227 x 227) is given then it goes to convolution layer the input is sharpen , the edges are enhanced and the parameter for the feature is identified through filters also known as Kernels.

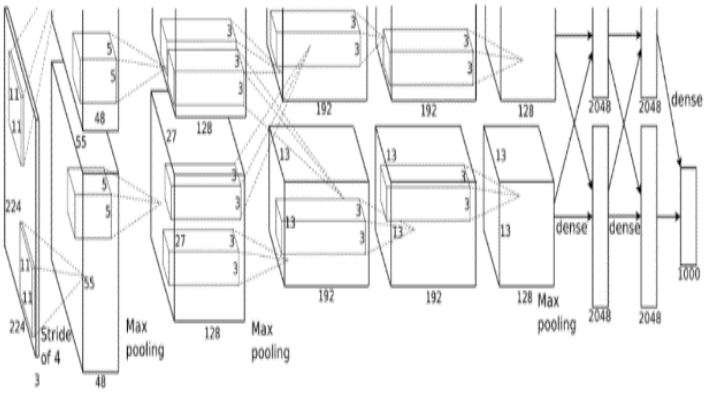

Fig 3.AlexNet Architecture

Next it goes to Rectified Linear Unit (ReLU) Layer which applies the function f(x)=max(0,x) to all input volume values. This layer just change all negative activation to 0.This layer is used for increasing the nonlinear property for model. Next it goes to 4th layer called cross Layer. This layer is used for feature extraction from multiple layers. Next it goes to Max Pooling layer .In this layer Which selects the maximum element from feature map covered by filter .The output after max pooling will have only the most prominent features. For example in the following diagram a 2x2 region is replaced by the maximum value in it.

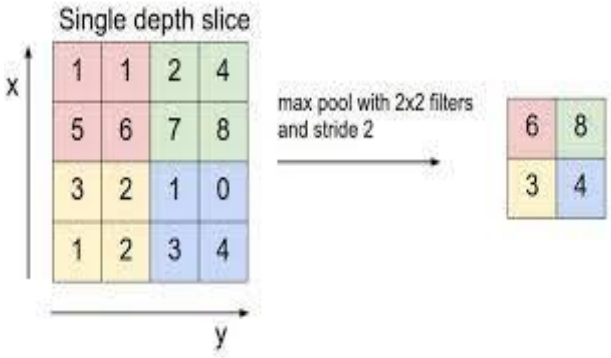

Fig 4.Max Pooling

Next from layer 3 to 5 will be repeated again .Then again it goes to convolutional layer.Next the layer 3 to 5 is repeated again (ie) ReLU ,Cross ,Max Pooling Layers and next goes to convolutional layer for further filtration .Then it goes to fully connected layer. This layer gets input from output of pooling or convolutional layer which is flattened and fed into this layer. This layer compiles the data extracted from previous layers to form an output .Then it goes to ReLU layer .Then it goes to Dropout layer which is used to prevent from overfitting then fully connected ,ReLU ,Dropout are repeated again and again. Next it goes to Softmax Layer .This layer converts vector of numbers to vector of probabilities .Next it goes to classification layer where the classification gets completed and it is the output layer. After training we test the system whether it gives the correct output for the respective hand gesture input.

3.1.3,3.2.3TEXT TO SPEECH TRANSLATION:

We have used TTS APIs , to transforming sign gesture to audio, It is one of the best text to speech APIs .The sign language is translated into text using above method and the result is fed to text to speech function which converts text to audio.

3.3TEXT TO SIGN LANGUAGE CONVERSION:

This system contain 2 main stage [1]Dataset collection for text and hand gesture[2]String Matching .The block diagram shows the working of the system.

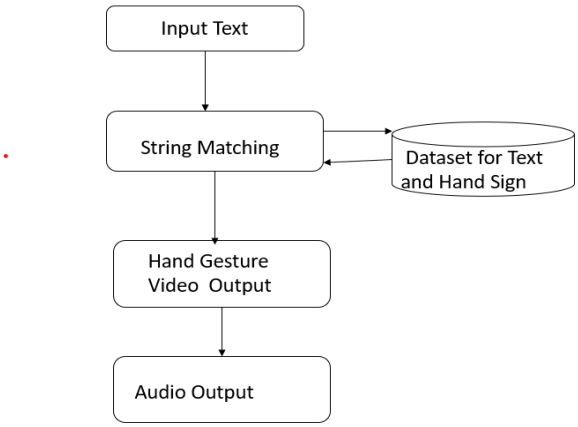

Fig 5 .Text to sign language flowchart

3.3.1DATASET COLLECTION FOR TEXT AND HAND SIGNS:

We have created the dataset for English words of sign language and collected the datasets for hand gesture video and text and it is stored in a database.

3.3.2 STRING MATCHING:

In text to sign conversion we have used the concept of string matching .For example if we give the input text as hello with the help of string matching the system checks the respective hand gesture video for the input text and gives the output as a video and audio.

4.CONFUSION MATRIX:

	Good	Help	L	No proper image detected	O	Thank You	Two	Why	Yes	Hello	Total	UA
Good	10	0	0	0	0	0	0	0	0	0	10	100
Help	1	9	0	0	0	0	0	0	0	0	10	90
L	0	0	10	0	0	0	0	0	0	0	10	100
No proper image detected	0	0		10	0	0	0	0	0	0	10	100
O	0	0	0	0	10	0	0	0	0	0	10	100
Thank You	0	0	0	0	0	10	0	0	0	0	10	100
Two	0	0	0	0	0	0	10	0	0	10	10	100
Why	0	0	0	0	0	0	0	10	0	0	10	100
Yes	0	0	0	0	0	0	0	0	10	0	10	100
Hello	0	0	0	0	0	0	0	0	0	10	10	100
Total	11	9	10	10	10	10	0	10	10	10	100	
PA	100	90	100	100	100	100	100	100	100	90	OA=	90%

FIG.6.Confusion Matrix

We have implemented around 10 sign gestures with 300 training images each .our accuracy is 90% it was calculated using confusion matrix.

5.CONCLUSION:

In this project, to deal with the common aiming problem for speech and hearing impairment persons . Our proposed system "GESTURAL TALK USING DEEP LEARNING(CONVOLUTIONAL NEURAL NETWORK)" is the advanced, reliable and robust version of interface between normal and deaf and dumb people to communicate with each other. The proposed GESTURAL TALK system we have collected the data of around 300 hand gestures per sign language. These hand gestures are trained using deep learning, still hand image frame is captured employing a digital camera. These frames are processed to urge increased options. Then feature extraction and classification algorithms are used to translate the sign language into English text. This translation is born-again to speech text to speech API. The system has implemented the interface between normal people and people with hearing impairment .This is done by using convolutional neural network (CNN).

REFERENCES:

1. Omkar vedak,Prasad zavre,

Abhijeet,Manoj(2019).Sign Language interpreter using image processing and machine learning in International Research Journal of Engineering and Technology(IRJET).

2.S.Kumutha,KPreethiMane(2020).Smart asisitant for deaf and dumb using flexible resistive sensor in Institute of Electrical and Electronic Engineering(IEEE).

3.J.Yashas,G.Sivakumar(2019).Hand Gesture Recognition : A Survey in Institute of Electrical and Electronic Engineering(IEEE).

4. K P VijayKumar,Ananth Nair,Nishanth Tomar(2020).Hand gesture to speech and text conversion device in (IJITEE).

5.Purva C. Badhe, Vaishali Kulkarni, "Indian Sign Language Translator using Gesture Recognition Algorithm", 2015 IEEE International Conference on Computer Graphics, Graphics and Information Security (CGVIS).