Two-way communication for dumb/deaf using Machine learning and CNN

Vignesh R
Vinayak B Bidari
Vishal Mise
Geetha B

School of Computer and engineering
REVA university
Bangalore, India

Abstract-- For several reasons, there has been a casual rise in several dumb and deaf cases. The importance of communication for them is that they communicate through sign language. A gesture is used by the dumb and deaf to convey a message or an emotion to others, the main part of the movement is the torso, hand, and movement segment. The dumb and the deaf can communicate among themselves using sign language. But ordinary people find it difficult to communicate with the dumb and deaf, as they are not aware of those sign languages. Even the average person cannot effectively communicate using sign language with the dumb and deaf. The main aim of this paper is to provide an effective way in which there is 2-way communication with the common person, the dumb, and the deaf. The proposed model is an actual-time application for hand gesture recognition that recognizes indications and then converts gesture images into text or speech. This conversion will allow them what they want to communicate in no time. The sign-to-speech conversion is done using CNN. This application can prove to be an effective communication tool for dumb and deaf people to use in daily life. This application can prove to be effective in a video chat application for communicating things online.

Keywords- Deaf and dumb, Hand Gesture recognition, CNN, YOLO, Sign to speech, sign to words

1. INTRODUCTION

Each day recurring, we will communicate with each other with the use of speech and hand gestures to make others understand our thoughts. Gestures are a better and more natural way for humans to engage with computer systems, and as a result, it creates a richer bridge between humans and technology. Sign language translation is one of the most rapidly evolving fields of study in recent years, and it is by far the most natural mode of communication for those with hearing impairments. We have developed a system that automatically converts the Sign language to text and speech. We have developed a huge set of samples to get more efficiency and understand the difficult sign language gesture by using a digital camera. This system proposes a very effective and easy-to-use system for the dumb and deaf which helps in verbal conversation and learning. Sign language conversion is a very important line of study in recent times. There have been many studies showing some information about the conversation between dumb and deaf. The shape, placement, motion of palms, further facial expressions, body actions, each plan y vital thing carrying records. signal language is not a normal language — in the entire USA. It Has its very individual language, and areas have dialects, just like the many languages are spoken anywhere in the internationally talking language, the detection price through the ASL language as in compared to the linguistic accuracy is 90 % proportion of organizations generally use Indian sign language The proposed model is CNN primarily based value which the changes sign language to text and speech. Here CNN is used because of its high mapping than the other comparison methods. This is the main application this system focuses on which converts the signal to text or speech with 95% accuracy. The first system is to locate the signal language and convert it to text. the main method here is to 1) Setup a UI to take his or her to enter from the person 2) Next is to test and train the information acquired from the user 3) form words and 4) conversion of text to speech. In the developing version, we use four layers Convolutional Layer, Pooling Layer, fully related Lanard and very last Output Layer. This version will help convert the sign to textual content conversion. you can then use Google APIs to convert text to speech.

2. LITERATURE SURVEY

[1] The machine proposed in this article is entirely based on a creative and scientific approach to hand popularity, which is more general, safer, and now does not require a database to learn about specific gestures. Hand gestures should be identified in different lighting conditions. Feature extraction and classification methods require several methods, and it is difficult to choose which method to apply. The proposed technique performs background segmentation of the hand according to the obtained statistics and then assigns specific gestures to special alphabets. This entails a feature extraction strategy for top-level calculations and hand gesture ratio calculations, sooner or later the gestures will be recognized and converted to speech and vice versa. Speech recognition machines are used to convert acoustic speech into the form of gestures. [2In this article, this article presents a convenient and time-saving new online platform for the deaf designed for web and Android applications. The utilization is used as an effective means of communication and learning. This model has a scientific 4-level function. Receive speech with PyAudio, convert speech to text using Google Speech to textual content API (text content ionization and process text content using NLP ideas), store text into visual symbols, phrases, and a series of processed text content for the deaf and deaf Matching video when displayed to male or female. [3] The system proposed in this paper is designed to recognize several

very important sign language elements and convert them into text content and sounds. American Sign Language is a visible language. Along with gestures, thoughts use language statistics, both figuratively and predictably. Sign language is not a common language in all states. Has its own sign language 6 and has a native language in the area For example, as a globally spoken language, multiple languages are spoken everywhere and for checking grammatical accuracy, detection fees in ASL languages are typically 90% for establishments that use Indian Sign Language. In this paper, there is an analysis of different methods and their efficiency. The best method of the analysis I then used and produced an android application that converts the Sign language to speech and text conversion.[5] This paper suggests a method in which speech is converted to sign using Microsoft Xbox Kinect 360s and unity 3d models to communicate. This first speech is taken as input this is then converted to text using Google API. Then this input is fed to models which show the signs.[6] this paper will give us another method that will be used to convert sign to speech conversion using CNN.[7] In this paper, The sign to speech is converted using an SVM classifier. In this method, different sensors are used such as flex sensors, MPU6050 sensor, Arduino Nano, and HCO5 Bluetooth module to detect the hand gesture to speech. In this accuracy of the system is 98.91% is achieved for the ASL database with 25% test data and 75% training data.[8] In this paper, there is a comparison between many approaches to object identification, here in this paper YOLO v3, R-CNN, and SSD are compared. In this comparison the R-CNN has high MAP than any other method, this makes it a key component to use in this project. YOLO has a high detection rate but R-CNN has a high MAP which is a key feature.

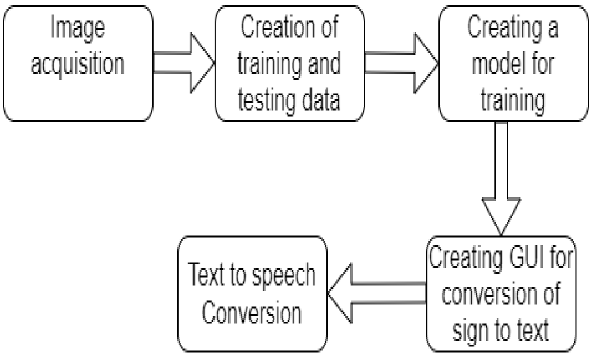
3. DESCRIPTION OF METHOD

SIGN TO SPEECH CONVERSION

First Process is to detect the sign language and convert it to text. The main process here is: 1) set up a UI to take his or her input from the user 2) Next is to test and train the data obtained from the user 3) form words and 4) conversion of text to speech.

A. Flow Diagram

This flow diagram explains the methodology of the project which is being implemented



Flow chart of the project

1) Image acquisition

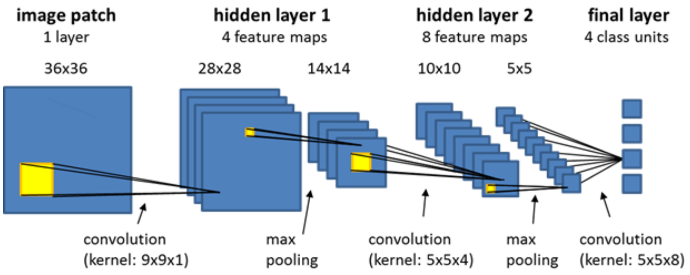
Image acquisition collects images linked to various sign languages. This contains all the alphabets, numbers, and other sign languages, which are all gathered and saved in a file. Approximately 600 to 700 photographs are shot for each sign.

2) Creation of testing and training data

Once the image has been captured, it must be separated into training and testing data. The image is gathered initially from the Region of Interest. These photos are then subjected to image processing to improve their accuracy.

3) Creating a model for training

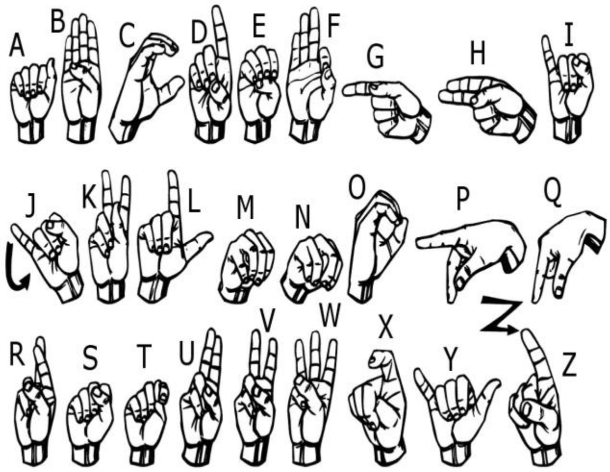
According to normal Neural Networks, there are three dimensions in neurons at CNN layers that are width, height, and depth. The neurons are not connected in a fully linked fashion but instead connected in the window size layer before it. By the conclusion of the CNN architecture, the output will be of several classes as we condense the full image into one vector of class scores. For creating the models, we use a convolutional layer for the input layer, and then further when we process two layers come into action that is hidden the first layer as well consists of 4 feature maps and the other one consists of 8 feature maps and the process continues further until we reach the final layer which contains the neurons that are like the images that are formed into more classes.



5) Creating GUI for conversion of sign to text

During the training, we turned our RGB input photos to grayscale and removed unwanted noise with gaussian blur. Then the captured image of the hand is resized to 128/128 pixels. We then train and test the data model. The image obtained is classified and the result between 0 to 1 is obtained. The sum of each such classified value equals 1. We achieved this result using the SoftMax function. As the prediction layer's results will sometimes not get in a clear manner it is improved by using labeled data for the training of the models. For the classification purpose, we make use of cross-entropy. It is a continuous function that is positive when the value is not the same as the labeled value and zeroes when the value is the same as the labeled value.

As a result, we increased the cross-entropy as close to zero as feasible to maximize the cross-entropy. I modified the weights of my neural network in the network layer to get this result.



6) Text to speech conversion

After that, the text is transformed into speech using the Google API. This is presented as an option because this is a two-way communication system where the user can choose between text and speech conversion.

4. Mathematics behind CNN

In CNN the convolution layer is the very important layer. It has the main proportion of the computational load. This layer performs a dot product between two matrices.

If we have an input of size W x W x D and a Dout number of kernels with a spatial dimension of F, stride S, and padding amount P, we can calculate the size of the output volume using the formula:

$$W_{out} = \frac{W - F + 2P}{S} + 1$$

The pooling layer uses a summary statistic of neighbouring outputs to replace the network's output at specific spots.

If we have an activation map of size W x W x D, a pooling kernel of spatial size F, and stride S, we can calculate the size of the output volume using the formula:

$$W_{out} = \frac{W - F}{S} + 1$$

IV. Results

The model is running with an efficiency of 95%. Hence this model proves to be very effective with the training and testing data can also be implemented for Indian sign language. This model gives us an effective way of sign languages. This model has more training and testing data sets than other papers mentioned which makes this model give more efficiency. This system can also be used to detect both American and Indian sign language. This system makes it unique from other systems as this can be used to detect other sign languages other than alphabets. This system can be used for a variety of other uses other than sign language because of its efficiency.

						P	r	e	d	i	c	t	e	d		V	a	l	u	e	s					
		A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y
	A	147	0	0	0	0	0	0	0	0	0	0	0	1	2	0	0	0	0	0	0	0	0	2	0	0
	B	0	139	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	11	0	0	0
	C	0	0	152	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	D	0	0	0	145	0	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	E	0	0	0	0	152	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	F	0	0	0	0	0	135	0	0	0	0	0	4	0	0	0	0	0	1	0	0	2	10	0	0	0
C	G	0	0	0	0	0	0	150	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
o	H	1	0	0	0	0	0	7	143	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1
r	I	0	0	0	33	0	0	0	0	108	0	2	0	0	0	0	0	0	0	0	7	1	0	0	0	0
r	J	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
e	K	0	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0	0	0	0	0	0
c	L	0	0	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0	0	0	0	0
t	M	0	0	0	0	0	0	0	0	0	0	2	0	152	0	0	0	0	0	0	0	0	0	0	0	0
	N	0	0	0	0	0	0	0	0	0	0	0	0	0	152	0	0	0	0	0	0	0	0	0	0	0
V	O	0	0	0	0	0	0	0	0	0	0	0	0	0	0	154	0	0	0	0	0	0	0	0	0	0
a	P	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	153	0	0	0	0	0	0	0	0	0
l	Q	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	147	1	0	0	0	0	0	0	0
u	R	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	150	0	0	0	0	0	0	0
e	S	0	0	0	0	1	0	0	0	0	0	0	0	0	1	10	0	0	0	132	0	0	0	0	8	0
s	T	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	151	0	0	0	0	0
	U	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	35	0	0	115	0	0	0	0
	V	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	151	1	0	0
	W	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	149	0	0
	X	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	148	0
	Y	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	151
	Z	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Algo 1

5. Conclusion

This application helps the dumb and deaf to communicate with others very effectively. As this is a two-way communication this will allow effectively communicate with others. This model is having 95 % efficiency which makes the model very effective, and the UI is also good where the user can communicate with. This can also be used for Indian sign language if we train with a lot of amounts of training and testing datasets. This system can be used for a variety of other use cases.

6. References

1] Shweta S. Shinde, Rajesh M. Autee, Vitthal K. Bhosale- Real Time Two-Way Communication Approach for Hearing Impaired and Dumb
Person Based on Image Processing. IEEE, VOL1, 2016, 1-5.

[2] Ritika Bharti, Sarthak Yadav. Sourav Gupta, Rajith Bakthula-Automated Speech to Sign language Conversion using Google API and NLP, ICAEEC. Vol 1, 2019. 1-9.

[3] Ankit Ojha, Ayush Pandey, Shubham Maurya, Abhishek Thakur, Dr. Dayanada P Sign Language to Text and Speech Translation in Real time using Convolution Neural network IJERT NCAIT, VOLI, 2020, 1-6

[4] Kohsheen Tiku, Aishwarya Ramesh, Jayshree Maloo, Indra R Real-time Conversion of Sign Language to Text and Speech.
IEEE, VOL1,2020,1-6

[5] Pankaj Sonawane, Karan Shah, Parth Patel, Shikhar Shah, Jay Shah Speech To Indian Sign Language (ISL) Translation System. ICCCIS.VOL1,2021,1-5

[6] Aishwarya Sharma, Dr. Siba Panda, Prof. Saurav Verma Sign
Language to Speech Translation IEEE 2020,1-8

[7] Malli Mahesh Chandra, Rajkumar S, Lakshmi Sutha Kumar Sign Languages to Speech Conversion Prototype using the SVM Classifier JEEE.VOL1,2019,1-5

[8] Lu Tan, Tianran Huangfu, Liyao Wu, Wenying ChenComparison of YOLO v3, Faster R-CNN and SSD for real Time Pill Identification.