[[1]](#footnote-1)

AUTOMATED SIGN TO SPEECH CONVERTION

DEBABRATA SWAIN , DIVSEHAJ SINGH ANAND , ARYAN KUMAR GUPTA , SHREYAS KHARE , DHEERAJ SHARMA , PRACHI TAPADIYA

DEPARTMENT OF INFORMATION TECHNOLOGY

*Abstract —* ***Unable to communicate verbally is a disability. So in order to communicate there are may ways ,one of the most popular method is use of predefined sign languages . The purpose of this project is to contribute recognizing American sign languages(ASL) to the field of automatic sign language recognition with maximum efficiency.This paper basically focuses on the recognition of ASL, the real time static gestures are collected from Laptop Webcam. The most challenging part in the design of an automatic sign language translator is the design of a good classifier that can classify the input static gestures with high accuracy. CNN architecture is used to design of classifier for sign languages recognition, in the proposed system . The system trained CNNs for the classification of twenty five(5) alphabets using 150-200 images. The system has trained the classifier with different parameter configurations and tabulated the results. Compared to previous literature the proposed work attained an efficiency of 99.35% for our classifier .The result shows that accuracy improves as we include more data from different subjects during training.=***

***Keywords*** - ***convolutional neural network, deep learning, gesture recognition, sign language recognition***

# INTRODUCTION

The speedy progress being achieved in the sector of information technology leads computer systems to play an important role in the every day life of human beings. Human computer interactions(HCI) is one of the main contributors towards the progress of the mankind and country in the current fast-moving world. The means of interaction ,are formed by the input devices such as, keyboards, mouse, joysticks, etc. Which also acts like the typical HCI modules. Since they limit the naturalness and speed of human-computer- interactions, Sign Language recognition system has gained a lot of importance. Convolutional Neural Network has been used in order to implement a prototype of the sign recognition model and is tested successfully. The primary goal of the system is to identifying a specific Sign Language and using it to convey information. The architecture, results and analysis are discussed in detail in this paper .

Sign language is a language that provides visual communication and allows individuals with hearing or speech impairments to communicate with each other or with other individuals in the community. According to the World Health Organization, the number of hearing-impaired individuals has recently reached 400 million. For this reason, recent studies have been accelerated to make disabled people communicate more easy.Deaf-mute people communicate by sign languages. Not everyone understands sign languages. It is possible only for those who have undergone special training to understand the language. The alternative way of communication , which is written communication is not convenient always.

Motion of any body part like face, hand is a form of gesture. Gesture recognition enables computer to understand human actions and also acts as an interpreter between computer and human. The automatic sign language recognition helps to understand the meaning of different signs without any need of expert person.

In our project, the feature extraction is automated by using convolutional neural networks (CNNs).

Fig.1 shows the basic sign language:

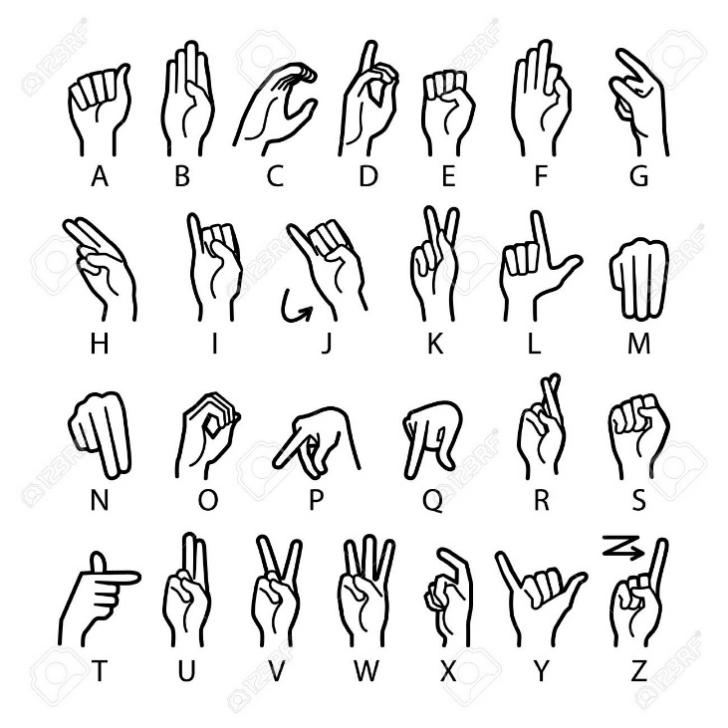


Fig.1 ASL finger alphabets

# Literature Review

Byeongkeun et al. using Convolutional Neural Networks [1] proposed a run time sign language finger. American Sign Language is specific but an essential pa rt of sign language recognition. Even though it used depth sensors which enable them to capture additional information to improve accuracy and processing time, their cache architecture is very complicated . A method for implementing a sign language to text/voice conversion system without using hand-held gloves and sensors ,by capturing the gesture continuously and converting them to voice. In this method only few images were captured for recognition. The design of a communication aid for physically challenged [2] has been created as a prototype.

In [19] an American Language recognized by system has been presented with a 30 words vocabulary. An appearance-based representations and hand tracking system was constructed that was classified by hidden Markov model (HMM). The RWTH-BOSTON-50 database has an error rate of 10.91%.

Another work related to this field was creating sign language recognition system by using pattern matching [5]. The main aim of this proposed work has been to create a system which will work on sign language recognition. Many researchers have already introduced about many various sign language recognition systems and have implemented using different techniques and methods. This proposed system is focusing on an approach which is to put the SLR system which will work on Signs as well as Text (which will be understandable by deaf and dumb persons and also by normal persons). The main task will be performed in two ways by the system. It will take input by the user in the form of text which will be then perform matching with the sign and vice-versa.

Sruthi Upendran [3] and *et.al* introduced “American Sign

Language Interpreter System for Deaf and Dumb Individuals”. Out of 24 static alphabets 20 alphabets could be implemented in the procedure. Due to occlusion problem, the letters A,M,N and S couldn’t be recognized. The number of images used were limited.

The same can be implemented using an optimized approach by implementing the famous algorithm with LBP feature i.e. Viola Jones to recognize hand gestures in real time environment. Using this algorithm, we created Indian sign language interpreter with android implementation [4]. The advantage of this approach is that it takes less computational power to detect the gestures.

Chenyang Zhang, Yingli Tian [5] *et.al* “multi-modality

American Sign Language recognition” .The main enlightened

features of the system is twofold: 1) the multiple

signal modalities like sequence of depth images,

RGB image-based hand shapes, facial expression attributes

and key point detections for both body joints and facial

landmarks were considered. 2)signing sequences were performed

by learning fluent ASL signers and annotations which were provided by

professional linguisticians, our system can recognize different

components such as English words and special ASL grammar

components, such as facial expressions or head movements that have grammatical meaning during sentences.

# Methodology/Experimental

## DATASET COLLECTION

It is a very crucial stage in the research. This a sole stage on which the training phase and the testing phase all are interdependent. Here, as our model has the ability to sense and detect 5 alphabets namely A,B,C,D,E. So,we have proceeded by collecting 400 images of each alphabets. Thus in total we have 2000 images out of which we have 300 images to the testing part equaling to the 15% and the rest 1700 images have been sent to the training part equaling to the 85% of the total images. Hence, these training part again has actual training and the validation split training part. So, in the actual training part we have 1500 images and, in the validation, split training part we have 170 images in total for it.

A Convolutional Neural Network (CNN**)** is a Deep Learning algorithm which can take in an input image, assign importance to various aspects/objects in the image and be able to differentiate one from the other. A CNN is a neural network that typically contains several types of layers, which are **convolutional layer**, **pooling layer** , relu layer **and fully connected**layer.

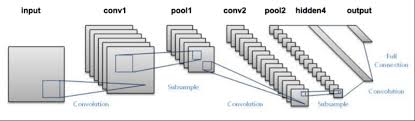


Fig 2 : Flow of CNN.

## Convolutional Layer

The convolutional layer is also known as filtering layer. As we take a image we effectively check for patterns in that section of the image. This works because of **filters,**stacks of weights represented as a vector, which are multiplied by the output values by the convolution.The content of image is predicted from the combinations of these high weights from various filters.

## ReLu Layer

ReLu is a abbreviation of Rectified Linear Unit. It is among one of the activation functions that CNN system possess.It overcomes the vanishing gradient problem, allowing model to learn faster and perform better.

Sigmoid Function

The sigmoid function is a activation function which is also known as squashing function. This function limits the output value between 0 and 1, which is useful in the prediction of probabilities. The following fig 4 shows the graph of sigmoid function. A screenshot of a cell phone

Description automatically generated

Fig 3 : Sigmoid function.

## Pooling Layer

Pooling layer is used to reduce the spatial size of convolved feature. Through dimensionality reduction the computational power required to process data decreases. There are two types of pooling viz. Max pooling and Average pooling.

Max pooling: Maximum pooling returns the maximum value from the portion of image covered by the kernel.

Average pooling: It returns the average of all the values from the portion of the image covered by the kernel.

## Fully Connected Layer

Now that we have converted our input image into a suitable form for our Multi-Level Perceptron, we shall flatten the image into a column vector. The flattened output is fed to a feed-forward neural network and backpropagation applied to every iteration of training.

## Training

We trained the model by varying the hidden layers from 1 to 4 and by varying the epochs 100 in each case. The number of nodes in each hidden layer also varies. Out of the 400 samples of each image we used 300 images for creating the training set, 60 for validation and 40 for testing. The data sets were pickled and were used for training. Given flowchart shows the training process. After the model was trained the net parameters were saved so that it can be used in the testing phase for testing the accuracy of the model and also for classification of the input symbols. A picture containing text, map

Description automatically generated

Fig 4: Training Phase Workflow.

## G. Testing

In the testing phase the accuracy of our trained model is

tested. The saved network parameter is loaded to test the dataset and determined the accuracy. The method used for all test cases and tabulated the obtained accuracies. The model allows the user to select the images of the sign language static gestures that need to be classified. The application used our trained CNN network to classify these symbols and produce their corresponding labels. The labels are then turned to their corresponding alphabets/numbers. The words/numbers are then spoken out using the python text to speech module.

# Results and Discussions

The model successfully recognizes the 26 English Alphabets and converts it into speech with a testing accuracy of 99.51%.

# Limitations

In the project we have not predicted the numbers from 0

to 9. Further the model fails in recognizing sentences and phrases. It is capable to take only sign at a time.The model also fails in detecting some expressions.

# Future Scope

In this project we have converted some alphabets and real time gestures into speech. Further by concatenating those alphabets we can convert phrases and sentences into speech.

By using high end processor we can have more number of data sets to increase the training and testing accuracy of the model.

# Conclusion

The main idea behind this project is to overcome the barrier of communication between Deaf-mute people and hearing majority. Using CNN we recognized different alphabets of a sign language and some real time gestures and converted it into text. Further this text is converted into speech. The

model allows users to select the static sign gestures as input and it will speak out the letters.

References

[1]

<https://www.researchgate.net › publication › 297757591_Sign_Language_R...>

[2]

<https://www.microsoft.com › uploads › 2016/02 › CNN_ASLPTrans2-14>

[3]

<https://www.intechopen.com › books › convolutional-neural-networks-for-...>

[4]

<https://www.researchgate.net › post › Where_can_I_find_a_handwritten_c...>

1. [↑](#footnote-ref-1)