SIGN LANGUAGE RECOGNITION SYSTEM

MINI PROJECT REPORT

Submitted in partial fulfilment of the requirements for the award of the degree
Of
BACHELOR OF TECHNOLOGY
In
INFORMATION TECHNOLOGY
By

Vaibhav Khera	Ayush Rastogi	Dinesh Kumar
35651203117	40151203117	352512013117

Guided by

Mrs. Shafali Dhall

Assistant Professor

Department of Information Technology
BHARATI VIDYAPEETH'S COLLEGE OF ENGINEERING
PASCHIM VIHAR, NEW DELHI

April 2020

CANDIDATE'S DECLARATION

It is hereby certified that the work which is being presented in the B. Tech Mini Project Report entitled **" SIGN LANGUAGE RECOGNITION SYSTEM"** in partial fulfilment of the requirements for the award of the degree of **Bachelor of Technology** and submitted in the **Department of Information Technology** of **Bharati Vidyapeeth's College of Engineering, New Delhi (Affiliated to Guru Gobind Singh Indraprastha University, Delhi)** is an authentic record of our own work carried out during a period from **January 2020 to April 2020** under the guidance of **Mrs. Shafali Dhall, Assistant professor.**

The matter presented in the B. Tech Major Project Report has not been submitted by me for the award of any other degree of this or any other Institute.

Vaibhav Khera	**Ayush Rastogi**	**Dinesh Kumar**
35651203117	**40135251203117**	**35251203117**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge. They are permitted to appear in the External Major Project Examination

Mrs. Shafali Dhall , Assistant Professor

The B. Tech Minor Project Viva-Voce Examination of **Vaibhav Khera (35651203117), Ayush Rastogi (40151203117) And Dinesh Kumar(35251203117),** has been held on**…….**

(Signature of External Examiner)

Mr. Arun Dubey
Project Coordinator

ABSTRACT

Gesture based communication is centre correspondence media to the general population which can't talk. It's anything but an all-inclusive language implies each nation has its very own gesture based communication. Each nation has its own punctuation for their communication through signing, word requests and articulation. The issues emerge when individuals endeavour to impart utilizing their language with the general population who are unconscious of this language sentence structure. Developing sign language Recognition System for deaf people can be very important, as they'll be able to communicate easily with even those who don't understand sign language.[1][2]

The gestures that have been translated include alphabets, words from static images. This becomes more important for the people who completely rely on the gestural sign language for communication and tries to communicate with a person who does not understand the sign language.

We aim at representing features which will be learned by a technique known as convolutional neural networks (CNN), containing four types of layers: convolution layers, pooling/subsampling layers, nonlinear layers, and fully connected layers. The new representation is expected to capture various image features and complex non-linear feature interactions.

ACKNOWLEDGEMENT

We express our deep gratitude to **Mrs, Shafali Dhall**, Designation, Department of Information Technology for his valuable guidance and suggestion throughout my project work. We are thankful to **Mr. Arun Dubey ,** Project Coordinator, for their valuable guidance.

We would like to extend my sincere thanks to **Head of the Department, Mrs. Vanita Jain** for his time to time suggestions to complete my project work. I am also thankful to **Dr. Dharmender Saini, Principal** for providing me the facilities to carry out my project work.

Vaibhav Khera	**Ayush Rastogi**	**Dinesh Kumar**
35651203117	**40151203117**	**35251203117**

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATION

SL – Sign Language

FC – Fully Connected

ASL – American Sign Language

CNN – Convolution Neural Network

CSV -Comma Separated Value

CHAPTER 1: INTRODUCTION

1.1 INTRODUCTION

Deaf is a disability that impair their hearing and make them unable to hear, while mute is a disability that impair their speaking and make them unable to speak. Both are only disabled at their hearing and/or speaking, therefore can still do much other things. The only thing that separate them and the normal people is communication. And the only way for them to communicate is through sign language. Sign language is not an universal language, and different sign languages are used in different countries, like the many spoken languages all over the world. Some countries such as Belgium, the UK, the USA or India may have more than one sign language. Sign Language (SL) substantially facilitates communication in the deaf community. Sign language is a visual language and consists of 3 major components: Finger-spelling (used to spell words letter by letter), Word level sign vocabulary (used for the majority of communication), Non-manual features facial expressions and tongue, mouth and body position [1].
Research for communication via gestures acknowledgment was begun in the '90s. Hand signal related research can be isolated into two classifications. One depends on electromagnetic gloves and sensors which decides hand shape, developments and introduction of the hand. Be that as it may, it is expensive and not appropriate for down to earth use [2]. Researchers want to find a way for the deaf-mute people so that they can communicate easily with normal person. The breakthrough for this is the Sign Language Recognition System. The system aims to recognize the sign language, and translate it to the local language via text or speech.

In order to diminish this obstacle and to enable dynamic communication, we present an SL recognition system that uses Convolutional Neural Networks (CNN) in real time to translate user's SL signs into text. With the use of Neural Networks in image processing the input image is compared with set of images in the dataset the word corresponds to the matched image will gives the output. The model is designed to recognise all non-dynamic gestures of American Sign Language with bare hands of different hand shapes and skin colours which makes it complex for model to correctly recognise the gesture [2].

1.2 AMERICAN SIGN LANGUAGE

American Sign Language is a complex visual-spatial language that is used by the deaf community in the United States and English-speaking parts of Canada. It is linguistically complete, natural language. It is native language of many deaf men and women, as well as some hearing children born into deaf families. It is the first language of the many North Americans United Nations agency are deaf and is one in all many communication choices utilized by those who are deaf or dumb [2].

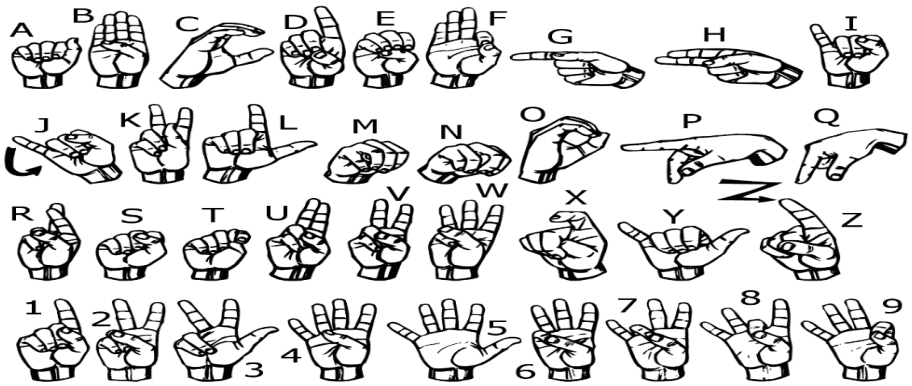

Figure No. 1.1 – list of Alphabet and number.

1.3 MOTIVATION

A good Sign Language Recognition (SLR) system can overcomes the barriers exists between the speech and hearing people with speaking society. The goal of SLR for developing systems and approaches for properly recognize the series of gestures and to know the meaning of the gestures .

SLR is a noticeable task due to its impact on humanoid society as the mute pupil facing huge communication gap with the speaking community. SLR is a challenging work because of its variation of hand gestures, facial expressions, body movements and many such variations and confines in this regard. A very less volume of work has been done in this lane to recognize the distance invariant, size invariant, rotation invariant, and race invariant ASL gestures with respect to background (plain and complex, uniform and non-uniform), location (indoor and outdoor), time (day and night), and light illumination (natural and artificial). Also, there is no noticeable work has been done predominantly for real time gesture recognition by considering various research disputes with respect to static and dynamic environment. There are huge amount of opportunities to carry out the research in recognizing ASL to make the communication easier between and mute and speaking community

1.4 OBJECTIVE

This project aims at identifying alphabets in Sign Language from the corresponding gestures.

- We aim to solve this problem using state of the art computer vision and machine learning algorithms.
- In addition to this, lack of datasets along with variance in sign language with locality has resulted in restrained efforts in Sign Language gesture detection.
- Our project aims at taking the basic step in bridging the communication gap between normal people and deaf and dumb people by developing a gesture recognition system using CNN to correctly interpret sign language.

1.5 CHALLENGES IN GESTURE RECOGNITION

- Gestures recognition involves complex processes such as motion modeling, motion analysis, pattern recognition and machine learning. It consists of methods with manual and non-manual parameters. The structure of environment such as background illumination and speed of movement affects the predictive ability. The difference in viewpoints causes the gesture to appear different in 2D space.
- In some research, signer wears wrist band or colored glove to aid the hand segmentation process, such as in [3].The use of colored gloves reduces the complexity of segmentation process.
- Several anticipated problems in a dynamic gesture recognition, includes temporal variance, spatial complexity, movement epenthesis, repeatability and connectivity as well as multiple attributes such as change of orientation and region of gesture carried out. There are several evaluation criteria to measure the performance of a gesture recognition system in overcoming the challenges. These criteria are scalability, robustness, real-time performance and user-independent.

1.6 RELATED WORK

Convolutional Neural Networks have been extremely successful in image recognition and classification problems, and have been successfully implemented for human gesture recognition in recent years. In particular, there has been work done in the realm of sign language recognition using deep CNNs, with input-recognition that is sensitive to more than just pixels of the images. With the use of cameras that sense depth and contour, the process is made much easier via developing characteristic depth and motion profiles for each sign language gesture [9].

The use of depth-sensing technology is quickly growing in popularity, and other tools have been incorporated into the process that have proven successful. Developments such as custom-designed color gloves have been used to facilitate the recognition process and make the feature extraction step more efficient by making certain gestural units easier to identify and classify [10].

Some neural networks have been used to tackle ASL translation [11]. Arguably, the most significant advantage of neural networks is that they learn the most important classification features. However, they require considerably more time and data to train. To date, most have been relatively shallow. Mekala et al. classified video of ASL letters into text using advanced feature extraction and a 3-layer Neural Network [11]. They extracted features in two categories: hand position and movement. Prior to ASL classification, they identify the presence and location of 6 “points of interest” in the hand: each of the fingertips and the center of the palm. Mekala et al. also take Fourier Transforms of the images and identify what section of the frame the hand is located in. While they claim to be able to correctly classify 100% of images with this framework, there is no mention of whether this result was achieved in the training, validation or test set.

In [12], new feature extraction techniques are proposed to recognition of static ASL signs of numbers 0 to 9 in plain background and obtained 74.69%, 82.92%, 87.94, and 98.17% of recognition rates using Statistical Measures Technique, Orientation Histogram Technique, COHST (Combined Orientation Histogram and Statistical Technique), and Wavelet Features Technique respectively. An Open-Finger Distance Feature Measurement and Neural Network Classification Technique is used to recognize the ASL Numbers in , and obtained 92.09% of recognition rate.

CHAPTER 2: PROPOSED METHOD

2.1 Approach:

There are basically two types of approaches for hand gesture recognition:

- Vision based approach
- Data glove approach

Here for Gesture recognition we are using image processing and computer vision.

- Gesture recognition enables computer to understand human actions and also acts as interpreter between computer and human. This could provide potential to human to interact naturally with the computers without any physical contact of the mechanical devices.
- Sign language can be performed by using Hand gestures either by one hand or two hands. It is of two type Isolated sign language and continuous sign language.
- Isolated sign language consists of a single gesture having a single word while continuous ISL or Continuous Sign language is a sequence of gestures that generate a meaningful sentence.

In this report we performed isolated ASL gesture recognition technique.
In this work, 36 different categories have been considered: 26 categories for English Alphabets (a-z) and 10 categories for Numerals (0-9).

The whole framework works in four stages :

- Image Acquisition: This is the initial step or procedure of the central strides of advanced picture handling
- Image Enhancement
- Image Restoration
- Color Image Processing
- Wavelets and MultiResolution Processing Compression [2]

2.2 Algorithmic Strategy :

- Convolutional Neural Network (CNN) works by consecutively modelling and small pieces of information and combining them deeper in work CNN follows Greedy approach In neural networks, Convolutional neural network (ConvNets or CNNs) is one of the main categories to do images recognition, images classifications. Objects detections, recognition faces etc., are some of the areas where CNNs are widely used.

- CNN image classifications takes an input image, process it and classify it under certain categories (Eg., Dog, Cat, Tiger, Lion). Computers sees an input image as array of pixels and it depends on the image resolution.

- Based on the image resolution, it will see h x w x d(h = Height, w = Width, d = Dimension). Eg., An image of 6 x 6 x 3 array of matrix of RGB (3 refers to RGB values) and an image of 4 x 4 x 1 array of matrix of grayscale image.

- Technically, deep learning CNN models to train and test, each input image will pass it through a series of convolution layers with filters (Kernals), Pooling, fully connected layers (FC) and apply Softmax function to classify an object with probabilistic values between 0 and 1.

- CNN works in following four stages:
    1. Convolution Layer
    2. ReLU Layer
    3. Pooling Layer
    4. Fully Connected Layer

2.3 Dataset:

The image dataset consists of ASL gestures. The dataset consists of 45500 images in 26 classes. Each character has 250 images in testing dataset and 1750 in training dataset. Each category represented a different character of ASL.. Out of this dataset 75% i.e. 39000 images were used for training and remaining 25% i.e. 6500 images were used for testing .

- We had an option of using the dataset as a Comma-separated values CSV or to generate the images from the pixel's values and then use them to train our dataset.

- We went with using the CSV file as it made the process of classification faster. All the missing values in the dataset were first handled and was made sure that there are no missing values in the dataset.

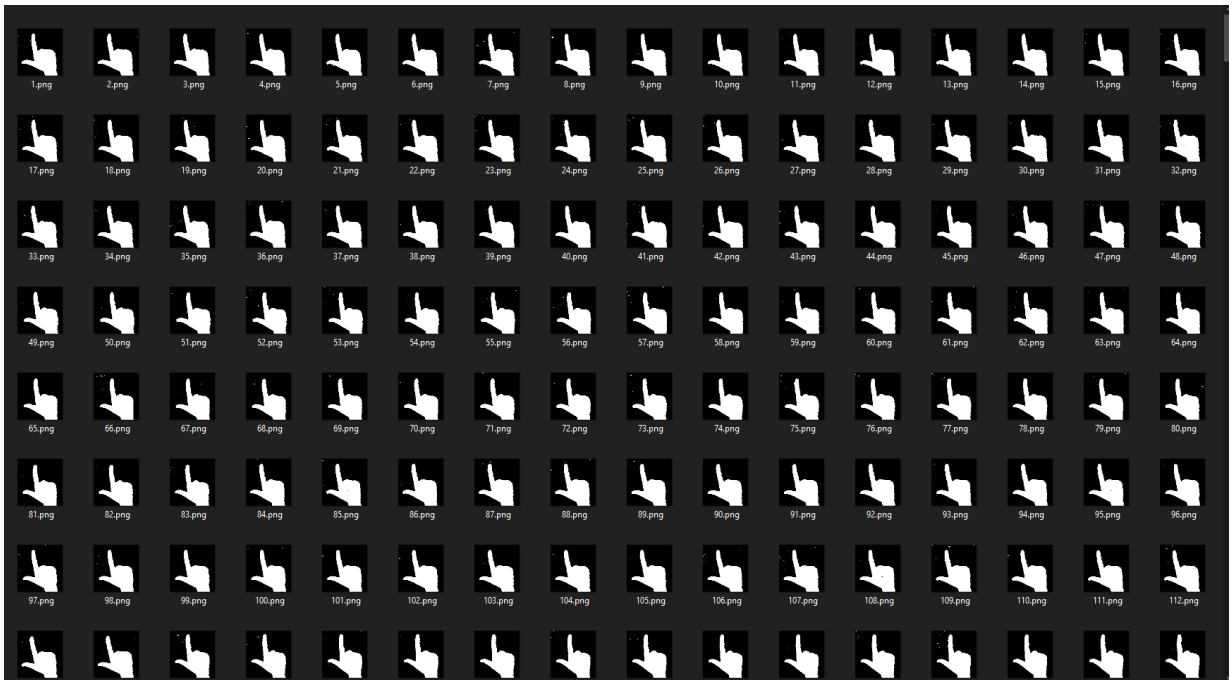

Figure No. 1.2 – Dataset of Alphabet L.

CHAPTER 3: PROCEDURE

The first step of sign language recognition system is to acquire the sign data. There can be various ways to get the data.

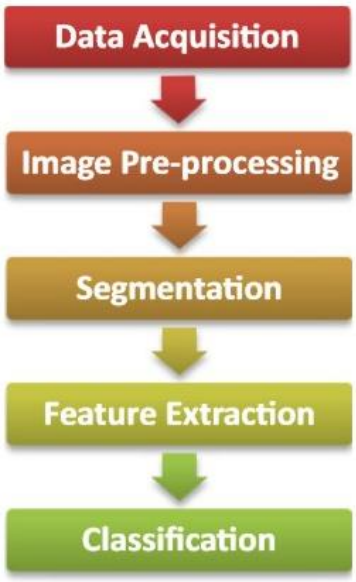

Figure No. 3.1 - Generalized block diagram of Image recognition system

- **Data Acquisition**: The process of capturing the photographic images, such as of a physical scene.

- **Data Pre Processing**: Goal of the pre-processing is an improvement of the input image data that suppresses unwanted noise or enhances image features important for further processing.

- **Feature Learning**: The Feature learning starts from the initial measured data to builds its derived features intendeds to be informative and facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations.

- **Classification**: Classification is the process related to categorization, the process in which ideas and objects are differentiated from others.

- **Recognition**: Focuses on the recognition of patterns and regularities in data. Pattern recognition is the process of classifying input data into objects or classes based on key features.[5]

3.1 Image Pre-processing

Main aim of pre-processing is an improvement of the image data that reduce unwanted deviation or enhances image features for further processing. The images in the data set were of a varying size and shape. Therefore the first step was to read and resize each of the image to the similar size of 224x224 pixel. Only when all of the images in the dataset are of the same size can the images be fed into a neural network for training.

Pre-processing is also referred as an attempt to capture the important pattern which express the uniqueness in data without noise or unwanted data which includes cropping, resizing and gray scaling.
The mean value of RGB over all pixels was subtracted from each pixel value. i.e in first pass the model will compute the mean pixel value of each channel over the entire set of pixels in a channel and in the second pass it will modify the images by subtracting the mean from each pixel value.

Subtracting the mean value from the pixels centers the data.

- The mean is subracted because the model involves multiplying weights and adding biases to the initial inputs to cause activations then back propagated with the gradients to train the model.

- It is important that each feature has a similar range, in order to prevent the gradients from getting out of control. Also CNN's involve sharing of parameters and if the inputs are not scaled to have similar ranged values sharing will not happen easily because one part of the image will have large value of weights while the other will end with smaller value.

3.1.1 Cropping

Cropping refers to the removal of the unwanted parts of an image to improve framing, accentuate subject matter or change aspect ratio.

3.1.2 Resizing

Images are resized to suit the space allocated or available. Resizing image are tips for keeping quality of original image. Changing the physical size affects the physical size but not the resolution.

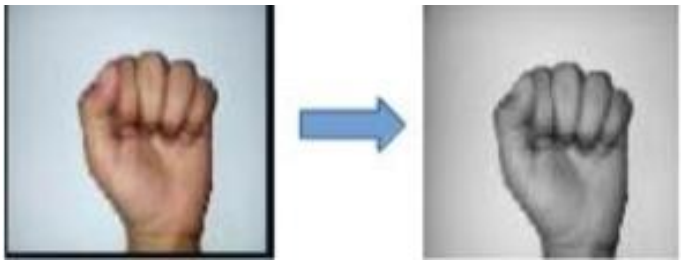

Figure No. 3.2 - Output of Pre-processing

3.2 Segmentation

Segmentation is the process of partitioning images into multiple distinct parts. It is a stage whereby the Region of Interest (ROI), is segmented from the remaining of the image. Segmentation method can be contextual or non-contextual. Contextual segmentation takes the spatial relationship between features into account, such as edge detection techniques. Whereas a non-contextual segmentation does not consider spatial relationship but group pixels based on global attributes.[4]

3.2.1 Skin color segmentation

Skin color segmentation are mostly performed in RGB, YCbCr, HSV and HSI color spaces . Several challenges toward achieving a robust skin color segmentation is sensitivity to illumination, camera characteristic and skin color . HSV color space is popular as the Hue of palm and arm differs greatly, hence palm can be segmented from the arm easily , it segments the face and hand in HSV color space. Research found that YCbCr is more robust for skin color segmentation compared to HSV in different illumination condition

3.3 Convolutional Neural Network Model

Convolutional Neural Network Model was proposed by LeCun, and has made a breakthrough in the field of image classification and target detection.

A convolutional neural network (CNN, or ConvNet) is a type of feed-forward artificial neural network in which the connectivity pattern between its neurons is inspired by the organization of the animal visual cortex.

Deep CNN's introduce a large number of hidden layers, thus reducing the dimensionality of the image and enabling the model to extract sparse image features in low dimensional space. [1]

- CNNs have repetitive blocks of neurons that are applied across space (for images) or time (for audio signals etc). For images, these blocks of neurons can be interpreted as 2D convolutional kernels, repeatedly applied over each patch of the image.
- For speech, they can be seen as the 1D convolutional kernels applied across time-windows. At training time, the weights for these repeated blocks are 'shared', i.e. the weight gradients learned over various image patches are averaged

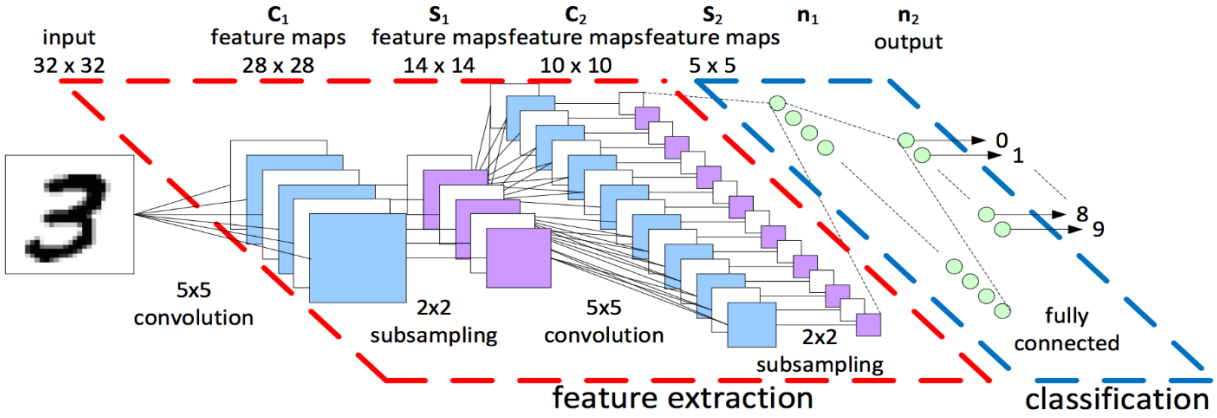

Figure No. 3.3 - Classification

There are four main steps in CNN: convolution, subsampling, activation and full connectedness.

3.3.1 Convolutional Layer :

The convolution is a special operation that extracts different features of the input. The first it extracts low-level features like edges and corners. Then higher-level layers extract higher-level features. For the process of 3D convolution in CNNs. The input is of size N x N x D and is convolved with the H kernels, each of them sized to k x k x D separately.

If the input signal looks like previous cat images it has seen before, the “cat” reference signal will be mixed into, or convolved with, the input signal. The resulting output signal is then passed on to the next layer.

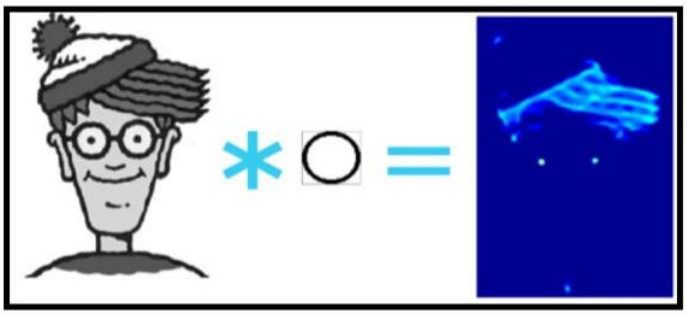

Figure No. 3.4 - Convolving Wally with a circle filter. The circle filter responds strongly to the eyes

Convolution of one input with one kernel produces one output feature, and with H kernels independently produces H features respectively. Starts from top-left corner of the input, each kernel is moved from left to right. Once the topright corner reached, kernel is moved one element downward, and once again the kernel is moved from left to right, one element at a time. Process is done continuously until the kernel reaches the bottom-right corner.

3.3.2 Sub-Sampling Layers:

This layer reduces the resolution of features. The features are robust against noise and distortion. Inputs from the convolution layer can be “smoothened” to reduce the sensitivity of the filters to noise and variations . This smoothing process is called subsampling.

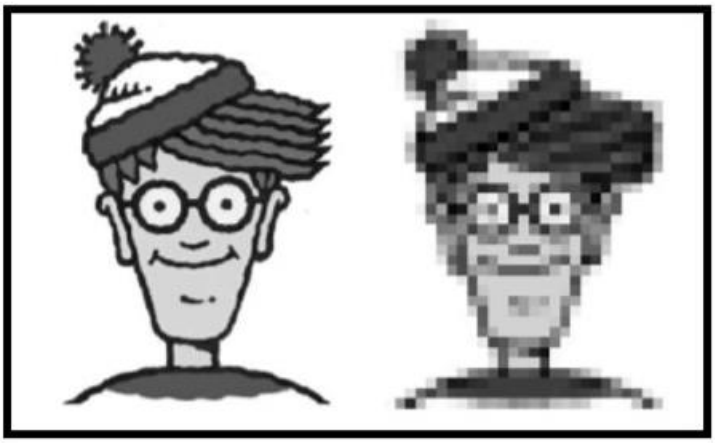

Figure No. 3.5 - Sub sampling Wally by 10 times. This creates a lower resolution image.

3.3.3 Pooling

- A pooling layer is another building block of a CNN.
- There exists two ways to do pooling: 1.max pooling and 2.average pooling. For both cases, the input is divided into non-overlapping dimensional spaces. For average pooling, the average of the given values in the region are calculated. For max pooling, the maximum value of the given values is selected.
- Its function is to progressively reduce the spatial size of the representation to reduce the amount of parameters and computation in the network. Pooling layer operates on each feature map independently.

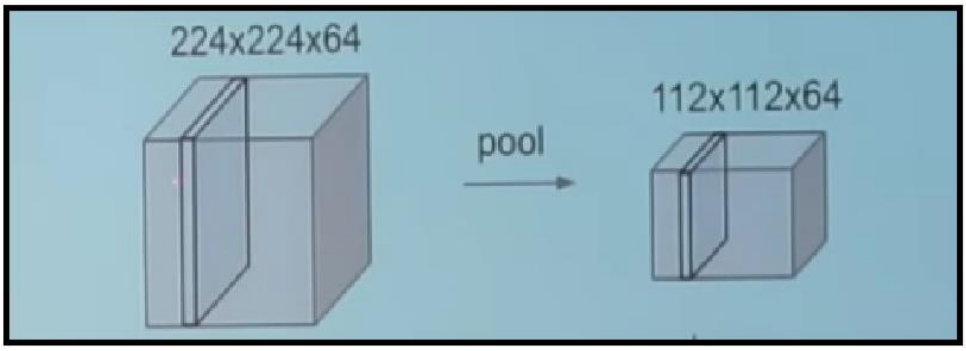

Figure No. 3.6 - Pooling to reduce size from 224x224 to 112x112

The most common approach used in pooling is max pooling in which maximum of a region taken as its representative. For example in the following diagram a 2x2 region is replaced by the maximum value in it.

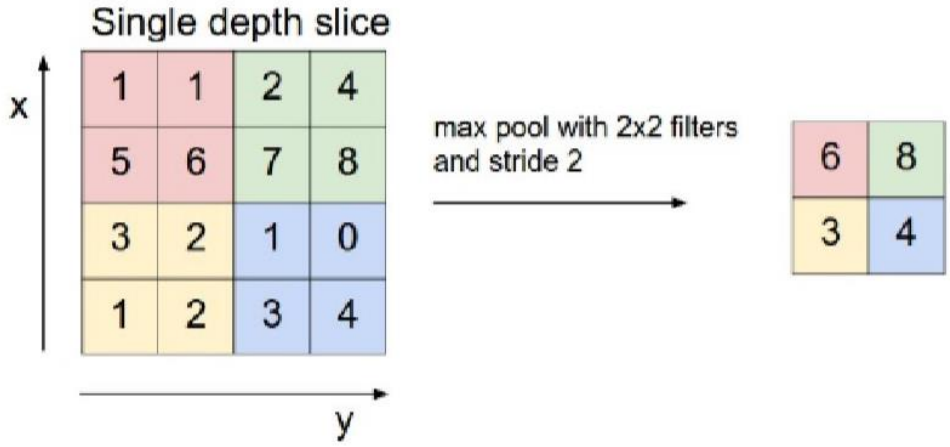

Figure No. 3.7 - Max Pooling

3.3.4 Activation

The activation layer controls how the signal flows from one layer to the next, emulating how neurons are fired in our brain. Output signals which are strongly associated with past references would activate more neurons, enabling signals to be propagated more efficiently for identification.

CNNs in particular rely on a non-linear —trigger function to signal distinct finding of likely features on each and every hidden layers. CNNs may use the variety of specific functions like ReLUs (rectified linear units) and continuous trigger functions to efficiently implement this non-linear triggering.

3.3.5 Fully Connected

The last layers in the network are fully connected, meaning that neurons of preceding layers are connected to every neuron in subsequent layers. This mimics high level reasoning where all possible pathways from the input to output are considered.

Fully connected layers are always used as the final layer. These layers are mathematically sums the weigh of previous layer of features, indicating the precise mix of —ingredients to determined specific target result. All the elements of all the features of the previous layer get used in the mathematical calculation of each element of the each output feature.

- **In training sequence**, the workflow of proposed system is as follows:

• The reference image is extracted and the individual image is pre-processed. In this pre-processing, filters are applied so as to enhance the useful content of the frame information and to reduce the unwanted information as much as possible.

• All the features of processed image are then extracted using CNN and stored in database associated with that sign.

• The same procedure is followed for all the sign to be included in the system. And the complete reference database is prepared.

➢ **Test Phase** The workflow of testing sequence can be outlined as below:

• The input image is extracted.

• The images are pre-processed in order to get the Processed image as similar to that in training sequence.

• After obtaining the Processed image in testing, the images are matched with the previously maintained database.

• The difference between them is measured depending feature. On finding the nearest\ match, the image is recognized as that particular sign and corresponding output is flashed on the screen

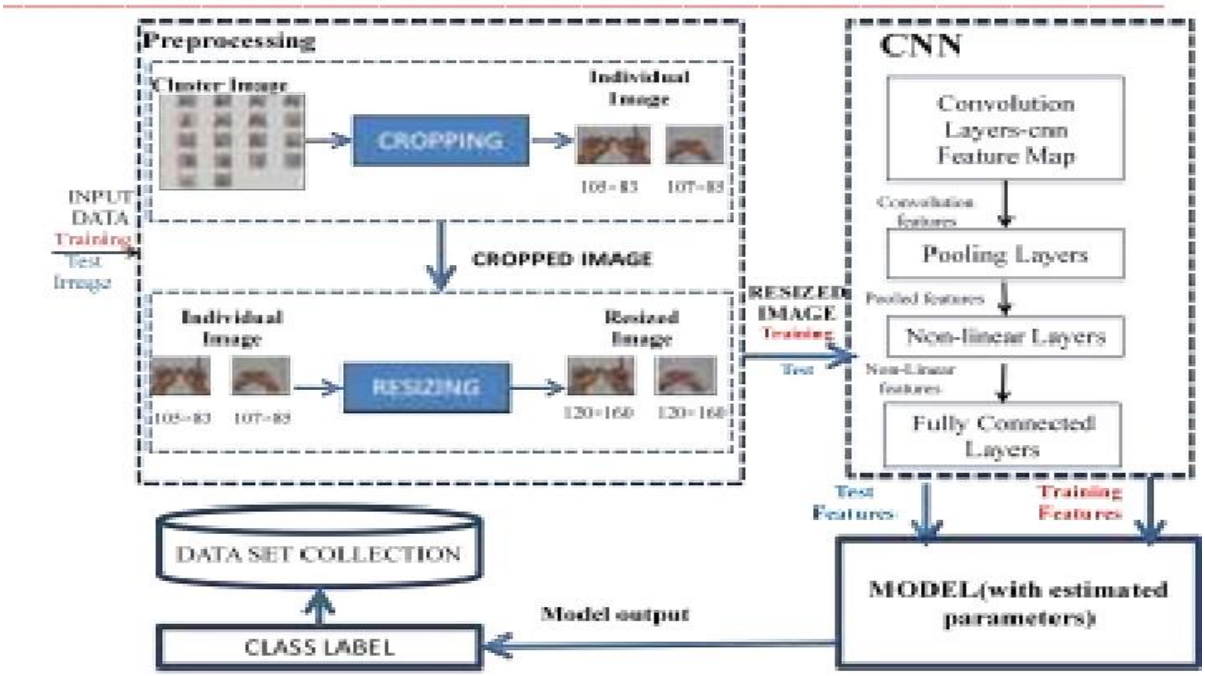

Figure No. 3.8 - Model Architecture

3.5 VGG16 Model

VGG 16[7] model is of deep convolutional neural network model proposed by K. Simonyan and A. Zisserman in their work [7]. The model was able to achieve 92.7% top-5 test accuracy in ImageNet.

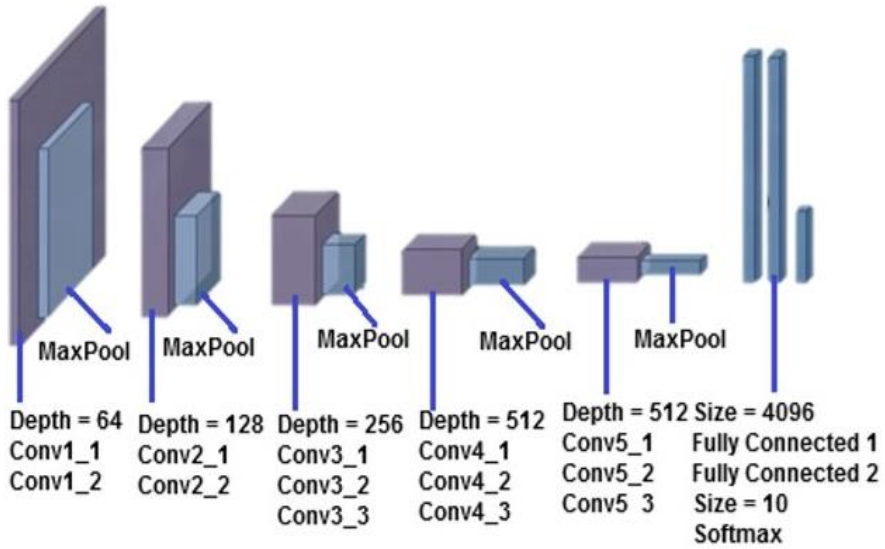

Figure No. 3.9 - VGG16 Architecture

- The macro architecture of VGG16 [7] can be seen in Fig 4. The input to the convolutional neural network during training is a fixed size 224 x 224 RGB image. The only pre processing we do in the training step is to subtract mean value of each channel (red, blue, green channels) which was computed on the training set, from each pixel. We pass each image through a stack of convolutional layers and each layer uses a very small receptive field of size 3 x 3. The convolution stride of 1 pixel is used. The spatial padding of convolutional layer is selected such that the resolution is preserved after convolution. The process of spatial pooling is carried out by max pooling layers which follow some of the convolutional layers.

- Three Fully-connected (FC) layer follows a stack of convolutional layers. The first two layers have 4096 channels and the third performs 1000-way ILSVRC classification of the input image. Finally, the last layer in FC layers is soft-max layer. All hidden layers are equipped with the rectification (ReLU) nonlinearity.

- Pre trained weights was obtained by training the VGG 16 model on the ImageNet database which contains over million on images. This was done so as to initialize the weights of the model. The original model contained 1000 channels in the soft-max channel i.e. it was designed to classify 1000 categories but for our purpose we needed only 36 categories (26 categories for alphabets and 10 for numerals). Therefore we deleted the last layer from the model and inserted a layer which would be able to categorize 36 different types of images. The rest of the model remained unchanged.

- The training of the model was done using stochastic gradient descent [8] with momentum. Batch size of 128 images and momentum of 0.9 was used. The learning and decay rates were initialized to 0.001 and 10-6 respectively.

CHAPTER 4: RESULTS AND DISCUSSION

4.1 Result

The samples were tested on VGG16. The Testing and training accuracies are tabulated below for 4 epochs**.** Overall, out of the 3696 images used for testing 3531 images were classified in to correct categories and the remaining 165 images were misclassified resulting in an average accuracy of 95.54 percent.

Characters	Total Sample	Correct Predictions	Incorrect Predictions	Percentage Accuracy
0	109	59	50	54.13
1	91	87	4	95.6
2	101	91	10	90.1
3	106	105	1	99.06
4	106	96	10	90.57
5	107	104	3	97.2
6	101	93	8	92.08
7	99	99	0	100
8	101	99	2	98.02
9	96	95	1	98.96
A	107	106	1	99.07
B	111	111	0	100
C	105	105	0	100
D	93	92	1	98.92
E	83	83	0	100
F	103	103	0	100
G	105	105	0	100
H	101	101	0	100
I	113	108	5	95.58
J	107	105	2	98.13
K	103	102	1	99.03
L	110	110	0	100
M	91	91	0	100
N	121	109	12	90.08
O	111	107	4	96.4
P	100	100	0	100
Q	104	104	0	100
R	102	98	4	96.08
S	100	100	0	100
T	84	83	1	98.81
U	99	98	1	98.99
V	110	103	7	93.64
W	97	66	31	68.04
X	93	90	3	96.77
Y	114	114	0	100
Z	112	109	3	97.32

Table No. 4.1 – Results obtained for individual accuracy of alphabets and numerals

Table 4.1 shows the number of correctly classified and misclassified samples for each symbol and the corresponding accuracy. These however do not provide a complete metric to analyze the work. The results also shows that while most of the symbols are classified correctly with high accuracy, zero and the alphabet “W‟ are misclassified as alphabet “O‟ and six respectively in significant cases.

- It was observed that model required very few epochs to converge in spite of having very deep model. A pre-trained model was used to initialize the model with weights and this step may have reduced the learning time to a considerable amount. The pretrained model was trained on a ImageNet database. ImageNet database was used because it contains around 1.2 million images. More the number of images, the more different initial features will be discovered and there will be a greater probability of input image features to be matched.

The validation loss and validation accuracy obtained during training is tabulated below:

Epoch	Validation Loss	Validation Accuracy
1	0.3883	81.29
2	0.1726	93.20
3	0.3165	94.31
4	0.0715	96.5

Table No. 4.2 – Overall Validation Loss and Accuracy

Table 4.2 depicts the validation loss and accuracy obtained during each epoch of training the model. During the first four epoch loss decreased and accuracy increased at each epoch. After the fourth epoch the loss started increasing and the model started.

Thus, we have designed a system using which it enables the deaf and dumb community to create recognition and also to give them a standard platform to communicate and express their opinions with every other individual. Using Convolutional Neural Networks (CNN) we have been able to get an accuracy of 95% which is higher than all the previously implemented systems.

```
Using TensorFlow backend.
C:\Users\omen\Anaconda3\lib\site-packages\ipykernel_launcher.py:12: UserWarning: Update your `Conv2D` call to the Keras 2 AP
I: `Conv2D(32, (3, 3), input_shape=(64, 64, 3..., activation="relu")`
  if sys.path[0] == '':
C:\Users\omen\Anaconda3\lib\site-packages\ipykernel_launcher.py:18: UserWarning: Update your `Conv2D` call to the Keras 2 AP
I: `Conv2D(32, (3, 3), activation="relu")`
C:\Users\omen\Anaconda3\lib\site-packages\ipykernel_launcher.py:22: UserWarning: Update your `Conv2D` call to the Keras 2 AP
I: `Conv2D(64, (3, 3), activation="relu")`
Found 45500 images belonging to 26 classes.
Found 6500 images belonging to 26 classes.
Epoch 1/4
800/800 [==============================] - 222s 278ms/step - loss: 2.2264 - accuracy: 0.3287 - val_loss: 0.3883 - val_accura
cy: 0.8129
Epoch 2/4
800/800 [==============================] - 245s 306ms/step - loss: 0.7024 - accuracy: 0.7670 - val_loss: 0.1726 - val_accura
cy: 0.9320
Epoch 3/4
800/800 [==============================] - 190s 237ms/step - loss: 0.4194 - accuracy: 0.8587 - val_loss: 0.3165 - val_accura
cy: 0.9431
Epoch 4/4
800/800 [==============================] - 196s 245ms/step - loss: 0.3117 - accuracy: 0.8924 - val_loss: 0.0715 - val_accura
cy: 0.9655
dict_keys(['val_loss', 'val_accuracy', 'loss', 'accuracy'])
```

Output –

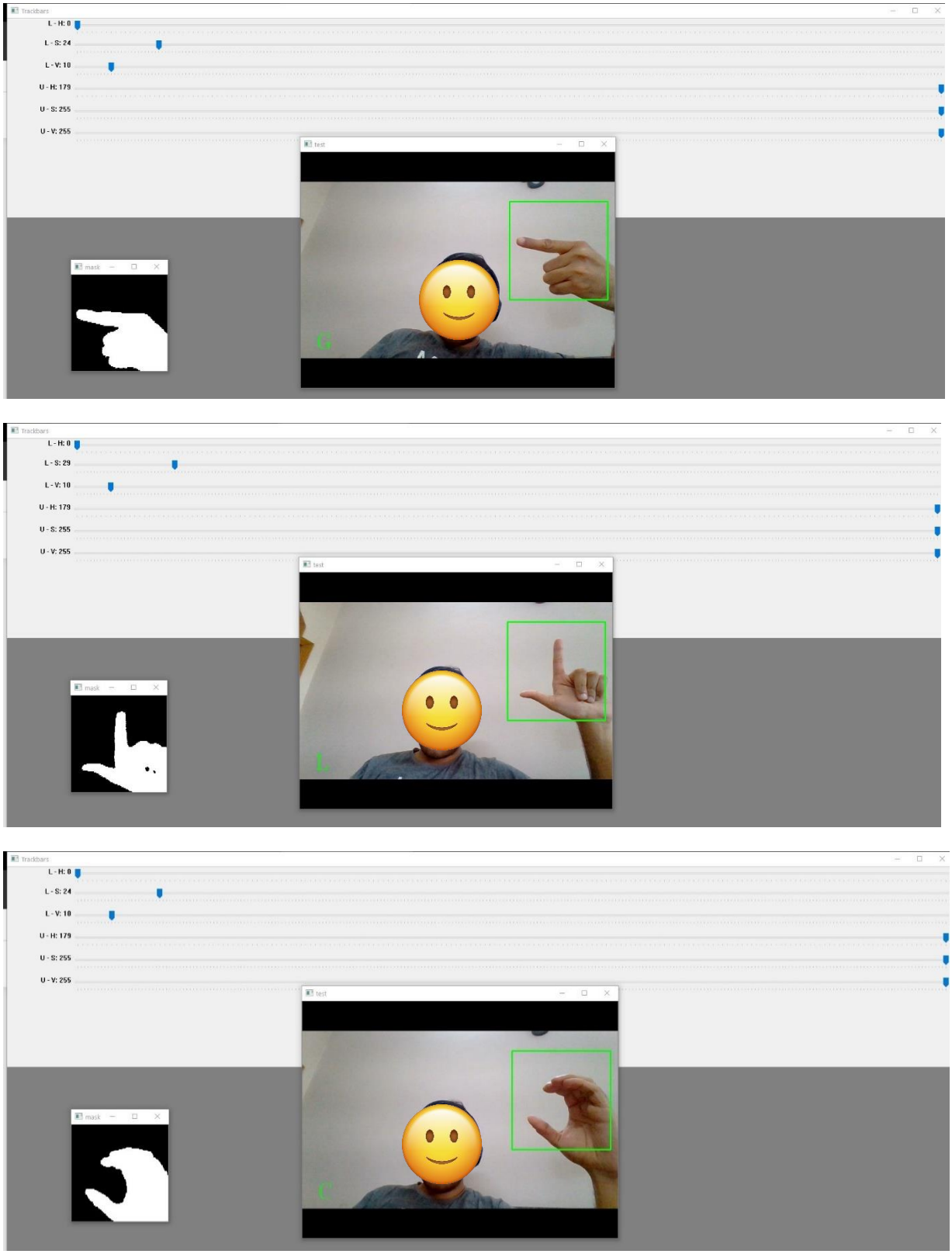

Figure No. 4.1 – Output

4.2 Discussions

The proposed system is able to detect the hand gestures with more accuracy when compared with the existing system

- It takes into consideration the detailed features including the shape, size, color of bare hand
- The processing time required for the proposed system is less compared to the existing system
- The proposed model successfully handles two way communication by converting sign language into text, speech and vice versa.

This work study and examined how to outwardly perceive every single static motion of American Sign Language (ASL) with uncovered hand. Diverse clients have diverse hand shapes and skin hues, making it progressively troublesome for the framework to perceive a signal. Gesture based communication Recognition is fundamental for the less privileged individuals to speak with other individuals.

CHAPTER 5: CONCLUSION

- The principal goal of this project is to determine gesture recognition that might enable the deaf to converse with the hearing people. The features extraction is one of the important task such as different gestures should result in different, good discriminable features.
- In this proposal, we have created an idea of translating the static image of sign language to the spoken language of hearing. The static image includes alphabet and some words, used in both training and testing of data. Feature representation will be learned by a technique known as convolutional neural networks.
- We use CNN algorithm trained dataset to detect the character from the gesture images. From the results obtained above we can conclude that Convolution Neural Network provides a remarkable accuracy in identifying the sign language characters including alphabets and numerals.

FUTURE WORK

- With the help of these features and trained dataset we can extend it to recognize ASL alphabets and numbers with accuracy in real time.

- The proposed system can be made available in multi languages making it more reliable and efficient. It could be made available entirely on the mobile devices which will help in the making the system handier and portable in the near future

REFERENCES

[1] Sarfaraz Masood, Harish Chandra Thuwal, Adhyan Srivastava, “American Sign Language Character Recognition using Convolution Neural Network” Department of Computer Engineering, Jamia Millia Islamia, New Delhi – 110025, INDIA. October 2018, DOI: 10.1007/978-981-10-5547-8_42.

[2] Rutuja J., Aishwarya J., Samarth S. , Sulaxmi R. , Mrunalinee P. “Hand Gesture Recognition System Using Convolutional Neural Networks” ,Volume: 06 Issue: 04 | Apr 2019 p-ISSN: 2395-0072

[3] Sobia Fayyaz, Yasar Ayaz, “CNN and Traditional Classifiers Performance for Sign Language Recognition”, January 25–28, 2019, Da Lat, Viet Nam © 2019 Association for Computing Machinery. ACM ISBN 978-1-4503-6612-0/19/01

[4] Ming Jin Cheok, Zaid Omar , Mohamed Hisham Jaward, “ A review of hand gesture and sign language recognition techniques”, 23 June 2016 / Accepted: 31 July 2017, DOI 10.1007/s13042-017-0705-5

[5] N.Priyadharsini, N.Rajeswari, “Sign Language Recognition Using Convolutional Neural Networks”, International Journal on Recent and Innovation Trends in Computing and Communication ,ISSN: 2321-8169 ,Volume: 5 Issue: 6

[6] Pigou, Lionel, Sander Dieleman, Pieter-Jan Kindermans, and Benjamin Schrauwen. “Sign language recognition using convolutional neural networks”, In Workshop at the European Conference on Computer Vision 2014, pp. 572578. Springer International Publishing

[7] Simonyan, Karen, and Andrew Zisserman. "Very deep convolutional networks for large-scale image recognition." arXiv preprint arXiv:1409.1556 (2014).

[8] Bottou, Léon. "Large-scale machine learning with stochastic gradient descent." In Proceedings of COMPSTAT'2010, pp. 177-186. Physica-Verlag HD, 2010

[9] Agarwal, Anant & Thakur, Manish. Sign Language Recognition using Microsoft Kinect. In IEEE International Conference on Contemporary Computing , 2013.

[10] Cao Dong, Ming C. Leu and Zhaozheng Yin. American Sign Language Alphabet Recognition Using Microsoft Kinect. In IEEE International Conference on Computer Vision and Pattern Recognition Workshops , 2015.

[11] P. Mekala et al. Real-time Sign Language Recognition based on Neural Network Architecture. System Theory (SSST), 2011 IEEE 43rd Southeastern Symposium 14-16 March 2011.

[12] Asha Thalange, Dr. S. K. Dixit, “COHST and Wavelet Features Based Static ASL Numbers Recognition”, 2 nd International Conference on Intelligent Computing, Communication & Convergence (Elsevier), pp.455-460, 2016.