SIGN LANGUAGE DETECTION

Pavitra Kadiyala[1]

[1]*SCOPE, Vellore Institute of Technology, Vellore, India*

Abstract - *Communication through signs has consistently been a significant way for communication among hearing and speech impaired humans, generally called deaf and dumb. It is the only mode of communicating for such individuals to pass on their messages to other human beings, and hence other humans need to comprehend their language. In this project, a sign language detection or recognition web framework is proposed with the help of image processing. This application would help in recognizing Sign Language. The dataset used is the Indian Sign Language dataset. This application could be used in schools or any place, which would make the communication process easier between the impaired and non-impaired people. The proposed method can be used for the ease of recognition of sign language. The method used is Deep Learning for image recognition and the data is trained using Convolution Neural Network. Using this method, we would recognize the gesture and predict which sign is shown.*

Key Words: Sign Language, Convolution Neural Network, Image Processing, Framework, Gestures

1.INTRODUCTION

Sign language is the most significant way of communication between the impaired people. Establishing an easy way of communication with deaf and dumb people is very important. Everyone should be able to understand sign language as it would be useful in case of any emergency. These individuals communicate through hand signals or gestures. Signals are essentially the actual activity structure performed by an individual to pass on the important data. People trained to know sign language would be able to communicate efficiently but this would be a problem for the untrained ones.

Communicating using signs is using gestures and using hands or expressions to depict a particular sign. Few people use signs for communicating when they are busy for example in a meeting or so. There are many types of signs depending on the language and region. This paper proposes a method for identifying the Indian Sign Language.

We know the field of Machine Learning, Artificial Intelligence, Image Processing is advancing and can be used for multiple domains nowadays. In this paper, Deep Learning is used.

There are many factors that as taken into consideration when it comes to signing language recognition. The angle of the gesture also plays a very important role. The type of dataset also plays a very vital role in the recognition model.

1.1 Existing Solutions

There are a few ways which are being used.

1. Searching in a book or online for the matching image of the sign represented. But this is a time-consuming method for searching.
2. To have an experienced and educated translator every time with you. But this is a very tough method to find a translator everywhere you go.
3. Non-vision-based which uses Sensor and Hardware-based devices for detecting.

1.2 Motivation of the Topic

Keeping in mind the above drawbacks of the existing method, this paper proposes a solution to it by using Deep Learning. Also, this method would help the ease of understanding, and even in public areas the communication is easy.

2. LITERATURE REVIEW

In this paper [1], a gesture-based communication fingerspelling letters in order ID framework was created by utilizing image processing and Artificial Intelligence. Specifically, they introduced 24 sequential images by a few blends. Histogram of Oriented Gradients (HOG) and Local Binary Pattern (LBP) highlights of each motion were taken from the training dataset. Then, they applied Multiclass Support Vector Machines (SVMs) to prepare this separated information. Additionally, Convolutional Neural Network (CNN) design has been used in the training dataset for correlation. The Massey Dataset is used in the training and testing periods of the entire framework. They have done five main methods. They are HOG-SVM, LBP-SVM, HOG-LBP-SVM, CNN, CNN-SVM.

This paper [2] audits alternate strategies received to lessen obstruction of correspondence by building up an assistive gadget for hard-of-hearing quiet people. The fundamental objective is to build up a constantly installed gadget for impaired people to help their communication. This paper proposes the use of ANN for the detection.

This paper [3] clarifies another method without sensors. A picture handling procedure called Histogram of angle (HOG) alongside CNN has been utilized to prepare the System. Web Camera is utilized to take the picture of various signals and that will be utilized as information to the Mat lab. The product will perceive the picture and recognizes the center's forthcoming voice yield which is played utilizing the voice replay unit. This paper clarifies both-way correspondences between the hard of hearing, moronic and typical individuals which imply the proposed framework is equipped for changing the gesture-based communication over to text and voice.

This paper [4] portrays another strategy for American Sign Language letters using Microsoft's Kinect device. First, they used a per-pixel classification-based algorithm. Then, a various leveled mode-chasing technique is created and carried out to limit hand joint situations. Last, a Random Forest (RF) classifier is worked to perceive ASL signs utilizing the joint points. They utilized a freely accessible dataset from Surrey University.

This paper [5] proposed use for a glove-based hard of hearing quiet correspondence mediator framework. The glove is inside furnished with five flex sensors, which tell a corresponding change in position for every signal, material sensors, and an accelerometer that estimates the direction of the hand. Arduino is used for handling the signal data. The glove incorporates two methods of activity – preparing mode to profit each client and an operational mode. The link of letters to shape words is additionally done in Arduino.

This paper [6] builds an ongoing framework for hand motion recognition which perceives hand signals, gestures of hands, and afterward, convert motion pictures into voice and the other way around. To execute this, they utilized a straightforward night vision webcam with a 20-megapixel force. The paper comprised of executing a framework utilizing computerized reasoning, Image Processing, and information mining to accept contribution as hand signals and produce text and voice output.

This paper [7] presents a Leap Motion regulator for Australian Sign Language (Auslan) recognition. Testing showed that the regulator can give exact following of hands and fingers, and to follow development. This identification loses precision when the hand moves into a place that hinders the regulator's capacity to see. When individual fingers of the hands are united, this device would not be accurate and may fail.

An overview of ongoing approaches is given in this paper [8]. In this paper, gesture recognition techniques such as contact-based and vision-based have been given. They also have given an overview of gesture technologies such as 3D models and appearance-based models. They have given 4 main human gesture recognition approaches that are feature extract, model-based, template match, and hybrid.

This paper [9] presents a novel framework to help in speaking with those having vocal and hearing incapacities. It examines a strategy for gesture-based communication and change of discourse to signs. The calculation is equipped for extricating signs from video successions under insignificantly jumbled and dynamic foundations utilizing skin shading division. It recognizes static and dynamic motions and concentrates the proper element vector. These are grouped utilizing Support Vector Machines. Discourse acknowledgment is based upon a standard module - Sphinx. Their trial results give the good division of indications under assorted foundations and moderately high exactness in motion and discourse acknowledgment.

This paper [10] recognizes hand motions into text and voice. For this interaction, RealTimeImage made by impaired individuals is caught and it is given as a contribution to the pre-processor. At that point, the component extraction measure by utilizing otsu's calculation and grouping by utilizing SVM(Support Vector Machine)is done. The acquired content is changed over into voice with the utilization of the

MATLAB work. In this way, hand motions made by impaired individuals had been converted into text and voice for better communication.

3. METHODOLOGY

Neural Networks can be used to recognize patterns and we would use a convolution neural network for the detection and recognition of signs. Libraries and languages for the deep learning model used are NumPy, pandas, TensorFlow, Keras, sklearn, python. We use flask and Html for our web framework.

The data set used is the Indian Sign Language Dataset andhas Alphabets from A-Z[11].

```
['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', 'z']
Found 11389 images belonging to 26 classes.
Found 1248 images belonging to 26 classes.
```

Fig. 1. A Snippet of the Dataset

There are 26 classes. Each class corresponds to one letter. We split the dataset into 2 sections. One is used for training and the other is used for testing the model.

Algorithm

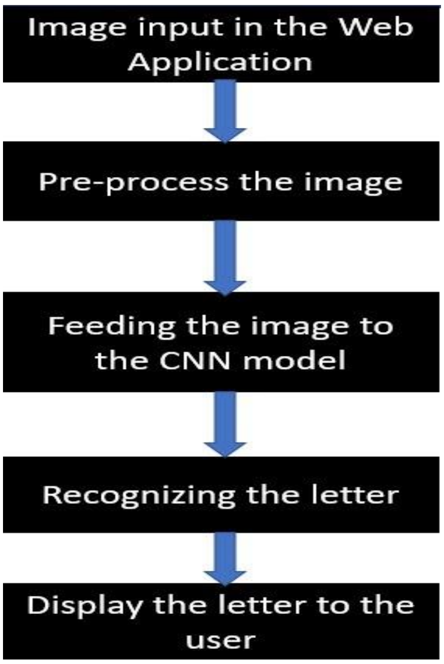

Fig. 2. Proposed Framework Flow

1) Install and import the necessary libraries.
2) Load and split the dataset for test and train.
3) Data Augmentation
4) Define the CNN Model
5) Train the training data on the model and evaluating in the test.
6) Storing the model and integrating with a website
7) Taking input from the user on the website and predicting the sign.

The input image is normalized. For the model, the convolution base is first created, and pooling is done. Then flattening layer is added and then the dense layers are added (hidden and output layer). The model is compiled compile. The optimizer used is adam and the loss function used is categorical cross-entropy. Data is trained and tested on this model for 10 epochs and then save the model. This model is integrated with the website.

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
conv2d (Conv2D)              (None, 62, 62, 32)        896
_________________________________________________________________
max_pooling2d (MaxPooling2D) (None, 31, 31, 32)        0
_________________________________________________________________
conv2d_1 (Conv2D)            (None, 29, 29, 32)        9248
_________________________________________________________________
max_pooling2d_1 (MaxPooling2 (None, 14, 14, 32)        0
_________________________________________________________________
conv2d_2 (Conv2D)            (None, 12, 12, 32)        9248
_________________________________________________________________
max_pooling2d_2 (MaxPooling2 (None, 6, 6, 32)          0
_________________________________________________________________
flatten (Flatten)            (None, 1152)              0
_________________________________________________________________
dense (Dense)                (None, 128)               147584
_________________________________________________________________
dense_1 (Dense)              (None, 26)                3354
=================================================================
Total params: 170,330
Trainable params: 170,330
Non-trainable params: 0
_________________________________________________________________
```

Fig. 3. Model Snippet

4. RESULT

The model was trained for 10 epochs and the result is as follows:

```
356/356 [==============================] - 102s 285ms/step - loss: 2.6681 - accuracy: 0.2234 - val_loss: 1.6202 - val_accura
cy: 0.5136
Epoch 2/10
356/356 [==============================] - 79s 222ms/step - loss: 1.3131 - accuracy: 0.6080 - val_loss: 0.7301 - val_accurac
y: 0.7917
Epoch 3/10
356/356 [==============================] - 73s 205ms/step - loss: 0.7404 - accuracy: 0.7695 - val_loss: 0.4125 - val_accurac
y: 0.8838
Epoch 4/10
356/356 [==============================] - 79s 221ms/step - loss: 0.4925 - accuracy: 0.8450 - val_loss: 0.3113 - val_accurac
y: 0.9038
Epoch 5/10
356/356 [==============================] - 77s 216ms/step - loss: 0.3513 - accuracy: 0.8905 - val_loss: 0.2435 - val_accurac
y: 0.9215
Epoch 6/10
356/356 [==============================] - 76s 215ms/step - loss: 0.2824 - accuracy: 0.9128 - val_loss: 0.2175 - val_accurac
y: 0.9375
Epoch 7/10
356/356 [==============================] - 79s 222ms/step - loss: 0.2332 - accuracy: 0.9252 - val_loss: 0.1270 - val_accurac
y: 0.9591
Epoch 8/10
356/356 [==============================] - 98s 274ms/step - loss: 0.1918 - accuracy: 0.9375 - val_loss: 0.0967 - val_accurac
y: 0.9696
Epoch 9/10
356/356 [==============================] - 88s 248ms/step - loss: 0.1602 - accuracy: 0.9467 - val_loss: 0.0639 - val_accurac
y: 0.9792
Epoch 10/10
356/356 [==============================] - 81s 229ms/step - loss: 0.1543 - accuracy: 0.9500 - val_loss: 0.0749 - val_accurac
y: 0.9728
```

Fig. 4. Model

The accuracy graph between the training and validation datais shown below.

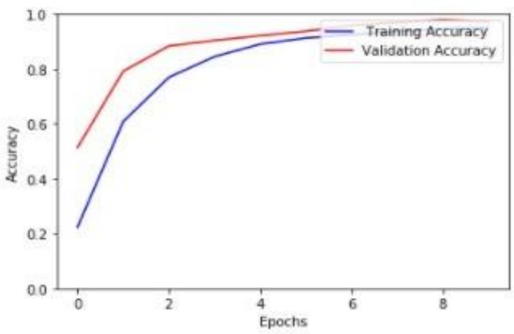

Fig. 5. Model Accuracy

Then the model was integrated with the web application.

Fig. 6. The initial web application image

The input is taken from the user and the input is Pre-Processedand then the letter is predicted using the model and then the model matched the sign with the classes and display the letter of the class matched. We have 26 classes.

Fig. 8. Selecting an Image

Fig. 8. Prediction for the letter H

Fig. 9. Prediction for the letter A

5. FUTURE SCOPE

We could use key point matching and other techniques for more accurate decisions. The model can be run further for more epochs. The data could be increased. We can add words, numbers as well for this. We could also do this by inputting a constant stream ofimages and that get a resultant string for a particular word. Wecould also put an input field for letters and get the sign as an output, if possible, to make it into a fully sign communication platform. We could get an android application for the same. Text to speech could be added as well. It could be made multilingual as well.

6. CONCLUSIONS

Sign Language is very important to impair and non-impairedpeople. This Web Application aims to ease the communication between both using Deep Learning. It is way so that sign language can be consequently caught, perceived, and the meanttext be showed on the web. There are different strategiesfor sign language change. Some of them utilize hardware and others utilize visual-based methodology. This framework preprocesses the image and then using the model displays the text on the website. This could be deployed and used anywhere using your phone.

REFERENCES

[1] H. B. D. Nguyen and H. N. Do, ”Deep Learning for American Sign Language Fingerspelling Recogni- tion System,” 2019 26th International Conference onTelecommunications (ICT), Hanoi, Vietnam, 2019, pp. 314-318, doi: 10.1109/ICT.2019.8798856.

[2] Kakde, Manisha Nakrani, Mahender Rawate, Amit. (2016). A Review Paper on Sign Language Recognition System For Deaf And Dumb People using Image Pro- cessing. International Journal of Engineering Research and. V5. 10.17577/IJERTV5IS031036.

[3] Sign Language and Gesture Recognition for Deaf and Dumb People by Hema B, Sania Anjum, Umme Hani, Vanaja P, Akshatha M

[4] Cao Dong, M. C. Leu and Z. Yin, ”American Sign Language alphabet recognition using Microsoft Kinect,” 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Boston, MA, USA, 2015, pp. 44-52, doi: 10.1109/CVPRW.2015.7301347.

[5] Rajamohan, Anbarasi Hemavathy, R. Dhanalakshmi, Munirathnam. (2013). Deaf-Mute Communication In- terpreter. International Journal of Scientific Engineeringand Technology. 2. 2277-1581.

[6] Shinde, Shweta Autee, Rajesh Bhosale, Vitthal. (2016). Real time two way communication approach for hearingimpaired and dumb person based on image processing. 1-5. 10.1109/ICCIC.2016.7919572.

[7] Leigh Ellen Potter, Jake Araullo, and Lewis Carter. 2013. The Leap Motion controller: a view on sign language. In Proceedings of the 25th Australian Computer-Human Interaction Conference: Augmentation, Application, Innovation, Collaboration (OzCHI ’13). Association for Computing Machinery, New York, NY, USA, 175–178. DOI:https://doi.org/10.1145/2541016.2541072

[8] Real Time Hand Gesture Recognition System: A Review1 Shaikh Shabnam, 2 Dr. Shah Aqueel

[9] A. Kumar, K. Thankachan and M. M. Dominic, ”Sign language recognition,” 2016 3rd International Conference on Recent Advances in Information Technology (RAIT), Dhanbad, India, 2016, pp. 422-428, doi: 10.1109/RAIT.2016.7507939.

[10] Translation of Sign Language for Deaf and Dumb People by Suthagar S., K. S. Tamilselvan, P. Balakumar, B. Rajalakshmi, C. Roshini

[11] Indian-Sign-Language-dataset by Ayeshatasnim Hannure