INDIAN SIGN LANGUAGE RECOGNITION USING CONVOLUTIONAL NEURAL NETWORK

Vrushali Chavan[#1], Ankita Sakhare[#2], Vaishnavi Pingale [#3], Dimple Thoke [#4], Prof. Mrs. Jayshree Patil[#5]

School of Computer Engineering and Technology, MIT Academy of Engineering Alandi (D). Pune

***Abstract*— Nowadays, many people are facing difficulties that are dumb and deaf, etc. They face difficulties while communicating with each other. Earlier developed techniques are communicated with each other. Earlier developed technologies are all based on sensors, and this technique is not given the perfect solution. This paper explains a new technique of the Convolutional Neural Network(CNN) for training and classifying the images. In this paper, a web camera is used to capture the image for different signs classification, and these images are used as input images. The system will recognize these images, and then it will predict the output of those images. This paper explains two-way communications between deaf, dumb, and normal people. This means the proposed system is capable of recognizing the Indian sign languages into a text.**

Keywords*— *Sign Language, Feature Extraction, Gesture Recognition, CNN, Virtual, Deaf, Dumb.

I. INTRODUCTION

The problems deaf and dumb people face when communicating with the system in the workplace because they cannot hear and it is too risky to visit a place alone because they cannot hear the sound of a car, bicycle, etc. can or other people. It can be difficult for them to adapt quickly to the environment and respond to other normal people and express themselves. The history of sign language dates back to the 17th century as a visual communication method for recording the history of the sign [1]. Sign language Recognition has two approaches: first, based on images and, secondly, based on sensors. Much research nowadays is based on image-based approaches as the complex device does not need to wear gloves. As required with a sensor-based approach, the sign language recognition system has multiple uses, such as in the area of the human user interface. Communication sector, multimedia and security sector. Related to understanding image patterns. There are two steps: the first step is character recognition and the second step is character recognition. In sign recognition, we extract a feature from gestures in relation to defined parameters. Recognizing characters helps to identify the contours that distinguish the object from the rest of the contours [2]. Language, especially in cases where no transfer options are available. The technical characteristics of sign language. The system is your social address, the meaning of your hand gestures. In this system, the webcam helps capture drawing images, and then a preprocessing method is applied to the images. The system webcam to eliminate background noise does preprocessing and also uses an algorithm based on slopes. The sign language also represents conventional gestures, the hand sign represents the letters of the alphabet and mimics the spelling of the figures. Represent a complete idea or phrase [5]. The aim of the sign language recognition system is to facilitate communication between deaf people and normal people by means of sign language hand movements without intelligence using a sensor for mute people and producing the text as output [4].

II. LITERATURE REVIEW

Sign language recognition is an essential application for recognizing and classifying gestures. Two different methods are often used in sign language recognition: the first is based on usage data, and the second is based on vision. Many different researchers have focused on the subject of a Sign Language Recognition System. These works can be represented as the following:

[3] Tubaiz, Noor, Tamer Shanableh, and Khaled Assaleh propose a glove-based Arabic sign language recognition system using a novel technique. In this system, manual labeling was carried out using a camera to identify word boundaries. This system required input must be in Arabic language. The main disadvantage is that every time signer needs to be wear glove.

[4] Ravi Patel propose vision/image-based method. This method uses a webcam to take video or pictures. The signal recognition process based on vision or image can be divided into two stages: training and the classifier must be trained using the training data set in the training stage. Data set creation, pre-processing, feature extraction, and classifier training are the main stages of the training phase. The testing phase includes video/image recording, pre-processing, function extraction, and classification.

[5] Garge, Pragati, Navin Agraval, and Sanjeev Sofat developed this system. This system implemented data glove method. This method uses various sensors to collect gesture signals. The gesture signal is analog. ADC is used to convert analog signals into digital form. It consists of a flexible sensor and an accelerometer. The flexible sensor records, the curvature signal.

III. METHODOLOGY

Four steps for sign language recognition, i.e., dataset generation, feature extraction, Gesture Recognition, interpretation. The first step is dataset generation, then that dataset needs training, so accuracy will increase.

A. DATASET CREATION

In this module, the data collection work is completed. For that, already implemented datasets are not suitable because the gesture must be made using two hands, and the Indian Sign language requires two hands. So using both hands, we prepared a new dataset by clicking more gestures. And that’s why OpenCV is used. By using OpenCV, two windows are created for capturing the colour images and one for grey images for threshold. Then applied a Gaussian blur filter to extract the features of the captured image like hand border, skin colour, etc.

B. GESTURE CLASSIFICATION

This approach uses two layers of algorithms to estimate the user's ultimate symbol. Apply the filter and thresholding to the image captured by OpenCV to get the image after extraction of feature. Then, these images are sent to the convolution classification model for prediction of the given images. And when a letter is observed for more than 50 images, the letter is printed and taken to form the word. The space between words is used for the empty symbol.

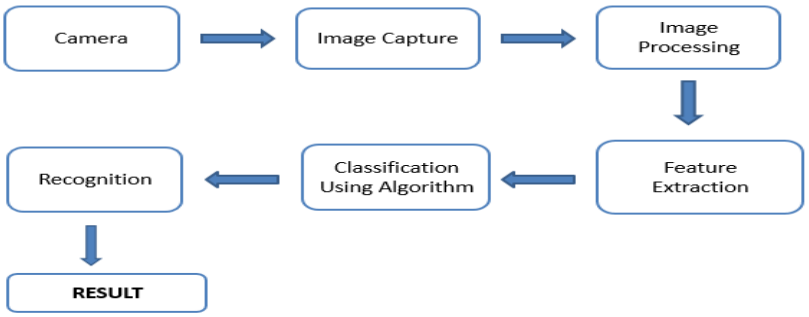

Figure1: Module Description

C. CNN MODEL

1. Keras First Convolution Layer: It is the first layer to extract features from the input image of 128 X 128 pixels. Here we define the kernel as the layer parameter. The output is 126 X126 pixel image.

2. Keras First Convolution Layer: It is the first layer to extract features from the input image of 128 X 128 pixels. Here we define the kernel as the layer parameter. The output is 126 X126 pixel image.

3. Keras Second Convolution Layer: Output of keras 1st pooling layer is becoming input for this layer. It has 32 filter weights, and its output is 60X60 pixel image.

4. Second Pooling Layer: here again max pooling with filter 2X2**.** its output is 30X30 pixel image.

5. First Keras Dense Layer: In these layers are fully connected with 128 neurons . Each node in this current layer is connected to the previous one which becomes densely connected. In these we are reusing output of previous layer that is 30X30X32 =28800 values. output of these layer is given to second keras Dense layer. Dropout is 0.4 to avoid overfitting.

6. Second Keras Dense Layer: The output from the first Dense Connected Layer will take as an input for the fully connected layer.

7. The Final Layer: The output of the second Dense Connected Layer taken as an input for the final layer which contains the 26 number of neurons as there are 26 numbers of classes we are classifying.

D. SENTENCE FORMATION IMPLEMENTATION

1. Every time the count of the detected letter surpasses a particular value is 50, and there is no other letter close to it, there is a threshold value which is 70. Then we print that letter and append it to the current string.

2. Otherwise, the current directory which holds the count of detected symbols is cleared to keep away from predicating the wrong letter.

3. Whenever the plain background is detected, then a certain value is exceeded, and if the current buffer is empty, no spaces will be found.

4. In other times it predicts the word with space and current string added to the sentence below.

IV. RESULT AND DISCUSSION

A. DATASET CREATION

Initially, the system captured approximately 400 images of each letter in SLRS for training and testing purposes, and approximately 240 images for each character used. In principle, every frame displayed by the system's webcam will be recorded. Each frame has a region of interest (ROI) represented by a blue square (see image below).

Figure2: Dataset Collection

Figure3: Dataset Collection

B. PREPROCESSING

Extract RGB color ROI from the image and convert it to a grayscale image. Then apply a Gaussian blur filter to all images to extract various features. The results after applying the Gaussian blur filter are listed below.

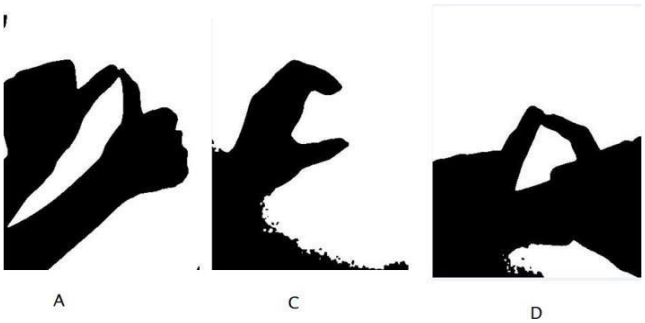

Figure4: After Preprocessing

C. TRAINING AND PREDICTION

This method uses two levels of algorithms to predict the user's final character: apply a filter and thresholding to the image captured with OpenCV to obtain a processed image after extracting the function. Send the processed image to the convolution model for prediction. If they found letter exceeds 50 frames, print it and take into account the composition of the word.

Figure5: Gesture prediction

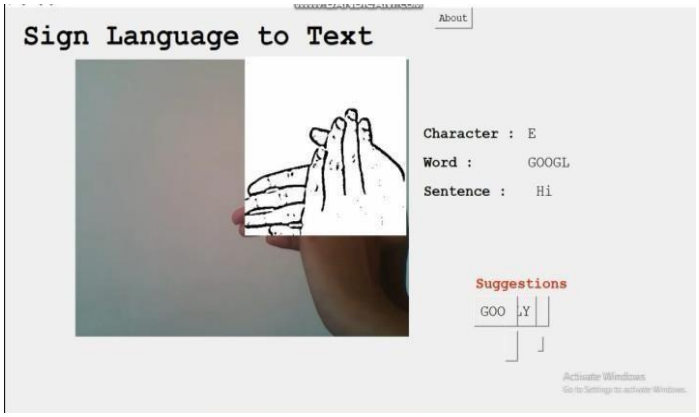

Figure6: Word Formation

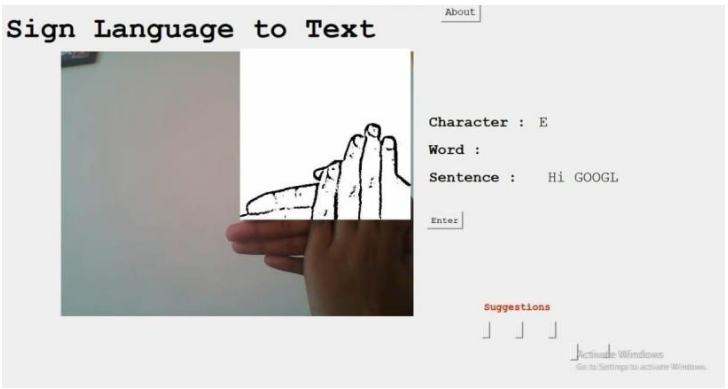

Figure7: Sentence Formation

V. CONCLUSION

The goal of the Indian sign language recognition is to establish direct communication between the deaf and dumb people. This project is useful for people with ideas and tools to communicate with each other. Sign language recognition that can recognize hands and fingers and give you the meaning of them. Recorded the input data with the help of a webcam. We developed a model for Indian Sign Language that recognizes words and sentences. We achieved 85% accuracy and our model shown good predicting the results.

ACKNOWLEDGMENT

We would like to give special thanks to our institute, the, the MIT Academy of Engineering. We would also like to thank the School of Computer Engineering and Technology.

REFERENCES

[1] Bell Laboratories RECORD (1969) A collection of several articles on the AT&T Picture phone Archived 2012- 06-23 at the Way back Machine (then about to be released) Bell Laboratories, Pg.134–153 & 160–187, Volume 47, No. 5, May/June 1969.

[2] Rajam, P. Subha, and G. Balakrishnan. "Real time Indian sign language recognition system to aid deaf-dumb people." 2011 IEEE 13th international conference on communication technology. IEEE, 2011.

[3] Tubaiz, Noor, Tamer Shanableh, and Khaled Assaleh. "Glove-based continuous Arabic sign language recognition in user-dependent mode." IEEE Transactions on Human-Machine Systems 45.4 (2015): 526-533.

[4] Ravi Patel B.E. Student , "A Review on Image based Indian Sign Language Recognition ", Department of Information Technology, Vishwakarma Government Engineering College, December 2018.

[5] Garg, Pragati, Naveen Aggarwal, and Sanjeev Sofat. "Vision based hand gesture recognition." World academy of science, engineering and technology 49.1 (2009): 972-977.

[6] Saurabh Kumar Soni, Namrata Singh, Sagarika Gupta, SachinGiri ,Prof. N.U. Khan "Vision Based Gesture Recognition" February 2018.