Indian sign language recognition using convolution neural network

Sukanya L[1*]*, Tharun* E[1*], *Anup Raj* G[1], *Shreyas Singh* T[1], and *Srinivas* S[1]

[1]Department of Information Technology, GRIET, India

Abstract. The goal of the project is to create a machine learning model that can classify the numerous hand motions used in sign language fingerspelling. Communication with deaf and dumb persons is frequently difficult. A variety of hand, finger, and arm motions that assist the deaf and hard of hearing in communicating with others and vice versa. Classification machine learning algorithms are taught on a set of image data in this userindependent model, and testing is done on a completely other set of data. For some people with particular needs, sign language is their only means of communicating their thoughts and feelings. It enables individuals to understand the world around them by visual descriptions and hence contribute to society. As a result, our model aids us in solving the problem more broadly. By watching the user's hand gestures, this transforms sign language to regular words.

1 Introduction

Hand signals and gestures are used by those who are unable to speak. Ordinary people have trouble understanding their own language. As a result, a system that identifies various signs and gestures and relays information to normal humans is required. It connects physically challenged humans with most people. there are various signal languages in the market, which include American signal Language (ASL), British signal Language (BSL), Indian signal Language (ISL), and many others. however, on these studies, we are figuring out and converting Indian sign Language into text. Humans find communication incredibly simple since it allows us to express ourselves. Speech, gestures, body language, reading, writing, and visual aids are all ways we communicate, with speech being the most common. sadly, there stays a communication gap for the minority of individuals who are deaf or hard of listening to. Fingerspelling, which spells out phrases character by using character, and phrase stage association, which includes hand motions that transmit phrase which means, are two varieties of sign language. Fingerspelling is an important tool in sign language because it allows people to communicate names, addresses, and other words that don't have a meaning at the word level. Despite this, fingerspelling is not frequently used since it is difficult to grasp and use. However, there is no universal sign language, and only a few individuals are familiar with it, making it an ineffective means of communication. The main goal of the

_____

*Corresponding author: ledalla.sukanya@gmail.com,

project is to recognize frames from live video using updated versions of TensorFlow and OpenCV, which are used in image processing the dataset's images.

2 Literature survey

This Sahil Jain and K V Sameer Raja worked on ISL Recognition with colored images. To extract features, they used strategies such as bag of visual words, Gaussian random, and the Histogram of Gradients (HoG). SVM was trained on three individuals and tested on a separate user with a 54.63 percent accuracy. For ISL, there is no standard dataset. As a result, Mukesh Kumar Makwana, an M.E. student at IISc, developed a dataset. There were 43,750 depth photos in all, with 1,250 for each of the 35 hand gestures. These were taken from five separate people. The alphabets (A-Z) and digits (0-9) are represented by gestures, with the exception of "2," which is identical to "v." The photos are grayscale and 320x240 pixels in size. H 2 A convolution neural network is a feed forward network. The CNN achieved accuracy in a variety of image processing applications, and several researchers used it to extract deep features in time series analysis. A convolutional layer, an activation layer, and a memory layer are the three layers of a traditional CNN. Pooling layers are divided into two types: layer and max pooling layers. To the traffic flow data, we apply a time and periodicity shift. A 1D convolution neural network was employed to carry out the training process, which encapsulates deep-learning.
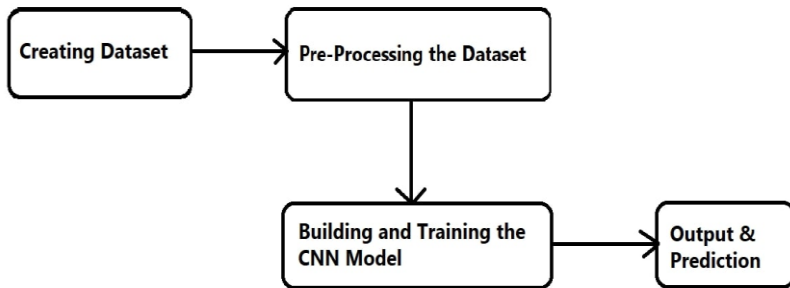
3 System architecture

Fig. 1. Functioning of our system

4 Methodology

4.1 Dataset

The dataset which is used in this project is created by hand gestures. Using OpenCV library, hand gestures given as an input to the system and save it to a file in our local machine and then this is to train model. This dataset has two main folders test and train each folder consists 36 elements that is 26 elements for Alphabets and 10 numbers which is considered as output of our model. We have collected 4628 images to train our model. The dataset we captured by our camera is first transformed into gray image and then threshold is applied in the images in order to decrease the file size this is done because to reduce the burden on the system and to increase the efficiency of the model. We can increase the training images in order to increase the efficiency and accuracy.

5   Implementation

5.1   Processing the dataset

So far, we have our dataset with 26 labels and 0-9 numeric value signs so we need to process the model. To process the model, we need a code and to code we must first import the required packages.

```python
# Coordinates of the ROI
x1 = int(0.5*frame.shape[1])
y1 = 10
x2 = frame.shape[1]-10
y2 = int(0.5*frame.shape[1])
# Drawing the ROI
# The increment/decrement by 1 is to compensate for the bounding box
cv2.rectangle(frame, (x1-1, y1-1), (x2+1, y2+1), (255,0,0) ,1)
# Extracting the ROI
roi = frame[y1:y2, x1:x2]
roi = cv2.resize(roi, (64, 64))

cv2.imshow("Frame", frame)

# do the processing after capturing the image!
roi = cv2.cvtColor(roi, cv2.COLOR_BGR2GRAY)
_, roi = cv2.threshold(roi, 120, 255, cv2.THRESH_BINARY)
cv2.imshow("ROI", roi)
```

Fig. 2. Processing the dataset code

This is sample and some part of the code which will execute and after this process the complete processed data will be converted into pickle files so that the processed data will be safe with us and if we want to move to any other system and work with this project then we can do that in any system that is the main theme of the pickle files format. In above code image is resized to 64*64 and then image converted into grayscale to reduce the weightage of the data. Once the image changed into grayscale from color, Threshold will be applied to the image.

5.2   Training the model using CNN

In the training of the model will be done after the preprocessing of the dataset. we will be training the model so that the image we will be selecting is trained and the parameters of image will be extracted from that and those calculations of the dataset will be shown at the time of training and that code will be executed with CNN. Then after the image will be trained under different parameters of the image that parameters.

Fig. 3. Training the model code

As a part of training the model, first we need to build the Convolution Neural Network (CNN). And the above code tells us how it is done. First, we need to initialize the CNN and build layers of Convolution layers and Pooling layer, here Max Pooling is used. After building Convolution layers, fully Connected layer will be added to CNN and compile CNN. After Building we need to use this CNN and prepare the train/test data, the following code will show how training of model will happen.

5.3 Working the model

The working model is so important and here we work on predicting and displaying it to the user. For that we need to work on previously saved model that is 'model-bw.h5' in Read Only mode. And also, we should import some libraries too. Run the code in command prompt or Console or terminal to predict the gestures.
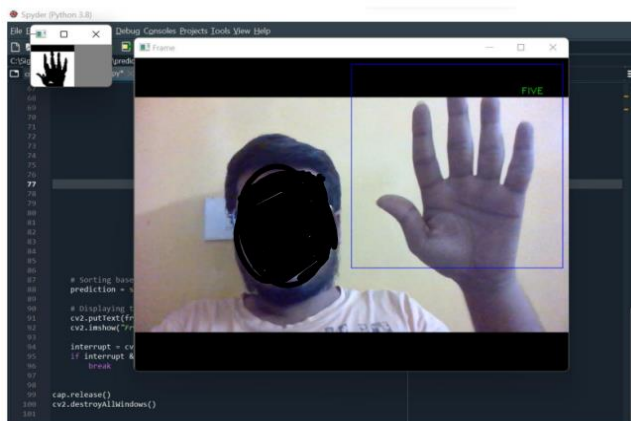
6 Results

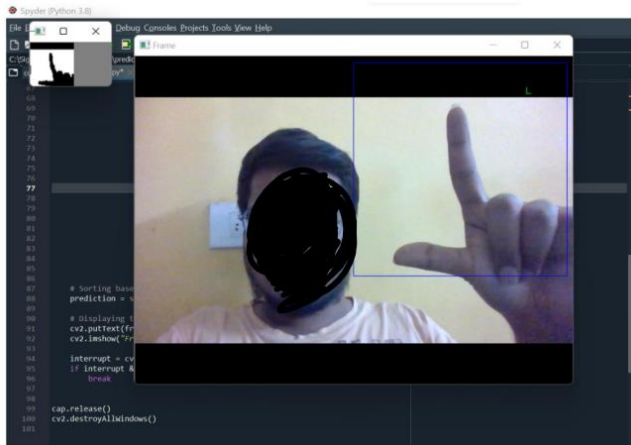**Fig. 4.** Predicting the numerical value

Fig. 5. Predicting the alphabet

7 Conclusion

Our methodology makes it simple for users to converse with physically challenged persons, which is critical when travelling to different countries. Even if we have problems expressing our opinions, this becomes an excellent instrument for doing so, as it is simple to master and allows us to communicate. When we consider all of the conceivable combinations of motions that a system like this must interpret and translate, sign language recognition becomes a difficult challenge. Although the system did not function well, it was shown that a first-person sign language translation system may be developed utilizing only cameras and convolutional neural networks.

References

[1] Utpal Nandi, Anudyuti Ghorai, Moirangthem Marjit Singh, Chiranjit Changdar, Shubhankar Bhakta & Rajat Kumar Pal, "Indian sign language alphabet recognition system using CNN with diffGrad optimizer and stochastic pooling," , 08 January 2022.

[2] Dushyant Kumar Singh. "3D-CNN based Dynamic Gesture Recognition for Indian Sign Language Modeling" , 14 July 2021.

[3] Image Processing, https://docs.opencv.org/4.7.0/d7/dbd/group__imgproc.html

[4] Prabhu, "https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148". 04 March 2018.

[5] Abubakr   masood,   "https://medium.com/analytics-vidhya/introduction-to-image-processing-with-opencv-2a1bbef8eda0". 22 February 2021.

[6] Priyanka Rawat, "All About Indian Sign Language" 11 February 2021.