American Sign Language(ASL) detection system using Convolutional Neural Networks(CNN) and Computer Vision

Submitted By

Siddhiben H. Shah

20MCEC15

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2022

American Sign Language(ASL) detection system using Convolutional Neural Networks(CNN) and Computer Vision

Major Project Part II

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Science and Engineering

Submitted By

Siddhiben H. Shah

(20MCEC15)

Guided By

Prof. Sapan H. Mankad

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2022

Certificate

This is to certify that the major project entitled **"American Sign Language(ASL) detection system using Convolutional Neural Networks(CNN) and Computer Vision."** submitted by **Siddhiben H. Shah (20MCEC15)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Science and Engineering of Nirma University, Ahmedabad, is the record of work carried out by her under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this Major Project Part-II, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Sapan Mankad
Internal Guide & Assistant Professor
CSE Department
Institute of Technology
Nirma University, Ahmedabad

Dr Sudeep Tanwar
Professor & PG Coordinator (M.Tech - CSE)
CSE Department
Institute of Technology
Nirma University, Ahmedabad

Dr Madhuri Bhavsar
Professor & Head
CSE Department
Institute of Technology
Nirma University, Ahmedabad

Dr Rajesh N. Patel
Director
Institute of Technology
Nirma University, Ahmedabad

Statement of Originality

I, **Siddhiben H. Shah**, **20MCEC15**, give undertaking that the Major Project Part II entitled **"American Sign Language(ASL) detection system using Convolutional Neural Networks(CNN) and Computer Vision."** submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Science & Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made.It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Signature of Student

Date:

Place:

Endorsed by

Prof. Sapan H. Mankad

(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. Sapan Mankad**, Assistant Professor, Computer Science and Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr Madhuri Bhavsar**, Hon'ble Head of Computer Science And Engineering Department, Institute of Technology, Nirma University, Ahmedabad for her kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr Rajesh N. Patel**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation she has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Science and Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Siddhiben H. Shah

20MCEC15

Abstract

As the communication is the very important part of our life, communication should be understandable and clear. In this world according to World Health Organization(WHO) there are 432 million adults and 34 million children have disabling hearing loss. So, using deep learning model called Convolutional Neural Networks(CNN), computer vision technologies and TensorFlow and Keras framework it is possible to detect and interpret hand gestures of American Sign Language(ASL). In this project there are approx 17K images of different gestures of conversation are used. Gaussian Filter is used to pre-process the data and as discussed we have used Convolutional Neural Networks to train the images of hand gestures of different people according to American Sign Language(ASL). To interpret and incorporate the voice modules Google Text to Speech API is used. It is observed that CNN is the one of the appropriate model for detection of ASL language. By using deep learning technology and libraries we are able to detect the ASL with 97.7% accuracy. So, for interpret ASL language for deaf peoples different libraries and deep learning techniques are very useful so that communication between disabled people becomes easy.

List of Figures

List of Tables

Contents

Chapter 1

Introduction

In this chapter we will discuss what is the concept behind to build a communication model for the people who are suffering from hearing loss and how deep learning technique is useful for this. The chapter discusses in brief about the data set that is used to train the Convolutional Neural Network , how to pre-process the data to get accurate result and other libraries to build this project more efficient.

1.1 General

Hearing loss is the one of the major problem increasing day by day due to noise pollution and excessive use of speaker/ headphones for a long time on high volume by adults. To help out by making communication between these people easier and efficient many technologies are useful but one of the more efficient and booming technology is deep learning , google libraries and API. Here in this project we have used approx 17K American Sign Language(ASL) gesture image database. First we have to pre-process the data images to smooth the given images and eliminate all irrelevant noise from the database.Here Gaussian Filter is used to do data pre-processing and double thresholding is implemented for consider only the strong and efficient edges in the images.After pre-processing the data we have used Convolutional Neural Networks(CNN) to build the model.So, here in this project Convolutionl Neural Network is the heart of the project and other Google libraries and Computer Vision is used.

1.2 Dataset

To train the model Convolutional Neural Networks(CNN) for American Sign Language(ASL),in this project we have used 36k images of different gestures of different kind of skin tones.The whole dataset have different 29 classes according to the American Sign Language(ASL).Total dataset is further splitted into three different categories Train, Test and validation.Training set contains 22k images and testing set contains 2k images and at last data pre-processing is done on this images to eliminate all the irrelevant noise.

1.3 Objective of Study

As the communication is the most important part of every one's life. In this world according to World Health Organization(WHO) 432 million adults and 34 million children have disabling hearing loss.So, the main objective of this project or study is to build a model for better, easy and efficient communication for those who are suffering from hearing loss.In this project we are trying to build a model which doesn't need any kind of wearing gadgets like wearing gloves or sensor etc.... So,with the help of deep learning technique convolutional Neural Networks this can be possible easily and accurately.So, main motive of this project is to build a highly ccurate model for deaf and dumb people so that they can communicate by just doing gesture of sign language and model will extract those sign and first convert those sign into the single word and then combinely one whole sentences with the help of different google libraries and framework without needing any kind of extra wearing gedgets for the disabled people.

Chapter 2

Literature Survey

2.1 Literature Summary

This chapter includes different experiments done by researchers to detect American Sign Language(ASL). This includes different technologies, ways and some use of different equipment to detect the ASL language. It is observed that most of the researchers have includes deep learning technology CNN and other version of CNN like 2D-CNN , 3D-CNN with approx 98% accuracy to detect the ASL language. Some of have used YOLO with 94.2% accuracy, SVM with 99.4% accuracy, R-CNN with 98.2% accuracy models for their research work. Equipment like sensible gloves are also used for detect the movement of the hand gestures. For database concern most of researchers have used ASL as it is the most common used sign language, besides this Myanmar sign language, BdSLimset(Bangladeshi Sign Langauage) is also used. For data pre-processing purpose some of has used direct edged images and some of have used edging technologies like guassian blur algorithm, Sobel, Canny, Prewitt algorithms.

Here are some valuable papers and their fruitful conclusion depict in following tables.

Paper Title	Year	Type	Author	Summary
Glove Based Deaf-Dumb Sign Language Interpreter	2021	Conference Paper	Santosh Kumar S., Ravi Gatti, Sunil Kumar K N, Nataraja N, Rajendra Prasad P, Sarala T	In this paper author has proposed the idea where the sign language is converted into the different text and voices using A Raspberry Pi based microprocessor to detect the sign language and triaxial accelerometer, flex sensor module is also used to make this project more efficient.[1]
Real Time Bangladeshi Sign Language Detection using Faster R-CNN	2018	Conference Paper	Oishee Bintey Hoque, Mohammad Imrul Jubair, Md. Saiful Islam, Al-Farabi Akash, Alvin Sachie Paulson	The main motive of this paper is to build a model so that Bangladeshi sign laguage can be detected and converted to the text or voices in real time video using Faster Region based Convolutional Neural Networks(R-CNN) which is the one of the technique in Deep Learning technologies.In this paper BdSLImset (Bangladeshi Sign Language Image Dataset) dataset is develop to train the model and 80% of data are used for training the model and 20% are used for testing the model and gain average accuracy rate of 98.2%.[6]

Table 2.1: Physical Glove and R-CNN model based literature summary

Paper Title	Year	Type	Author	Summary
Hand Gesture Detection based Real-time American Sign Language Letters Recognition using Support Vector Machine	2019	Conference Paper	Xinyun Jiang , Wasim Ahmad	The paper is about building a concept which can be helpful to deaf and dumb people by detecting American Sign Language into the text using machine learning concept Support Vector Machine(SVM).In this paper author has only experiments few Signs like B,D,F,L and U and these data are collected from USB connection camera connected to the computer. For preprocessing the data auther has used skin color algorithm in HSV color space to find the Region of Interest (ROI).In this model author has achieved 99.4% accuracy.[2]
Real-Time Sign Language Recognition Based on Video Stream	2020	Conference Paper	Kai Zhao, Kejun Zhang, Yu Zhai, Daotong Wang, Jianbo Su	The main motive of this paper is to build a model so that real time Chinese Sign Language can be detected from the video stream and converted to the text or voices in real time video using 2D-CNN, 3D-CNN as well as You Only Look Once model(YOLO) which is the one of the technique in Deep Learning technologies.For the dataset purpose they have collected total more than 5000 words and demos of Chinese vocabulary/words and the videos are collected from 720P RGB camera of 640*480 pixels frames.By this model 94.2% recognition accuracy can obtained.[7]

Table 2.2: SVM, CNN and YOLO model based literature summary

Paper Title	Year	Type	Author	Summary
Hand Gesture Recognition Software Based on Indian Sign Language	2019	Paper	Mr. Sanket Kadam, Mr. Aakash Ghodke , Prof. Sumitra Sadhukhan	The concept behind this paper is to build the model for who suffering from hearing loss by proposing the concept of converting Indian Sign Language to the text by using Convolutional Nueral Networks(CNN) and Deep Neural Networks(DNN) for train the model and enhance the model, here data are collected with the help of the digital camera and color gloves, hand gestures performed by the participants.By performing this 88.25% accuracy can be achieved.[3]
Myanmar Sign Language Classification using Deep Learning	2019	Paper	Sai Myo Htet, Bawin Aye, Myo Min Hein	The main aim of this paper is to build a model for detecting the Myanmar sign language and convert it into the text using Deep Learning technique called Convolutional Neural Networks(CNN) and the whole system is implemented by MATLAB.In this project author have also performed the skin color enhancement and all the data are collected through webcam with resolution of 640*480 pixels and includes 3 different classes and different 38 hand gestures styles.Here in this paper proposed algorithm is able to distinguish and detect motion sign of Myanmar Sign Language using Machine Learning techniques.[5]

Table 2.3: CNN and DNN based literature summary

Chapter 3

Proposed Research Work

A language translator is rapidly used by the people who are suffering from hearing loss and this one of the best blessing from technology to them so that they can communicate easily and without any hesitation. So, this paper proposed a model that can detect the American Sign Language(ASL) and translate into the text in real time without using any type of extra wearing device like glove. This can be possible with the help of deep learning technology Convolutional Neural Network(CNN) model, Computer Vision, Keras and Tensorflow framework.

The motive of this paper is not only detect the sign language single-single words but also to predict the whole string. With the help of Hunspell library we are also trying to suggest the appropriate whole word for prior given some letter of characters. So, with the help of this model users are able to detect the sign language very easily and conveniently.

Figure 3.1: American Sign Language 29 different classes

The proposed research work's main motive is to detect the American sign language which contains 29 different classes as we can see in Fig:3.1 with accurate results that can

be done starting from data resizing and data pre-processing using edge detection technique called GaussianBlur filtering, for training of given input data using deep learning model CNN, Computer vision and using some libraries like Keras and Tensorflow

Chapter 4

Concepts and Methodology

This chapter includes concepts and methodologies used by our model to detect the American Sign Language(ASL). To pre-process the images Gaussian blur algorithm has been used. In this model We have used CNN to train the given images of the model. Here are the detailed concepts of used technologies to build the model.

4.1 Concepts

The concept of this model is to classify and detect these 29 classes according to live inputs from video capture, for this task first of all we have to resize the given database images from 200*200 to 128*128. After resizing of image next step to build the model is to pre-process the given data. For that firstly images are converted from RGB to gray-scale images and then the Gaussian Blur algorithm for edge detection has applied on it to train the model. After doing these steps we are able to convert the images from RGB to Gray-scale images with the edge detection. For enhancing the functionality of detection here we have used the Hunspell library for suggesting the word according to the given letters. For training purpose, we have reduced the size of the database from 36K to around 24K due to the limitation of the device.

Gaussian Filtering is used to preprocess the data images as it can smooth the images by "blur" the images and can discard noise and detail of the image.It uses Gaussian distribution Function at different filters which describes the edge of the Gaussian curve.

4.1.1 Edge detection technique: Gaussian Blur Algorithm

Gaussian blur algorithm is one of the image pre-processing techniques which is based on liner filter. The main concept behind the Gaussian Blur algorithm is to reduce the noise of the image by blurring the image using the Gaussian function. Most of the sign language researchers have used the Gaussian blur algorithm for smoothing the image and reduce the noise by blurring the image. So, the Gaussian Blur algorithm is mainly used to improve the quality of digital images by pre-processing images. Two-dimensional(2D) Gaussian filter equation can represent by:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Figure 4.1: 2D Gaussian Filter equation

After preprocess the data images, training of model took place.To train the model we have used deep learning concept called Convolutional Neural Networks(CNN).

4.1.2 Convolutional Neural Network

Convolutional Neural Network is the application of Deep learning technologies and it has a wide range of applications in today's world. It is the learning algorithm that takes input as an image, video, etc..and according to given inputted data features it assigns the value and trains the model. After training, these values are used to differentiate the features for new inputs. Convolutional Neural Network has generally an input layer, 2D convolutional layer, pooling layer(Max pooling or Min pooling), and at last fully-fledged connected layer. With the help of convolutional operation, the convolutional layer extracts the given input data. After that to prevent overfitting with optimum epochs pooling layer will extract the significant features which will further be given to the fully connected layer. By this CNN can automatically do feature extraction which is the most important feature of this model than other image processing techniques. The basic structure of CNN model is described in the figure :4.2.

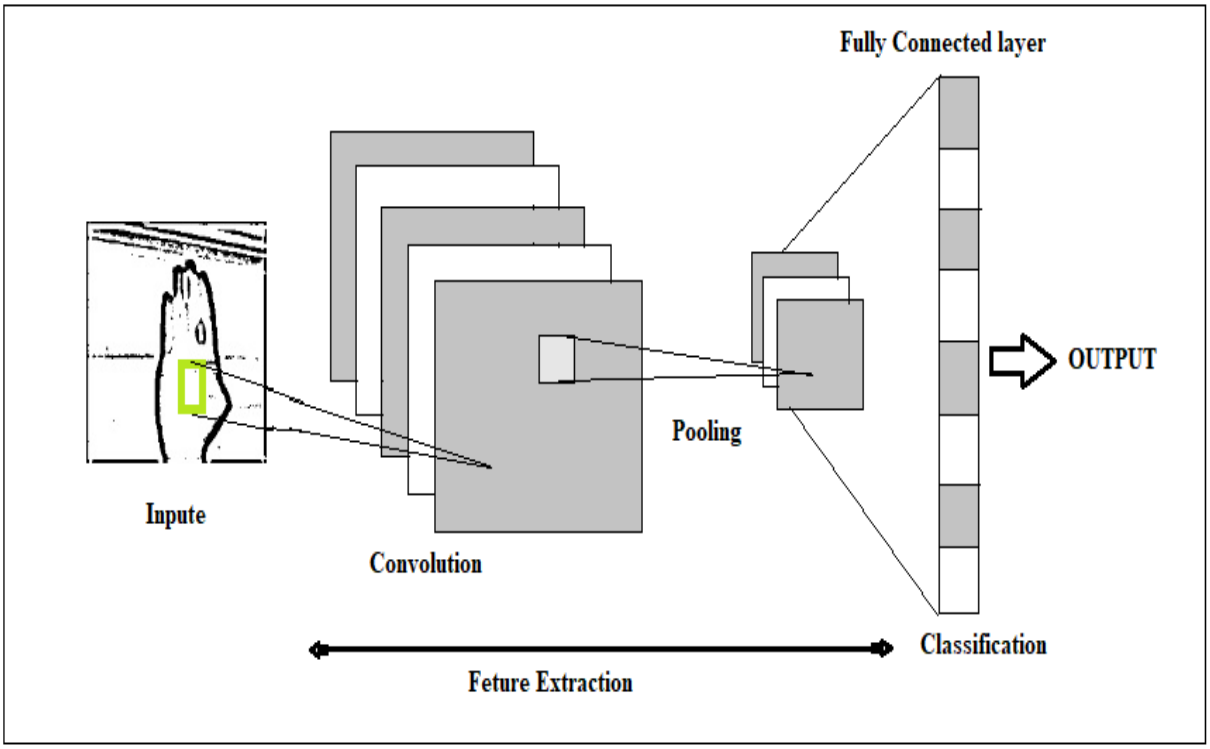

Figure 4.2: Convolutional Neural Networks

4.2 Methodology and Workflow

To train the model, database is further divided into mainly two parts Train and Test. After training the model and gaining certain accuracy with the help of CNN, the outcome of feature extraction for classifying input images are stored as checkpoints. For further to take an input of sign language different gestures from a user, Graphical User Interface(GUI) has created which is the video capturing interface from where again sign language gesture is taken from a user and after that Gaussian Filtering is applied for edge detection and finally model can detect the given input on the bases of the trained model. After detecting the text of the given input further it will be converted into the speech using google speech library and the python module Playsound. The whole workflow is shown in the figure: 4.3

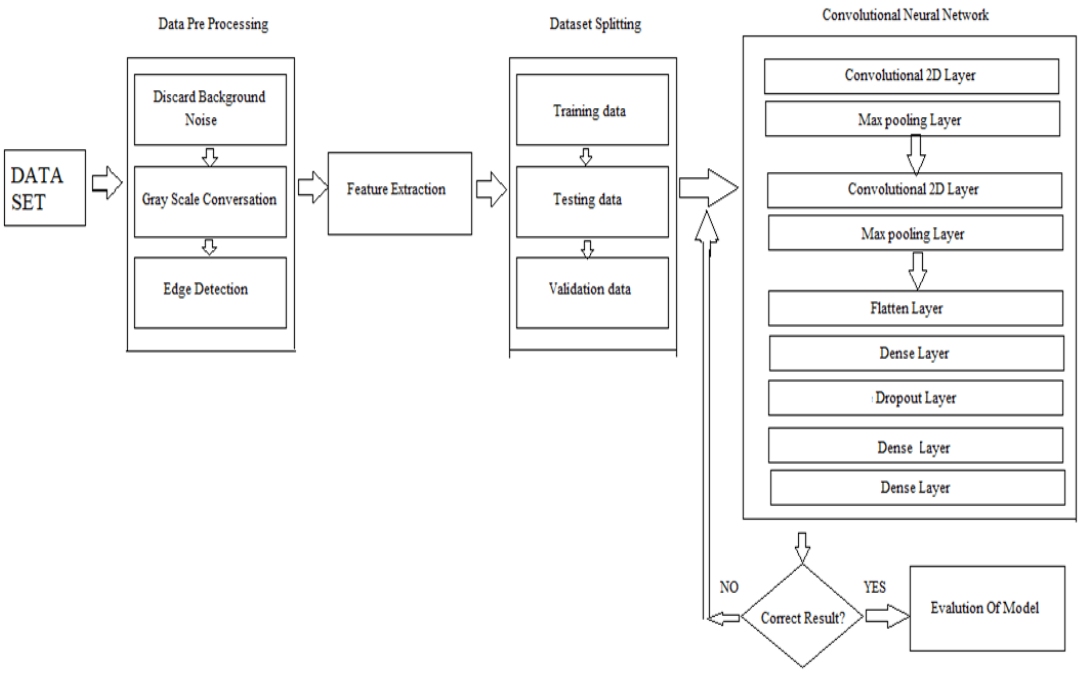

Figure 4.3: Work flow of the model

Chapter 5

Execution and Implementation

5.1 Dataset Selection and Pre-processing

For the dataset concern we have used Kaggle ASL-Sign language dataset which contains at most 36K data with 29 different classes for A to Z, del, space and nothing but because of machine limitation dataset size has reduced to around 24K.This database is further divided into the Train and Test. Training and testing both have same 29 classes in which training dataset contains around 22K and Testing dataset contains 2K data.

The very first step is to data pre-process.We have used Gaussian Blur algorithm for that and convert RGB based American Sign Language into the gray-scaled sign language with only edges where only edges of any hands are visible so that model which we are going to train will not got biased towards the color of skin. After converting RGB image to gray-scaled edge image as given in figure:5.1. Figure:5.2 shows the image of 29 different classes of dataset converted from RGB to gray-scale.

Figure 5.1: Image before and after data pre-processing

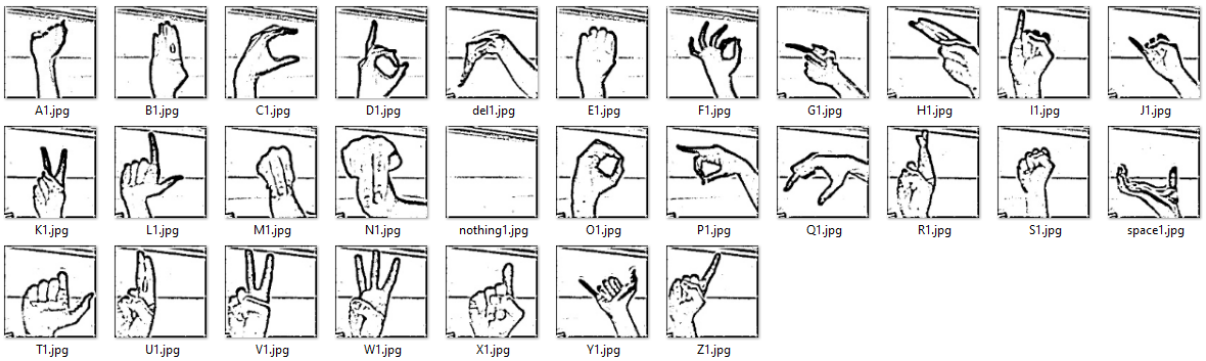

Figure 5.2: 29 different classes after data pre-processing

5.2 Execution and Result

After data pre-processing, we have trained the CNN Model with two Convolutional layers and two max-pooling layers, four fully connected dense layers, as activation functions, "Relu" and "Softmax" and as a optimizer, we have used "Adam" optimizer". The whole model is trained for 25 epochs and in between, we have saved the checkpoints of the trained model for accuracy. After training the model we have got 97.7% accuracy. With the help of this model we are able to predict the given sign text string as we can see in the figure:5.3

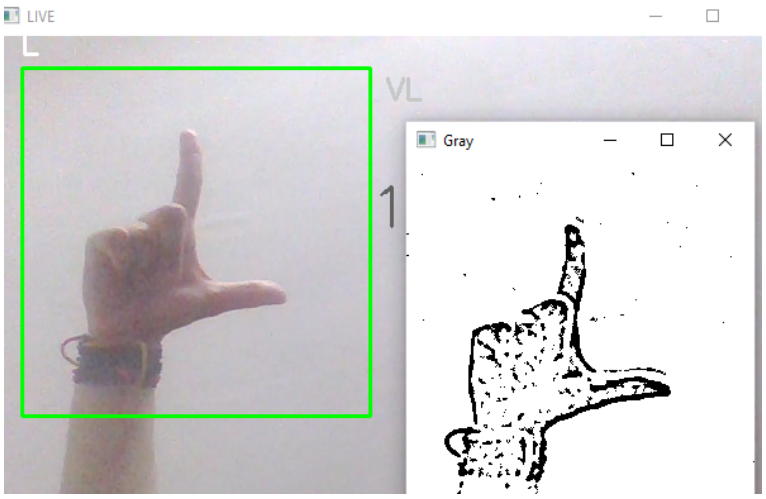

Figure 5.3: Result

For capture the real time signs hand gestures given by the users we have used Computer Vision technology. For detection, we have used the Hunspell library for suggesting the whole word according to the given letter so that the user can predict the next word accurately. After detecting the string text next step is to convert this text into the au-

dio/speech, for that we have used the python module Playsound. So, at last, we are able to convert these text strings given by the user in the form of American sign language into speech so that disabled as well as normal both the persons can communicate in a better way.

So, After all of the steps done, we have come up with 97.7% accuracy of model to detect the American Sign Language(ASL).

Chapter 6

Conclusion

In this paper, the model for American Sign language detection has been performed to make disabled people's life more convenient and easy with the help of deep learning technology and other libraries. After training the model for 25 epochs we are able to detect the signs with 97.7% of accuracy.To enhance the accuracy of model, very high quality computation device is required.We are also able to guess the whole words according to prior letters using Hunspell library.This is observed that there is a conflict wrong detection of some signs like between A-E and M-N.

Chapter 7

Future Work

Following are the work that needs to be done in future:

- Can be done on various languages too.
- In future we can try to add fillings so that who are suffering from blindness can also communicate easily.
- For future purposes, we can try some combination of deep learning models like CNN with LSTM, RNN, etc... for getting more accuracy.
- Also, it is possible to build a multi-language sign detection model for better communication across the world.

Bibliography

[1] S. S. Kumar, R. Gatti, S. K. N. Kumar, N. Nataraja, R. P. Prasad and T. Sarala, "Glove Based Deaf-Dumb Sign Language Interpreter," 2021 International Conference on Recent Trends on Electronics, Information, Communication Technology (RTE-ICT), 2021, pp. 947-950, doi: 10.1109/RTEICT52294.2021.9573990.

[2] X. Jiang and W. Ahmad, "Hand Gesture Detection Based Real-Time American Sign Language Letters Recognition using Support Vector Machine," 2019 IEEE Intl Conf on Dependable, Autonomic and Secure Computing, Intl Conf on Pervasive Intelligence and Computing, Intl Conf on Cloud and Big Data Computing, Intl Conf on Cyber Science and Technology Congress (DASC/PiCom/CBDCom/CyberSciTech), 2019, pp. 380-385, doi: 10.1109/DASC/PiCom/CBDCom/CyberSciTech.2019.00078.

[3] S. Kadam, A. Ghodke and S. Sadhukhan, "Hand Gesture Recognition Software Based on Indian Sign Language," 2019 1st International Conference on Innovations in Information and Communication Technology (ICIICT), 2019, pp. 1-6, doi: 10.1109/ICI-ICT1.2019.8741512.

[4] B. B. Atitallah et al., "Hand Sign Recognition System based on EIT Imaging and Robust CNN Classification," in IEEE Sensors Journal, doi: 10.1109/JSEN.2021.3130982.

[5] S. M. Htet, B. Aye and M. M. Hein, "Myanmar Sign Language Classification using Deep Learning," 2020 International Conference on Advanced Information Technologies (ICAIT), 2020, pp. 200-205, doi: 10.1109/ICAIT51105.2020.9261775.

[6] O. B. Hoque, M. I. Jubair, M. S. Islam, A. -F. Akash and A. S. Paulson, "Real Time Bangladeshi Sign Language Detection using Faster R-CNN," 2018 International Conference on Innovation in Engineering and Technology (ICIET), 2018, pp. 1-6, doi: 10.1109/CIET.2018.8660780.

[7] K. Zhao, K. Zhang, Y. Zhai, D. Wang and J. Su, "Real-Time Sign Language Recognition Based on Video Stream," 2020 39th Chinese Control Conference (CCC), 2020, pp. 7469-7474, doi: 10.23919/CCC50068.2020.9188508.

[8] L. Huynh and V. Ngo, "Recognize Vietnamese Sign Language Using Deep Neural Network," 2020 7th NAFOSTED Conference on Information and Computer Science (NICS), 2020, pp. 191-196, doi: 10.1109/NICS51282.2020.9335904.

[9] A. S. Nikam and A. G. Ambekar, "Sign language recognition using image based hand gesture recognition techniques," 2016 Online International Conference on Green Engineering and Technologies (IC-GET), 2016, pp. 1-5, doi: 10.1109/GET.2016.7916786.