Transfer Learning model for Gesture Recognition based on deep features extracted by CNN

A Project Phase-2 Report

Submitted to the APJ Abdul Kalam Technological University

in partial fulfillment of requirements for the award of degree

Bachelor of Technology

in

Information Technology

by

Ananthakrishnan K.M(TRV19IT010)

Arjun S(TRV19IT014)

Sanora Teressa Raju(TRV19IT049)

Vishnu V(TRV19IT059)

DEPARTMENT OF INFORMATION TECHNOLOGY

GOVERNMENT ENGINEERING COLLEGE, BARTON HILL,

THIRUVANANTHAPURAM-695035

KERALA

June 2023

DEPARTMENT OF INFORMATION TECHNOLOGY

GOVERNMENT ENGINEERING COLLEGE, BARTON HILL,

THIRUVANANTHAPURAM-695035

2022 - 23

CERTIFICATE

This is to certify that the report entitled **Transfer Learning model for Gesture Recognition based on deep features extracted by CNN** submitted by **Ananthakrishnan K.M** (TRV19IT010), **Arjun S** (TRV19IT014), **Sanora Teressa Raju** (TRV19IT049) & **Vishnu V** (TRV19IT059) to the APJ Abdul Kalam Technological University in partial fulfillment of the B.Tech. degree in Information Technology is a bonafide record of the project phase-1 work carried out by them under our guidance and supervision. This report in any form has not been submitted to any other University or Institute for any purpose.

Dr. Vijayanand K.S
(Guide)
Professor and Head
Dept.of IT
Government Engineering College,
Barton Hill, Trivandrum

Dr. Haripriya A.P
(Project Coordinator)
Associate Professor
Dept.of IT
Government Engineering College,
Barton Hill, Trivandrum

Dr. Vijayanand K.S
Professor and Head
Dept.of IT
Government Engineering College,
Barton Hill, Trivandrum

DECLARATION

We hereby declare that the project report **Transfer Learning model for Gesture Recognition based on deep features extracted by CNN** , submitted for partial fulfillment of the requirements for the award of degree of Bachelor of Technology of the APJ Abdul Kalam Technological University, Kerala is a bonafide work done by us under supervision of **Dr. Vijayanand K.S**.

We also declare that we have adhered to ethics of academic honesty and integrity and have not misrepresented or fabricated any data or idea or fact or source in my submission.We understand that any violation of the above will be a cause for disciplinary action by the institute and/or the University and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been obtained. This report has not been previously formed the basis for the award of any degree, diploma or similar title of any other University.

Ananthakrishnan K.M

Arjun S

Sanora Teressa Raju

Vishnu V

Trivandrum

22-12-2022

Abstract

Transfer learning is a machine learning technique where a model trained on one task is re-purposed on a second related task. This can be done by taking the pre-trained model and adding additional layers on top to solve the second task, or by fine-tuning the weights of the pre-trained model on the second task.

One way to use transfer learning for gesture recognition is to train a convolutional neural network (CNN) on a large dataset of images and extract the deep features from the CNN. These deep features can then be used as input to a separate classifier, which can be trained to recognize gestures using the extracted features.

By using transfer learning in this way, it is possible to take advantage of the CNN's ability to learn rich features from images and apply that knowledge to the task of gesture recognition. This can lead to better performance and faster training times compared to training a model from scratch on the gesture recognition task.Here the approach is to use a pre-trained CNN model and fine-tune it on a dataset of hand gestures. This can be done by unfreezing a few of the layers of the pre-trained model and training those layers on the hand gesture dataset.

Using a CNN for hand gesture recognition can be effective because CNNs are able to learn rich features from images and can handle variations in the size, position, and orientation of the hand gestures.

Acknowledgement

We take this opportunity to express our deepest sense of gratitude and sincere thanks to everyone who helped us to complete this work successfully.We express our sincere thanks to **Dr. Vijayanand K.S**, Head of Department, Information Technology, Government Engineering College, Barton Hill, for providing us with all the necessary facilities and support.

We like to express our sincere gratitude to **Dr. Haripriya A.P** Department of Information Technology, Government Engineering College, Barton Hill,Trivandrum for their support and co-operation.

We would like to place on record our sincere gratitude to our project guide **Dr. Vijayanand K.S**, Professor and Head, Information Technology, Government Engineering College, Barton Hill, for the guidance and mentorship throughout this work.

Finally we thank our family, and friends who contributed to the successful fulfilment of this project work.

Ananthakrishnan K.M

Arjun S

Sanora Teressa Raju

Vishnu V

Contents

List of Figures

Chapter 1

Introduction

Communication between disabled people and communication between disabled with others is a challenge . The people who are deaf and dumb communicate with others using sign language and for those doesn't know sign language need a translator to help others understand their words. But not everyone has that capacity to employ a translator and thus they may not communicate with others easily, so we have developed a project to help deaf and dumb people to translate their signs with others using concepts of deep learning

Transfer learning is a machine learning technique where a model trained on one task is re-purposed on a second related task. In the context of gesture recognition, transfer learning could be useful if we have a limited amount of data available for training a model to recognize gestures. Hand gesture recognition is the process of identifying and interpreting hand gestures made by a human user. It is a field of study within computer vision and human-computer interaction, and has a wide range of applications including virtual and augmented reality, sign language interpretation, and gaming.Hand gestures can be classified into different categories based on the number of fingers used, the orientation of the hand, and the motion of the hand. Some examples of hand gestures include waving, pointing, making a fist, and holding up a number of fingers.

There are several techniques that can be used to recognize hand gestures, including machine learning algorithms, computer vision techniques, and depth sensing technologies. Machine learning algorithms such as neural networks and support vector

machines can be trained on large datasets of hand gestures to recognize patterns and classify new gestures. Computer vision techniques such as feature extraction and template matching can be used to analyze the shape and motion of the hand in an image or video.

Gesture recognition technology has been more useful in recent years because gesture communication is a unique form of nonverbal communication that uses body motions and gestures to convey ideas. Recognize the human gestures using mathematical algorithms. A few modes of human-computer interaction (HCI) exist. These devices has their own limitations. Build user-friendly interfaces. Gestures can be originated from any bodily motion or state. Enables users to interact with the devices

Hand gesture recognition using convolutional neural networks (CNNs) is a machine learning approach to identifying and interpreting hand gestures made by a human user. CNNs are a type of neural network that are particularly well-suited to image classification tasks and have been successful in a wide range of applications, including object recognition and facial expression analysis.

To recognize hand gestures using a CNN, a dataset of images of hand gestures is first collected and labeled with the corresponding gesture. The CNN is then trained on this dataset, learning to recognize patterns in the images that correspond to different hand gestures. Once trained, the CNN can be used to classify new images of hand gestures, determining which gesture is being made in the image.

Using a CNN for hand gesture recognition has several advantages. CNNs are able to learn rich features from images and can handle variations in the size, position, and orientation of the hand gestures. They also have the ability to learn from large datasets, allowing for robust gesture recognition even with complex or subtle gestures.

By using the CNN as a feature extractor, we can leverage the knowledge it has learned from the original image classification task and apply it to the gesture recognition task. This can significantly improve the performance of the gesture recognition model, especially if we have a limited amount of data available for training.

In summary, transfer learning for gesture recognition involves using a pre-trained CNN to extract deep features from input images, and training a separate classifier on these features to recognize gestures. This approach can be effective in cases where we have limited data available for training a model from scratch.

Chapter 2

Literature Review

In [1] the authors stated that the hand gesture recognition possesses extensive applications in virtual reality, sign language recognition, and computer games. Dataset is relied on American Sign Language(ASL) (fig 3.1), which helps to obtain the aacuracy rate of about 98.34%. Sign language can involve combining orientation and movements of the hands, arms or body, hand shapes, and facial expressions to express thoughts and words which can be used for communication mostly by Deaf and Dumb people.The hand gesture identification is done in 3 step approach. The hand gesture is first placed for image segmentation and is then sent to feature extraction where the cross-correlation coefficient is applied on the gesture to recognise it.

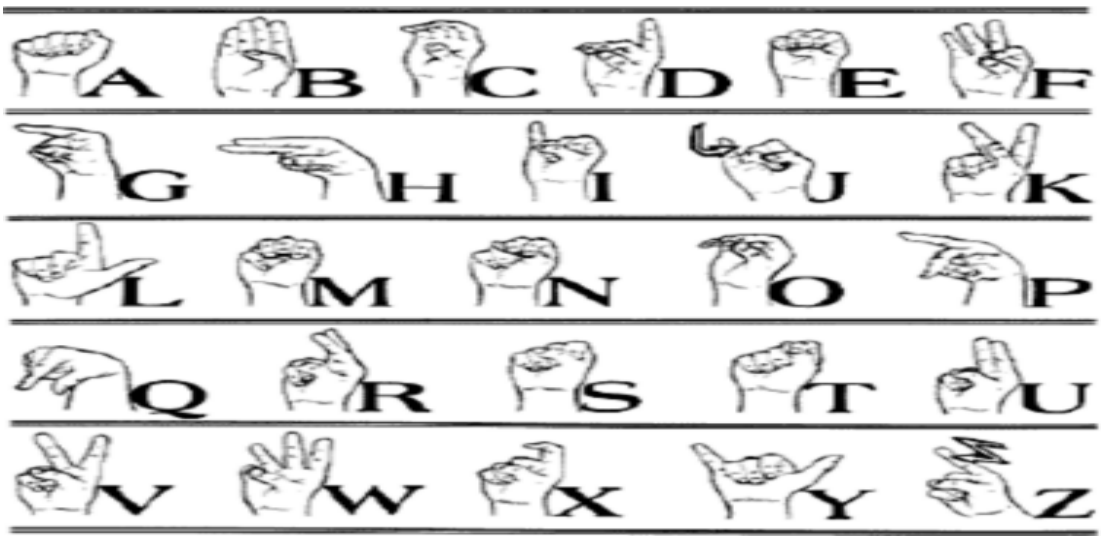

Figure 2.1: American Sign Language

In [2] Real-time human-computer interaction system based on hand gesture which consists of three components I.hand detection II.gesture recognition III.human computer interaction. Nowadays two methods are used primarily to perform gesture recognition. One is based on professional, wearable electromagnetic devices, like special gloves. The other one utilizes computer vision. The nature of gesture recognition is a classification problem. There are lots of approaches to handle 2D gesture recognition, including the orientation histogram, the hiddenMarkov model, particle filtering, support vector machine (SVM), etc. The system employs a CNN to learn features and to recognize gestures only using one cheap monocular camera.The accuracy rate of the recognition can reach over 99.8%. The CNN architecture simplifies the classification process to run faster and is also made runnable in real time.

In the stage of gesture recognition, the CNN classifier feeds on the processed, binary images, where the hand contour is centered and adjusted to a fixed size, and then produces a probabilistic result. The system recognizes the gesture as the one with the highest probability. All features are extracted, or say learned, by the CNN itself. Such a characteristic often leads to a better classification result when lacks an effective mean to extract features. The CNN, which used here to recognize gestures, contains two convolutional layers, each of which is followed by a max-pooling layer, and two fully connected layers. It uses rectified liner unit (ReLU) as the activation. The preprocessing steps threshold, resize and center the hand image and thus introduce contrast, scale and translation invariants to some degree during the CNN learns features.

In [3] introduced a system for detection of hand gestures.The proposed system consists of two modules which are the detection and prediction module. The detection module is implemented using Java while the prediction module is achieved through neural networks in Python. The detection module is responsible for detecting the hand with which gestures. The first step is Background subtraction which is to remove the background such that the hand is focussed. The next step is to convert the video feed into a HSB video feed that is the video feed is pre-processed such that the areas with skin pixels are converted to white and the other colour gradients are converted to black . The images for data set is also taken from the video , the images are taken and saved

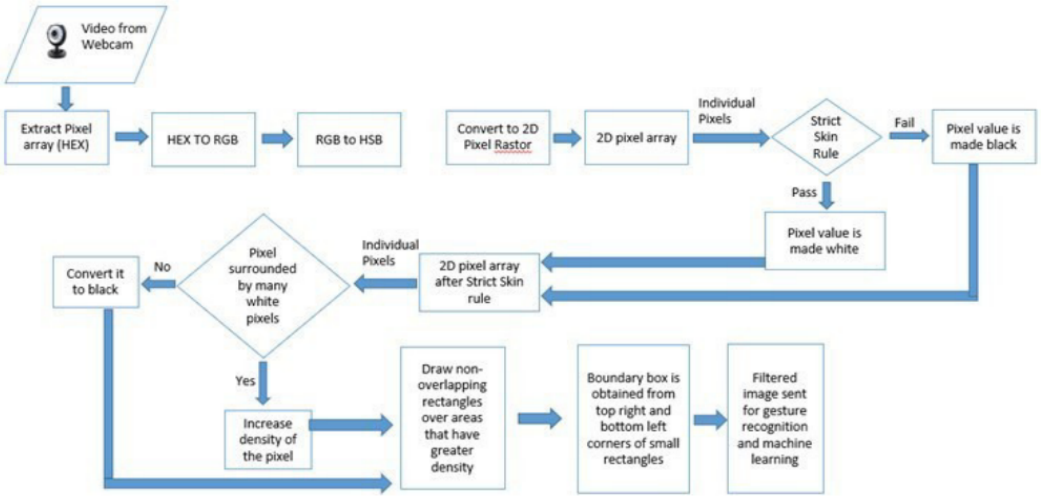

Figure 2.2: System architecture of Java (Filtering) environment

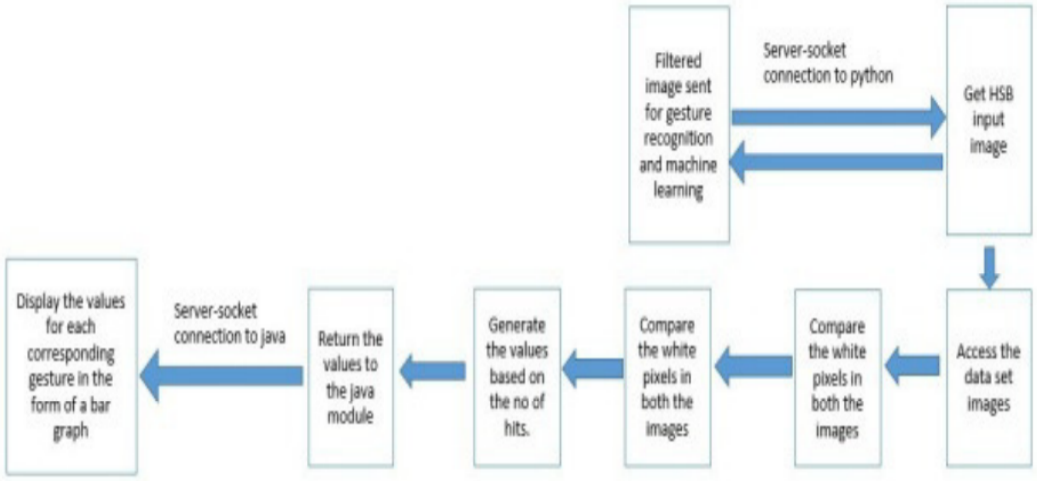

Figure 2.3: System architecture of Python (Machine Learning) environment

as PNG file . Then the neural network is then trained with the image and the prediction is done using the dataset

[4],is about creating a device that helps three kinds of people who are deaf, dumb and blind. The device helps visually challenged people by taking a image input of the text then converting it into audio . For vocally impaired people the input is taken as form of text from devices like keyboard and converting it into audio. And for audibly challenged the input is taken as audio taken in using a microphone then converting it into a text and display it. The device is build using a Raspberry Pi supported by Google API as the main unit, and also consists of a camera, microphone, speaker and a screen. The device consists of three modes and a three-way slider to change mode. Each mode is separately dedicated for the blind, deaf and dumb respectively in the device

In [5]the authors explained that for non verbal communication gestures play an important role.Pattern recognition and gesture recognition are thus so important.Hand Gesture recognition system provides us an innovative, natural, user friendly way of interaction with the computer which is more familiar to the human beings. Gesture Recognition has a wide area of application including human machine interaction, sign language, immersive game technology etc. In this paper the authors aims to present a real time system for hand gesture recognition on the basis of detection of some meaningful shape based features like orientation, centre of mass (centroid), status of fingers, thumb in terms of raised or folded fingers of hand and their respective location in image. The approach introduced in this paper is totally depending on the shape parameters of the hand gesture.

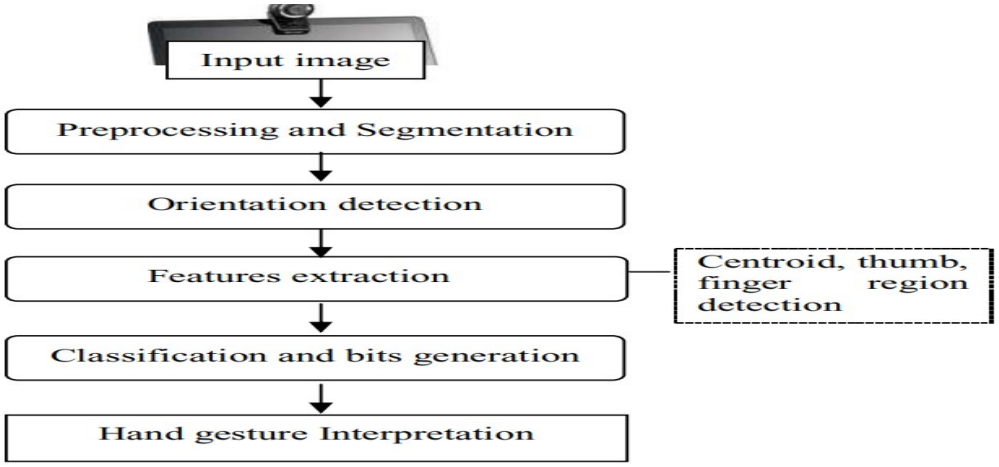

Figure 2.4: Flowchart of the Algorithm proposed in [5]

In [6], authors pointed out that CNN has been used for various applications like image segmentation, classification,recognition, semantic segmentation. Various CNN architectures have been introduced for increasing the performance of the models.Famous architectures are VGG-19, VGG-16 ,AlexNet.These architectures are trained using he dataset ImageNet and can be imported for using in any machine learning applications. These include tasks like object recognition , classification.The model was trained using VGG-19 and AlexNet architecture. Also for comparison the model was trained with SVM(Support Vector MAchine) classifier. Results of these three aprroaches where taken. VGG_19 showed greater % improvement in recall, precison, and f1score.

The tremendous growth in the domain of deep learning has helped in achieving breakthroughs in computer vision applications especially after convolutional neural networks coming into the picture.Use of a convolutional neural network to reduce the feature extraction process and parameters being used has been discussed in [7].Gesture recognition is a computer science technology that helps a user in interacting with their digital devices using simple and natural body gestures. Gesture recognition technology can be beneficial at many places like automated home appliances, hand signal interpretation, automobiles, etc. Hand gesture recognition is a part of gesture recognition that is based on recognizing the movements of hands meant to be delivered, for example: showing a forefinger could denote the number “1” or a thumbs up could be an indication of agreement.

[8] demonstrated a description for Automatic sound recognition. It has received heightened research interest in recent years due to its many potential applications. These include automatic labeling of video/audio content and real-time sound detection for robotics. While image classification is a heavily researched topic, sound identification is less mature.Here the advantage of the robust machine learning techniques used for developing images are taken for classification and apply them on the sound recognition problem. Raw audio data from the Freesound Dataset (FSD) provided by Kaggle is first converted to a spectrogram representation in order to apply these image classification techniques. Then testing and comparing the two approaches using deep convolutional neural networks (CNNs): 1.) Using CNN architecture 2.) Transfer learning using the pre-trained VVG19 network.

Chapter 3

Proposed System

This system consists of 2 models which focuses on communication between a differently-abled and . Gesture shows an expressive movement of body parts such as physical movements of head, face, arms, hand or body which convey some message. In order to communicate with someone blind,voice can be only used.To communicate with deaf person images are used. used. The models represents how a differently-abled person can identify what the other person is telling while communicating . We represent two method for this purpose.

3.1 Hand Signs to Text and Audio

The first model is trained with 26 characters from the American Sign Language(ASL) along with a custom sign for 'space'.Camera captures the real time hand signs and provide it to the model for recognition. Model identifies the landmarks on the image and classifies the image accordingly with the help of the The convolutional neural network (CNN).For this process the neural networks are pre-trained with dataset containing images of the characters along with some custom images.The dataset is from Kaggle.The full dataset is modified by adding landmarks to the images. The recognised hand signs are saved into a text file. This image of the text helps to communicate with deaf people.

A library called PYTTSX3 is applied to the text file for converting the words into voice. This can be used for communicating with blind people since they could only hear voices.This could be useful in a variety of applications, such as assistive

technology for individuals with speech impairments or as a means of controlling devices through hand gestures.

Figure 3.1: Hand Sign captured by the camera

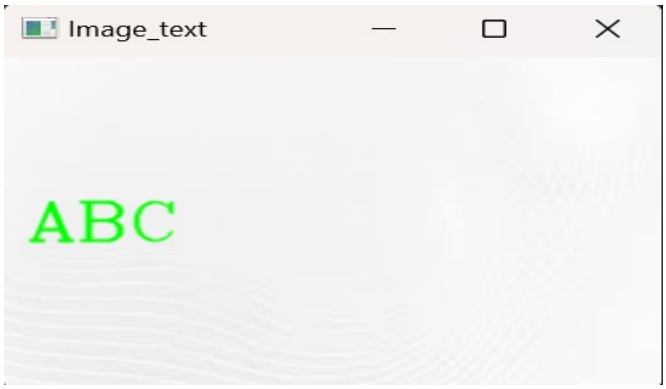

Figure 3.2: Characters saved into the text file.

3.2 Hand Gestures to Voice Speech

The second model is trained using gestures. The dataset contain 6 hand gestures including 'start','stop','left','right','yes','no'.The images for this gestures are custom created. This model help in recognition of the gestures using the convolutional neural network.The gestures identified corresponds to a particular pattern of vibration.The vibrations are produced using a adruino module which is integrated to the model.This different pattern of vibrations can be used for communicating with a person who is both deaf and blind.

3.3 Block Diagram

The visualization of the above proposed model can be represented using a block diagram.

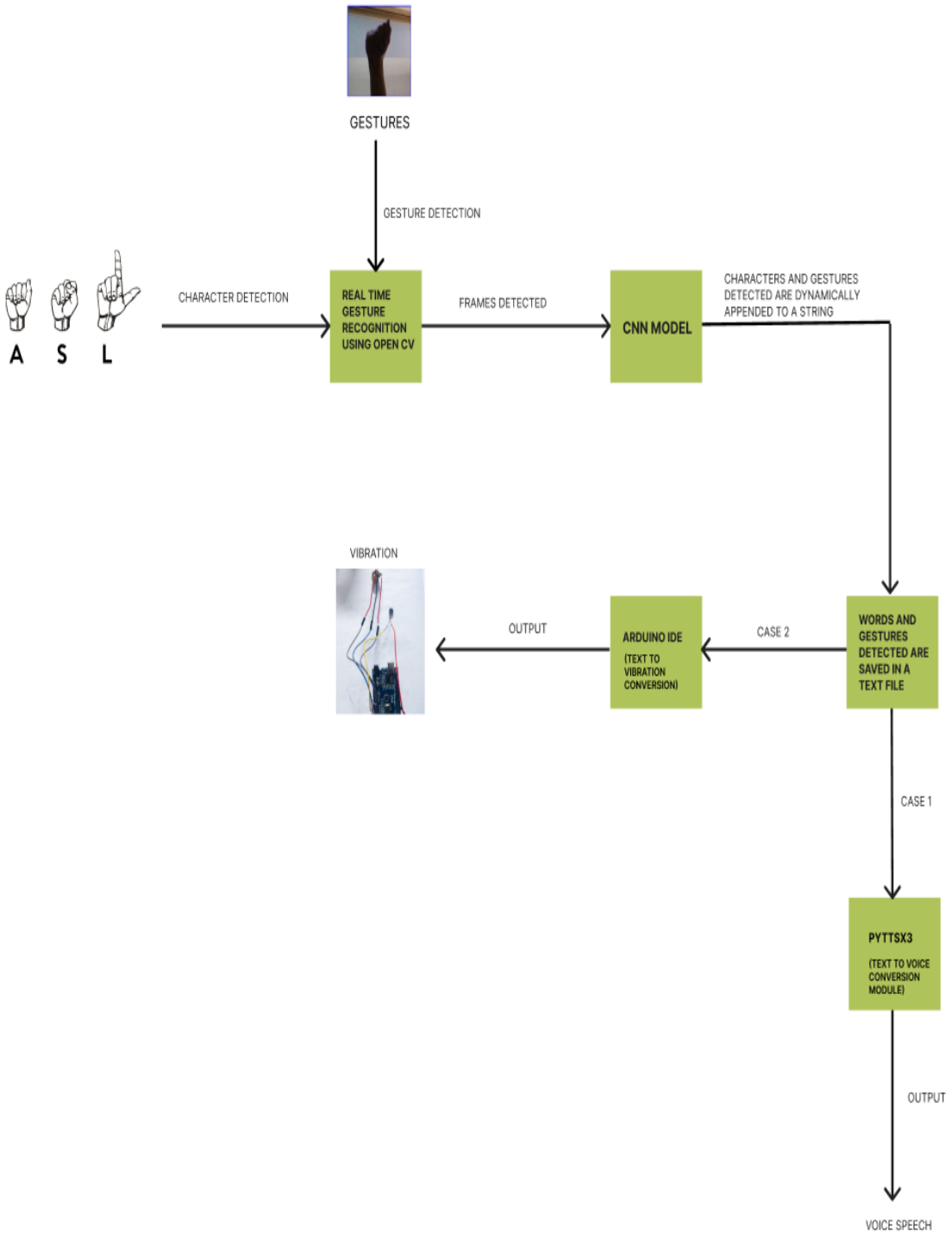

Figure 3.3: Block Diagram

3.4 Working Model

The entire working model which aims to be the end product has been visually represented below.

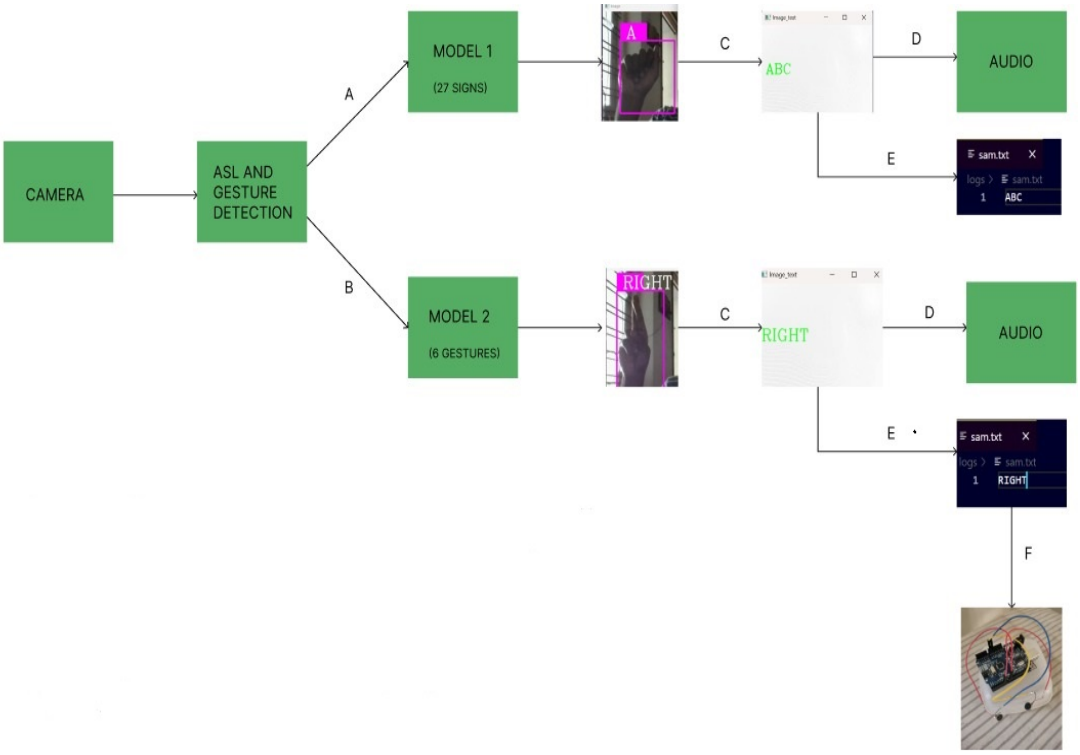

Figure 3.4: Overall working model

- A - Switch between two models. Model 1 contains 26 alphabets and a space character.
- B - Model 2 contains 6 gestures namely; LEFT, RIGHT, START, STOP, YES and NO.
- C - Detected sign/gesture saved to text box.
- D - Detected sign/gesture forms a word to output Audio.
- E - Detected sign/gesture from text box is saved to a .txt file.
- F - Saved gesture from .txt file is used to perform vibration patterns in arduino module.

Chapter 4

Implementation

4.1 Dataset

Dataset include images of American Sign Lanugae(ASL) alon with some custom made images of the hand signs/gestures. ASL is a visual-spatial language that uses handshapes,body movements,facial expressions,and other gestures to convey meaning.ASL can be used for communicating with the differently-abled community.It is the widely used and taught sign language in most of the educational institutions and workplaces.

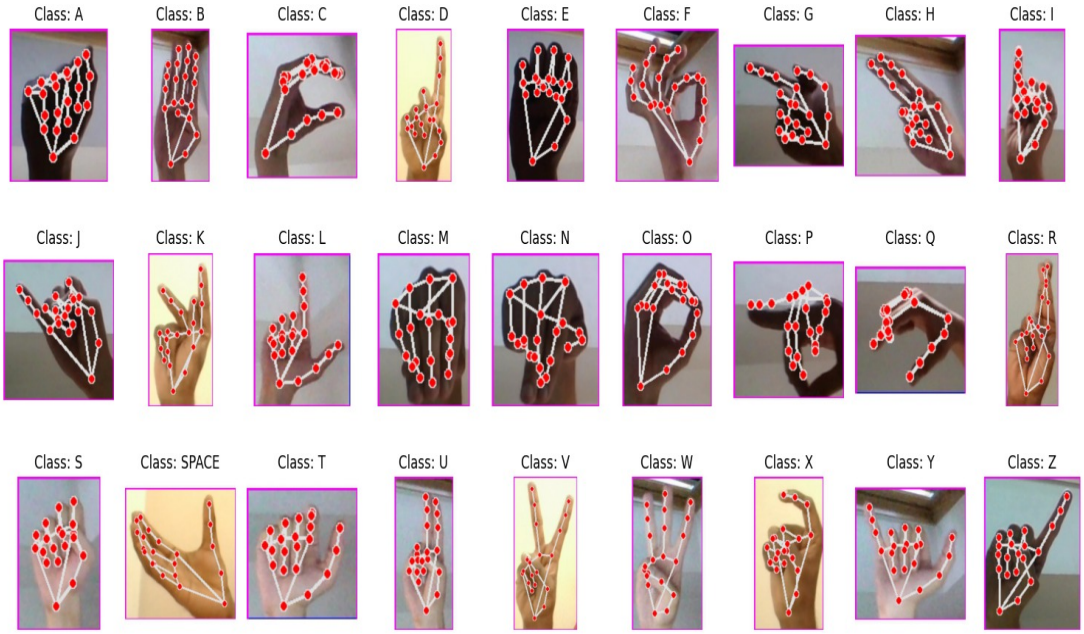

Figure 4.1: Dataset for characters : American Sign Language

Along with ASL ,custom made gesture dataset containing 6 gestures is used for identification of the gestures.

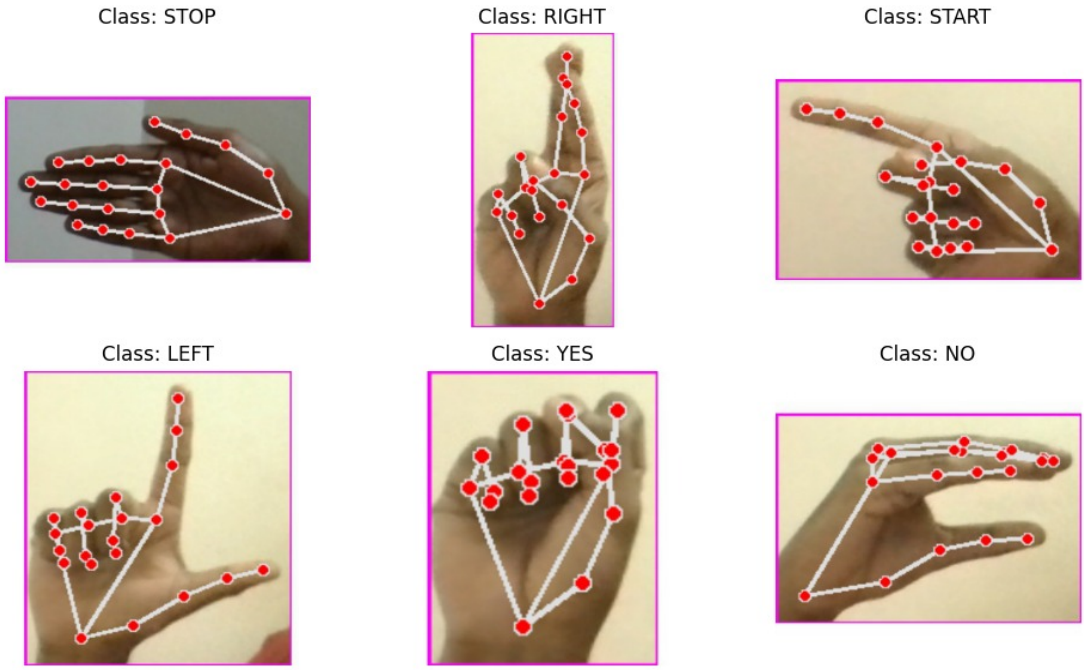

Figure 4.2: Dataset for gestures : American Sign Language

4.2 Deep Learning Models

4.2.1 Tranfer Learning Algorithm

VGG-16 is the convolutional neural network architecture which has been selected as the tranfer learning algorithm for training the models.It was developed by the Visual Geometry Group (VGG) at the University of Oxford. It is named ”VGG-16” because it consists of 16 weight layers, including 13 convolutional layers and 3 fully connected layers. VGG-16 is known for its simplicity and uniformity in design, making it easy to understand and implement.It is a pretrained model which classifies images into 1000 classes, such as ”Zebra,” ”Dalmatian,” and ”Dishwasher” in the ImageNet dataset. The model is trained using a variant of the ImageNet dataset, which consists of about 1.2 million images and 1000 classes.

Overall, the VGG-16 architecture is characterized by its deep stack of convolutional layers and relatively small filter size. This architecture has been influential in the

development of subsequent CNN architectures, and it has achieved strong performance on various image classification tasks.

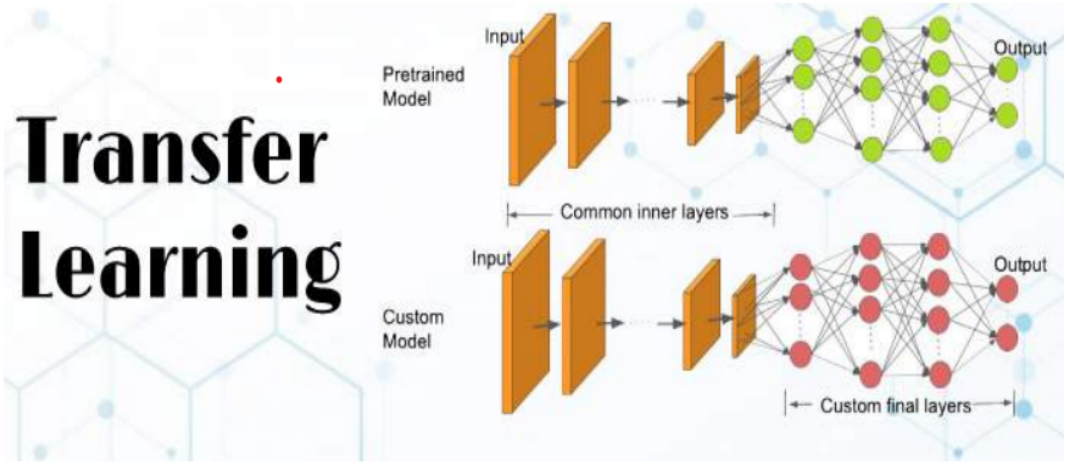

Figure 4.3: Transfer Learning Layers

4.2.2 CNN

Convolutional neural networks (CNNs) run using Google Colab, which is a cloud-based platform that allows users to develop and run machine learning models that use transfer learning.In the context of hand gesture recognition using a CNN and transfer learning, Google Colab can be used in several ways:

- Training a CNN: Google Colab provides access to powerful GPUs, which can be used to accelerate the training of a CNN for hand gesture recognition.
- Fine-tuning a pre-trained CNN: Google Colab can be used to fine-tune a pre-trained CNN on a new dataset of hand gestures, allowing us to leverage the knowledge learned by the pre-trained model on a related task.
- Extracting deep features: Google Colab can be used to extract deep features from a trained CNN, which can then be used as input to a separate classifier for hand gesture recognition.

- Training a classifier: Google Colab can be used to train a classifier on the deep features extracted from the CNN, allowing you to classify new hand gestures using the trained model.

Overall, Google Colab can be a useful tool for developing and training machine learning models for hand gesture recognition using a CNN and transfer learning.

4.3 Hardware Components

4.3.1 Adruino Uno

The Arduino Uno SMD is a version of the Arduino Uno, but uses an surface mount version of the Atmega328P instead of the through-hole version. This version was made in response to a shortage in supply of the through-hole Atmega328P. The board is based on the ATmega328. Arduino board includes a microcontroller, and this microcontroller is what executes the instructions in your program. The ATmega328 microcontroller is the MCU used in Arduino UNO as a main controller.

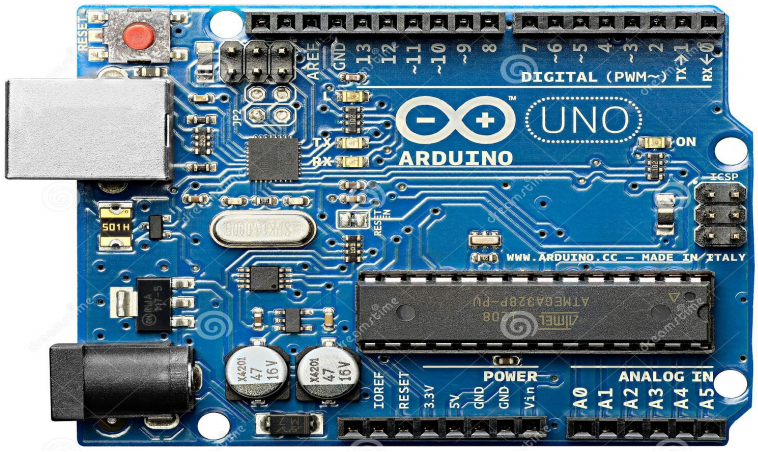

Figure 4.4: Adruino UNO

4.3.2 Adruino IDE

An integrated development environment (IDE) is a software suite that consolidates the basic tools developers need to write and test software. Arduino is a prototype platform (open-source) based on an easy-to-use hardware and software. It consists of a circuit

board, which can be programed (referred to as a microcontroller) and a ready-made software called Arduino IDE (Integrated Development Environment).

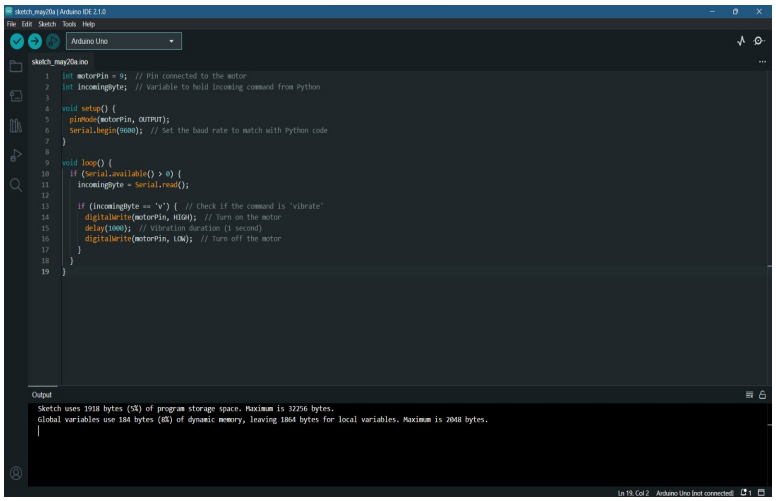

Figure 4.5: Adruino IDE

4.3.3 Vibration Motor

A compact integrated vibration motor module is a vibration motor. After providing 5V electricity, we may use a digital signal to regulate the motor's ON/OFF or vibration strength.Signals from the Colab's geture sensors trigger the server by data publish that sends impulse signals to the vibration motor attached to the microcontroller. The motor operate vibrating corresponds to the predicted class of image.

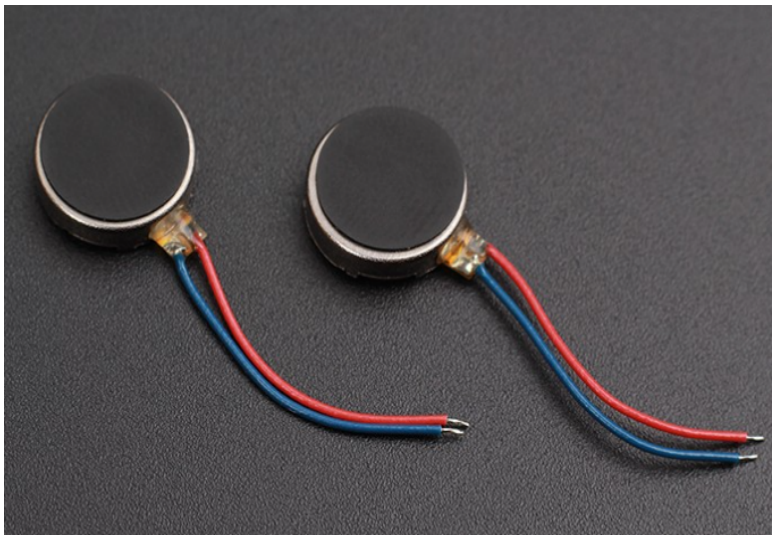

Figure 4.6: Vibration Motor

Chapter 5

Results

The proposed models recognized and classified the hand gestures/signs of ASL images.There are 27 classes for the images of ASL characters,and 6 classes for the gestures .They all are recognized and classified successfully.The model is helpful for differently-abled people,it says how they can identify the gestures through CNN model using Transfer Learning concept.

5.1 Software Evaluation

The first model classifies the captured images to corresponding hand signs . This signs are saved into a text file.For finding the best prediction algorithm we conducted a comparison study of 2 algorithms.

5.1.1 Comparison of model parameters

Among the parameters the comparison done based on the trainable parameters. Inception v3 model had way less trainable parameters when compared to VGG-16 model but eventhough the model is complex in inception v3. And the convergence to minimal loss solution got perfectly in VGG-16 model eventhough the model took more time to train. After experimenting with both of these model VGG-16 was found to be the optimal architecture for extracting the features and providing an accurate model.

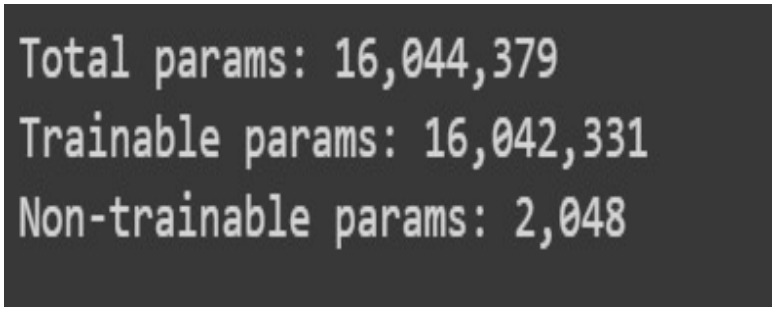

Figure 5.1: Parameters of VGG16 Algorithm

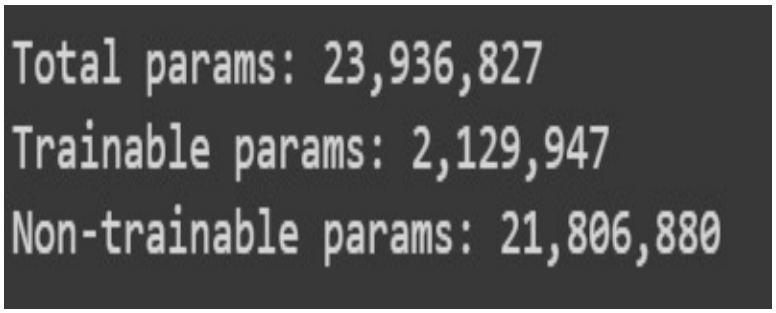

Figure 5.2: Parameters of InceptionV3 Algorithm

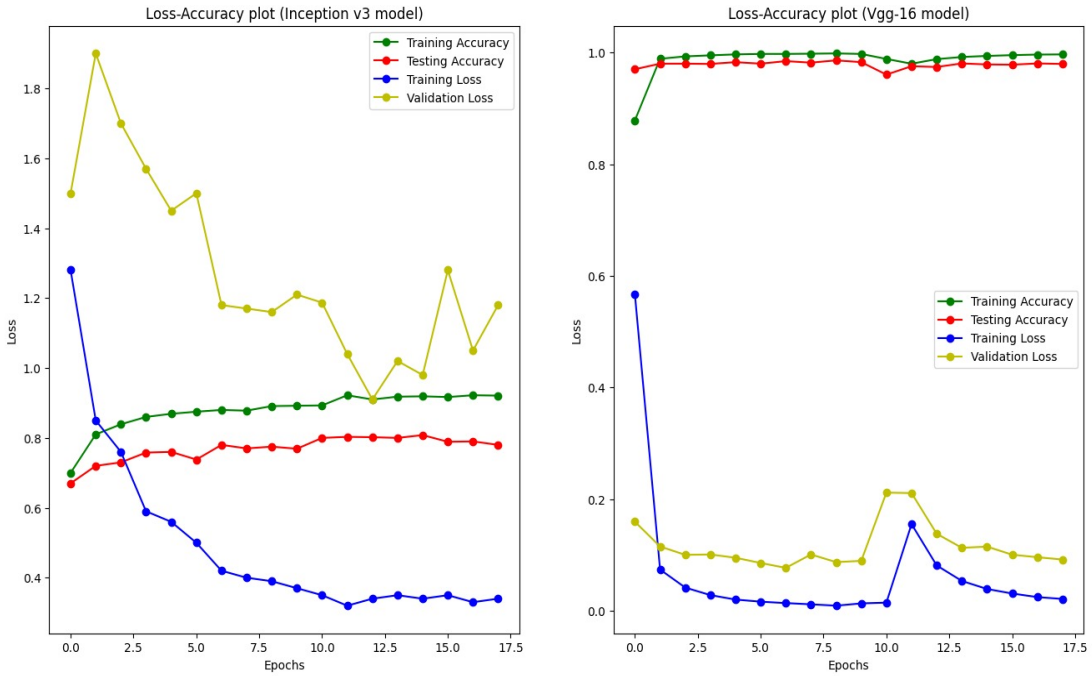

Figure 5.3: Comparison between VGG16 and InceptionV3

5.2 Hardware Evaluation

The proposed second model correctly identifies the captured gestures.It recognizes the image and saves the corresponding gesture to a text file.The file is then provided as

an input to the arduino module which in turn produces various vibration pattern using 2 vibration motor V_1andV_2 which corresponds to left and right hands respectively.The vibrations patterns for the gestures are as follows :

- 'Right' : one vibration at V_2
- 'Left' : one vibration at V_1
- 'Yes' : two vibration at V_2
- 'No' : two vibration at V_1
- 'Start' : three vibration at V_2
- 'Stop': three vibration at V_1

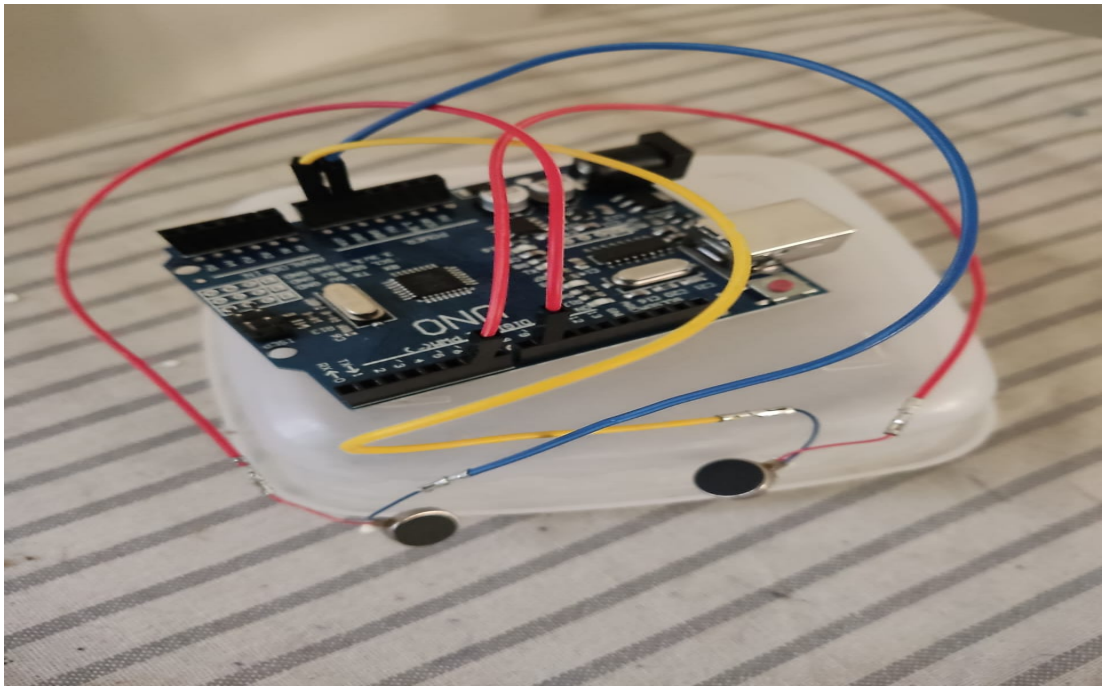

Figure 5.4: Vibration Module

Chapter 6

Conclusion

In conclusion, the system described aims to improve hand gesture recognition for the deaf and blind community. Hand gesture recognition systems are needed to be improved significantly for communication. Transfer learning using deep features extracted by CNN is used for Hand gesture recognition. The main aim of our system is to accurately identify the input hand signs to provide corresponding audio/text output and hand gestures to provide corresponding vibration pattern.

There are 2 models for the hand sign/gesture recognition.We have trained and evaluated the model on the basis of existing available dataset(American Sign Language) along with some custom made images.

The first model accepts hand sign images and stores it inside a text file.With the help of a python library PYTTSX3, the text is converted into an audio output for blind people. For deaf people the text file can be used for communicating.

The second model takes custom made images of the hand gestures and identifies each gestures.The gestures are saved into a text file.For each gesture saved into the text file, a pattern of vibration is created using a arduino module.

By combining image recognition, text conversion, audio output, and vibration patterns, this system aims to bridge the communication gap for the deaf and blind community. It provides both audio and tactile feedback for the recognized hand signs and gestures, enabling effective communication for individuals with sensory impairments.

References

[1] Reza Azad, Babak Azad and Iman tavakoli kazerooni, ”Real-Time and Robust Method for Hand Gesture Recognition System Based on Cross-Correlation Coefficient”.

[2] Pei Xu, Department of Electrical and Computer Engineering, University of Minnesota, Twin Cities, ”A Real-time Hand Gesture Recognition and Human-Computer Interaction System”.

[3] P. Suganya, R. Sathya, K. Vijayalakshmi. “Detection and Recognition of Gestures to Control the System Applications by Neural Networks.” International Journal of Applied Mathematics, vol. 118, no. 10, pp. 399-405, January 2018.

[4] Karmel A, Anushka Sharma, Muktak pandya, Diksha Garg, “Iot based assistive device for deaf ,dumb, blinf people” , International Conference on recent trends in advanced computing 2019,ICRTAC 2019

[5] M. Panwar, ”Hand Gesture Recognition based on Shape Parameters”, in Proceedings of IEEE International Conference, 2012, pp. 1-6

[6] Manali Shahal, Meenakshi Pawar, “Transfer learning for image classification”, Proceedings of the 2nd International conference on Electronics, Communication, 2018

[7] Ishika Dhall, Shubham Vashisth, Garima Aggarwal , ”Automated Hand Gesture Recognition using a Deep Convolutional Neural Network model” , 10th International Conference on Cloud Computing, 2020

[8] Boyang Zhang, Jared Leitner, Sam Thornton , “Audio Recognition using Mel Spectrograms and Convolution Neural Networks” ,IEEE Signal Processing Letters 24.3(2018).