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

Ananthakrishnan KM[1] Arjun S [2] Sanora Teressa Raju [3]
Vishnu V [4]

Abstract

Our project focuses on developing a real-time gesture and sign recognition model using transfer learning techniques, specifically using the VGG-16 algorithm. By utilizing the power of transfer learning, we are able to build upon the knowledge and pre-trained weights of the VGG-16 model, significantly reducing the amount of training data required and accelerating the model development process.

The primary objective of this model is to facilitate communication between differently-abled individuals and others. To achieve this, we constructed two distinct models: one dedicated to identifying signs and the other to recognizing gestures. Both models incorporate significant modifications to the last layers of the VGG-16 architecture.

The system provides three essential outputs. Firstly, for individuals who are deaf, we generate image outputs that visually represent the recognized gestures or signs, ensuring effective communication through visual cues. Secondly, for individuals who are blind, the system generates audio outputs, allowing them to perceive and understand the recognized gestures or signs through sound. Lastly, we incorporated an Arduino-based vibration mechanism that translates the recognized gestures/signs into physical feedback. This feature serves to assist both deaf and blind individuals, enhancing their ability to comprehend and respond to communication.

By utilizing transfer learning and adapting VGG-16, our real-time model offers a versatile and efficient solution for gesture and sign recognition. The integration of image, audio, and physical vibration outputs ensures inclusive communication for differently-abled individuals, bridging the gap between them and the wider community.

Literature review

The hand gesture recognition has several uses in computer games, virtual reality, and sign language interpretation. Orientation and movement of the hands and facial expressions can be combined to represent thoughts and words in sign language, which has been primarily used for communication by Deaf and Blind people [0].

Research gaps

The field of computer vision become more developed and emerged after the development of deep learining and CNN. CNN has been used to lessen the trainable parameters and the process of feature extraction become much for simpler. This is being discussed in[0]

Research objectives

The present study investigates the following objectives:

- **Objective 1:** To create a model that recognizes sign/gestures and convert it into voice and image for deaf people, blind people respectively.
- **Objective 2:** To create a model that recognizes sign/gestures and convert it into vibrations for people who are both deaf and blind.

Study methodology

Visualization of the model using a block diagram

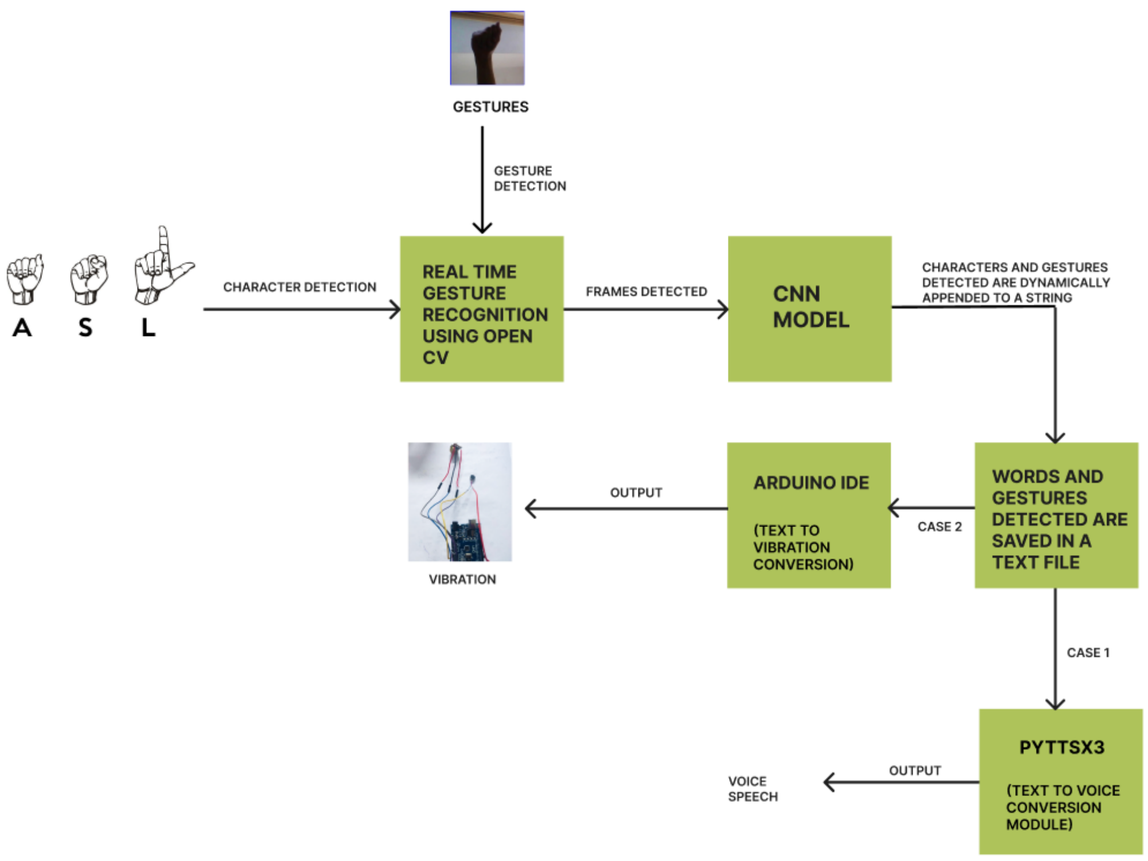

Figure 1. Block Diagram

Descriptive statistics

- Precision: Precision measures the proportion of correctly predicted positive instances out of all instances predicted as positive. It indicates the model's ability to avoid false positives
- Recall: Recall measures the proportion of actual positive instances that are correctly identified by the model. It indicates the model's ability to avoid false negatives.
- F1 Score: The F1 score is the harmonic mean of precision and recall. It provides a balanced measure that considers both precision and recall.
- Accuracy: Accuracy measures the overall correctness of the model by calculating the proportion of correctly predicted instances out of all instances.

Results and discussion

Model estimates

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.

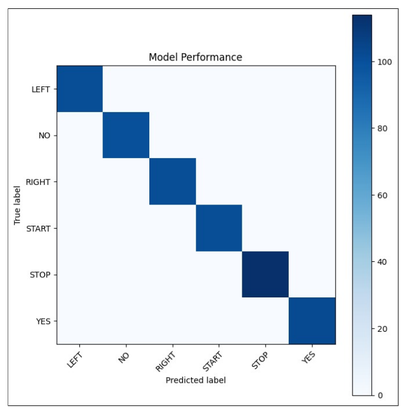

Figure 2. Confusion Matrix of Gestures

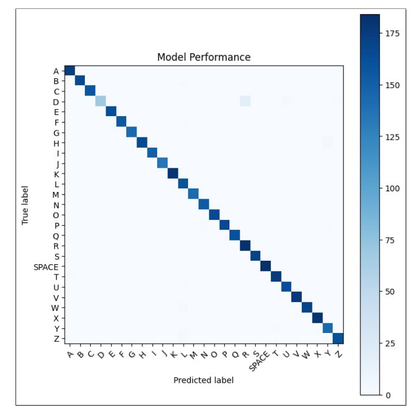

Figure 3. Confusion Matrix of Signs

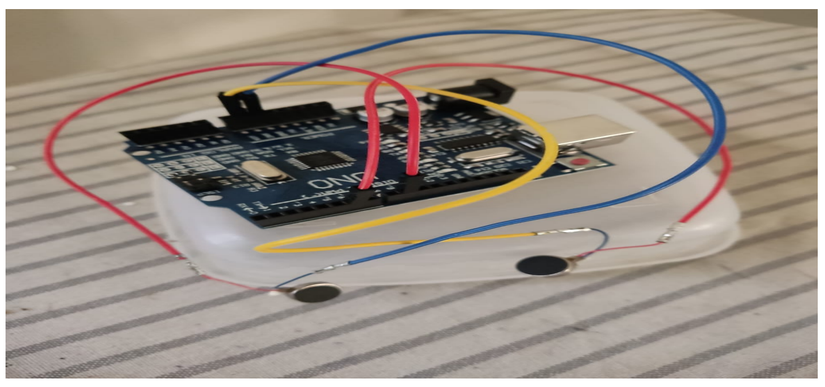

Figure 4. Vibration Module

Conclusions

- The system described aims to improve hand gesture recognition for the deaf and blind community.
- 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 an arduino module.

What is already known about this subject?

- Sign Language recognition using deep learning models that work on skin masked images.
- Using Artificial Intelligence,motion sensors or hand gloves are used to detect the positions of different fingers accurately.
- Use of pose estimation models such as OpenPose for sign language recognition

What does this study add?

- Data collection using Mediapipe approach.
- Transfer Learning model for Sign recognition.

Practical implications

Real world applications include:

- Easy communication with deaf and blind people using voice and image respectively.
- Easy communication with people who are both deaf and blind using vibration patterns.

References

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

[2] Babak Azad Reza Azad and Iman tavakoli kazerooni. "real-time and robust method for hand gesture recognition system based on crosscorrelation coefficient".