Mini Project Presentation

Sign Language Recognition Using CNN

- N180164 - K.Satish
- N180122 - V.Bhargavi
- N180632 - K.Srisha
- N180431 - E.Jeswanth
- N180353 -Y.Ajay Kumar

Table of Content

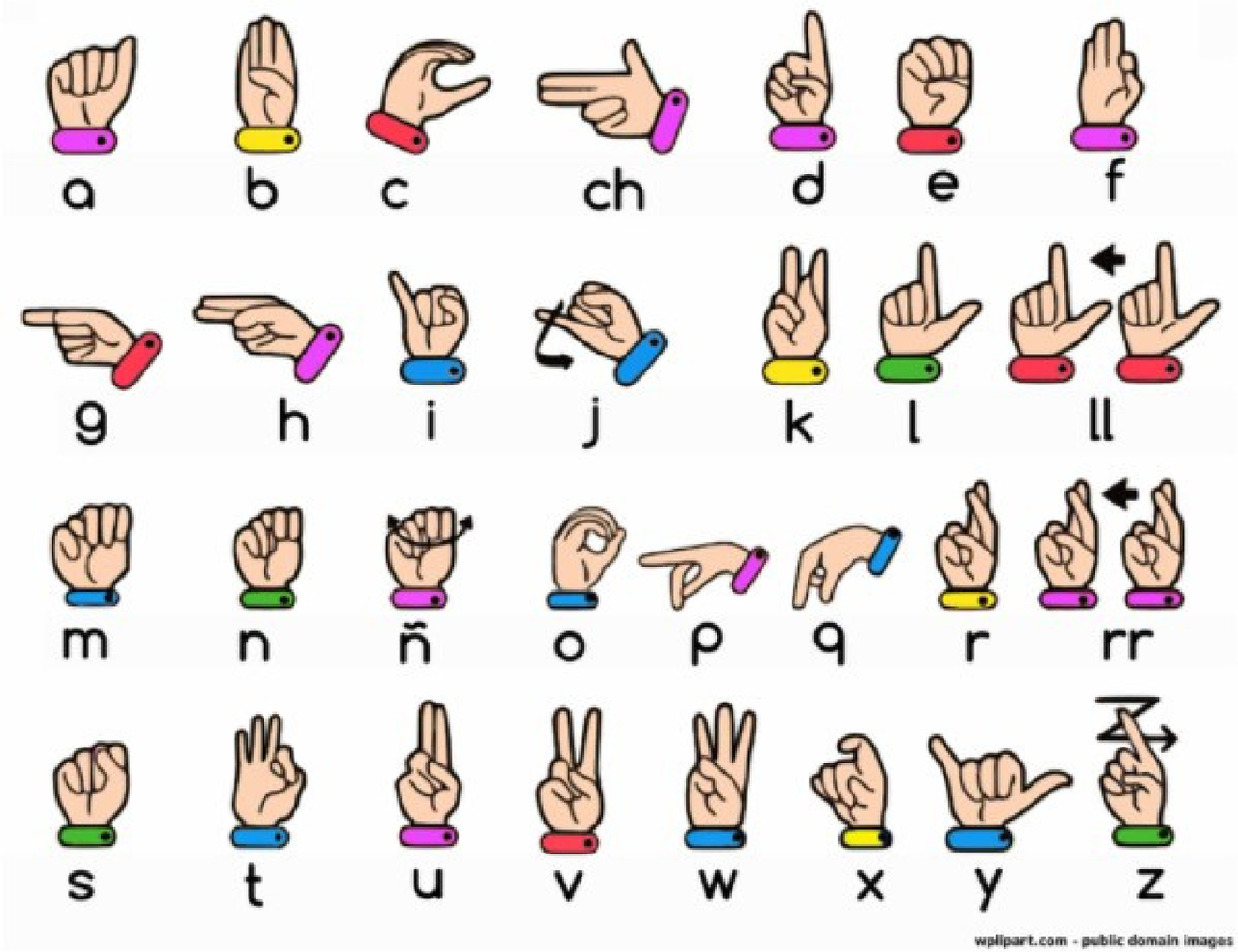
Abstract

We have developed a method using neural networks to recognize finger spelling in American Sign Language in real time. Sign language is a natural and ancient form of communication, and our goal is to use computer vision techniques to automatically recognize hand gestures captured by a camera. To achieve this, we propose using a convolutional neural network (CNN) to analyze the position and orientation of the hand in an image. The CNN is trained and tested using data obtained from the hand's position and orientation. The hand is first processed through a filter, and then passed through a classifier that predicts the type of hand gesture. The images used to train the CNN are adjusted for accuracy.

Introduction

American Sign Language is the main way of communication for the people who are deaf and dumb. Since they can't use spoken language, they rely on sign language to express themselves. Communication is how we share our thoughts and messages using different methods like speaking, signals, behavior, and visuals. Deaf and dumb people use their hands to make gestures and convey their ideas to others. These gestures are a form of nonverbal communication that relies on vision. It's known as sign language, and that's how deaf and dumb people communicate without using words.

American Sign Gestures



https://www.wpclipart.com/sign_language/Spanish_sign_language_alphabet.png.html

Related Works/Existing Works

Few existing works or approaches in sign language recognition are :

Glove-based Sensor Systems: In this system a specially designed gloves are equipped with sensors. These sensors are strategically placed on the glove to track the movements of the user's hand and fingers. By capturing data, this systems aim to accurately represent the hand configuration during sign language gestures.

https://www.degruyter.com/document/doi/10.1515/comp-2022-0240/html?lang=en

Related Works/Existing Works

Electromechanical Data Extraction: Once the sensor-equipped glove captures the hand movements, electromechanical devices are used to extract the relevant information. Algorithms like Gesture Recognition Algorithms are employed to interpret the sensor data and extract meaningful features that represent the sign language gestures.

Related Works/Existing Works

Challenges with Glove-based Approaches:

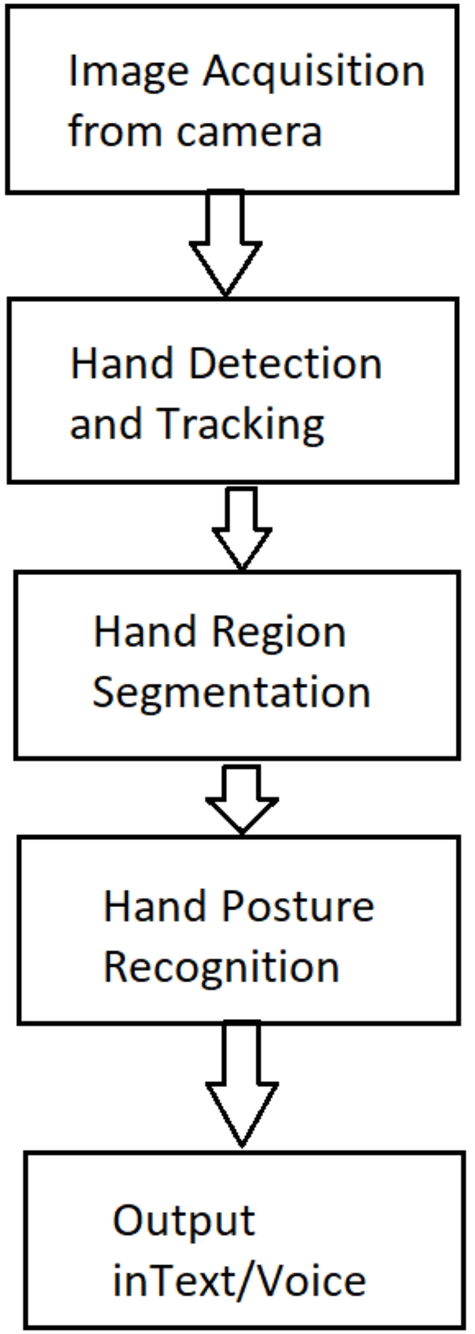
Glove-based sign language recognition approaches are limited by cost due to expensive sensors , as well as user-friendliness challenges caused by the need for consistent wearing a glove regularly for everyday use.

Our aim is to develop a user-friendly human computer interface (HCI) that recognizes American sign language.

This Project will help the dumb and deaf people by making their life easy.

Proposed Method

System Flowchart:



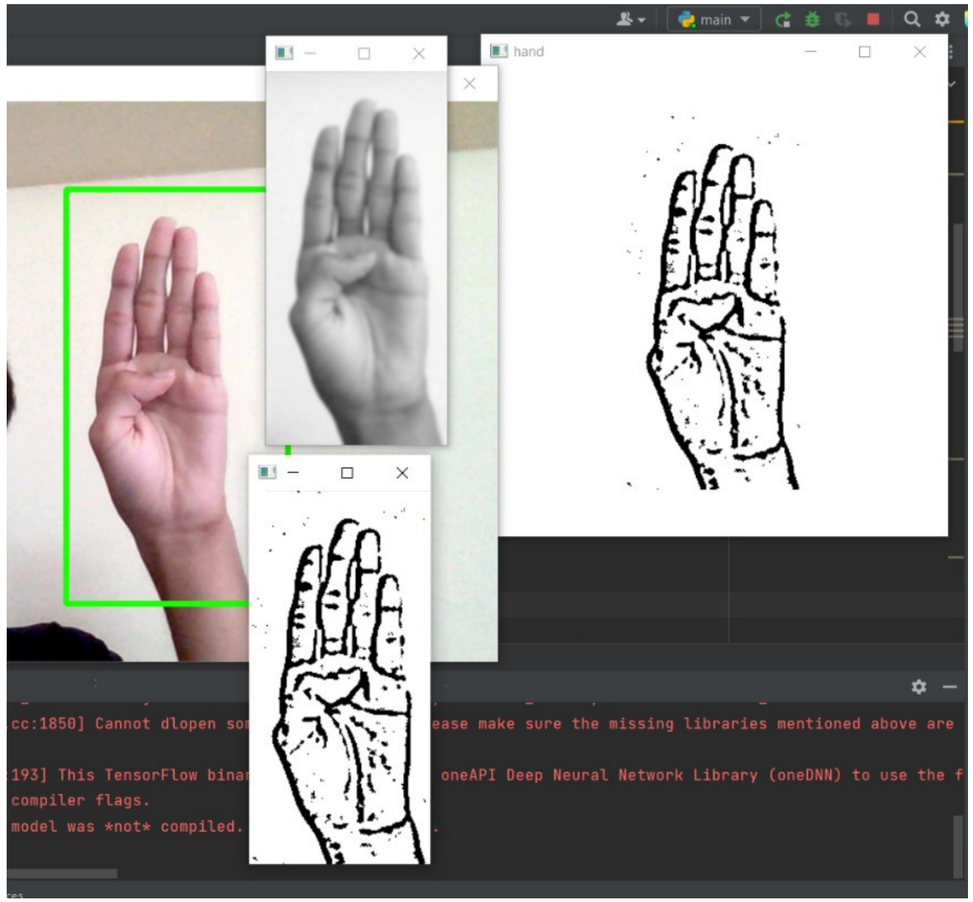
Data Acquisition

We have used Vision-Based approach to collect the data about hand gestures. This approach is advantageous due to its cost-effectiveness and user-friendliness, as it solely relies on a camera without the need for additional devices. By analyzing the visual input from the camera, information about the positions and movements of hands and fingers can be extracted.

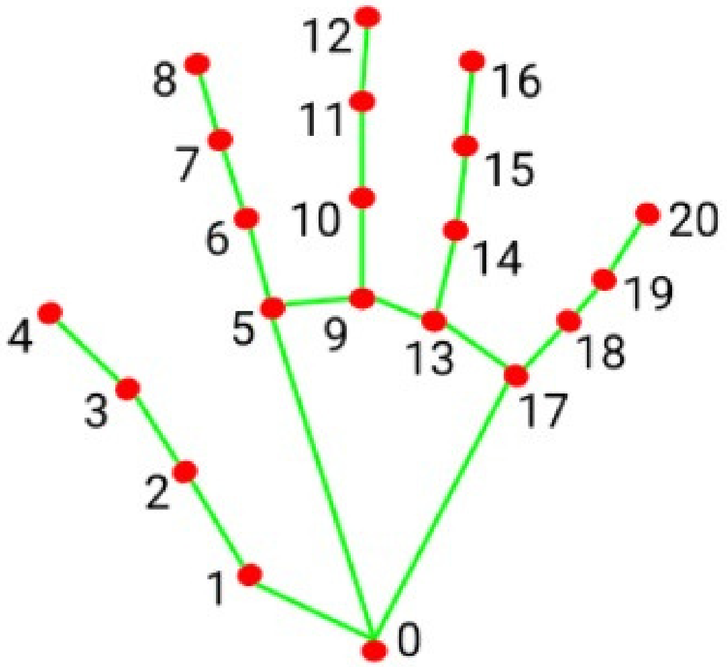
Data Pre-Processing and Feature Extraction

In this method for detecting hands, we start by using a webcam to capture an image. To identify the hand in the image, we utilize a library called MediaPipe, which helps with image processing. Once we have detected the hand, we focus on a specific area of the image called the region of interest (ROI). We crop this area and convert it to a gray-scale image using another library called OpenCV. To enhance the image, we apply a technique called Gaussian blur using OpenCV. Then, we further simplify the image by converting it to a binary image using methods called thresholding and adaptive thresholding. These steps help us analyze the hand more effectively.

Data Pre-Processing and Feature Extraction



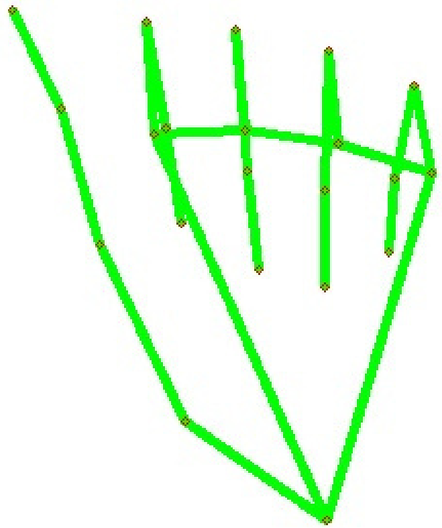
Mediapipe Landmark System



0. WRIST
1. THUMB_CMC
2. THUMB_MCP
3. THUMB_IP
4. THUMB_TIP
5. INDEX_FINGER_MCP
6. INDEX_FINGER_PIP
7. INDEX_FINGER_DIP
8. INDEX_FINGER_TIP
9. MIDDLE_FINGER_MCP
10. MIDDLE_FINGER_PIP
11. MIDDLE_FINGER_DIP
12. MIDDLE_FINGER_TIP
13. RING_FINGER_MCP
14. RING_FINGER_PIP
15. RING_FINGER_DIP
16. RING_FINGER_TIP
17. PINKY_MCP
18. PINKY_PIP
19. PINKY_DIP
20. PINKY_TIP

https://www.researchgate.net/figure/The-order-and-labels-for-keypoints-that-exist-in-the-hands-of-MediaPipe-30-For-Pose_fig2_364279614

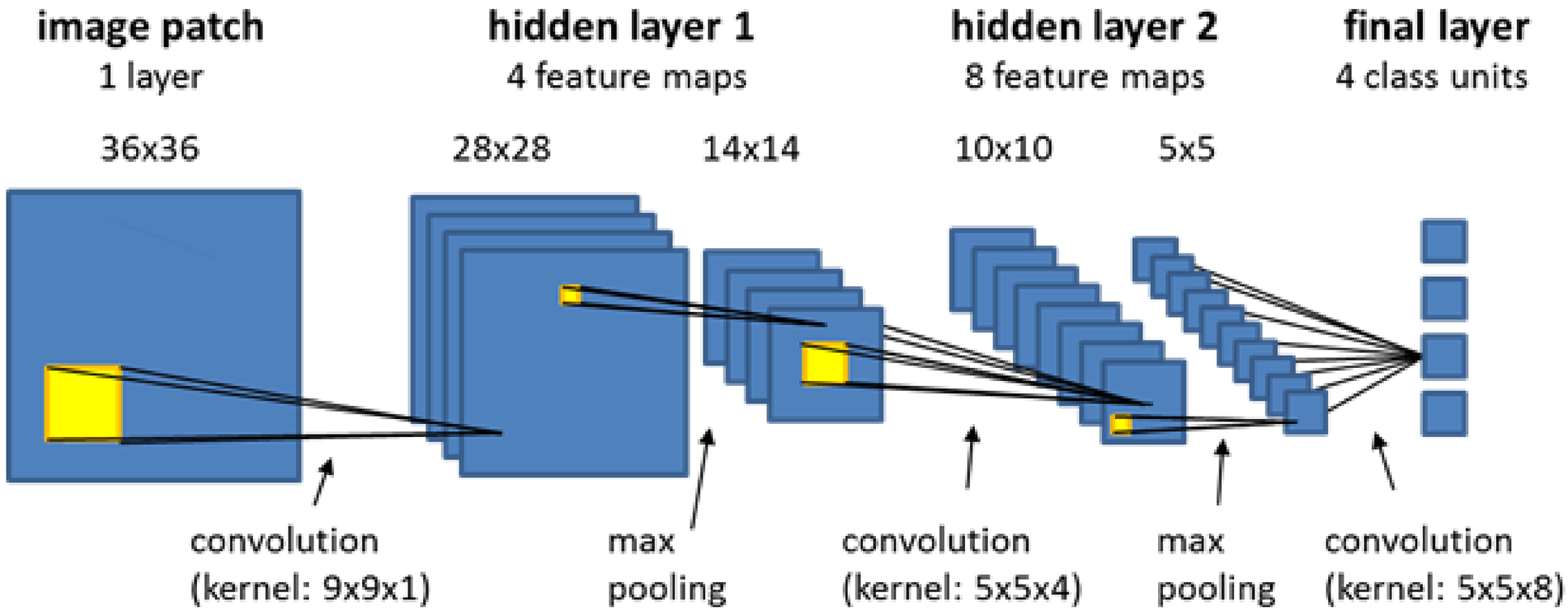
Mediapipe Landmark System



Convolutional Neural Network (CNN)

CNN (Convolutional Neural Network) is a deep learning model commonly used for image recognition. Inspired by the visual cortex in the brain. The network architecture includes layers like convolution, max pooling, flatten, and fully connected layers. These layers work together to identify features in an image.

Convolutional Neural Network (CNN)



https://jacobheyman702.medium.com/different-pooling-layers-for-cnn-4652a5103d62

Some key Terms to Know

Kernal / Filter : The kernel is a matrix that moves over the input data, performs the dot product with the sub-region of input data, and gives output as the matrix of dot products.

Padding : Padding involves adding extra pixels around the border of the input feature map before convolution. It allows more accurate analysis of images.

Stride: Stride denotes how many steps we are moving in each steps in convolution.By default it is one.

Convolutional Layer

This is the first layer in CNN. A convolution layer applies set of filters or kernels to input data. By convolving these filters across the input, it extract features. It captures local patterns and hierarchically learns complex features in images. The size of the kernels/filters is usually smaller than the actual input size.

Pooling Layer

The purpose of the pooling layers is to reduce the dimensionality of the matrix by preserving features in it. It reduces the computational cost. It is not always necessary to perform pooling after convolutional layer.

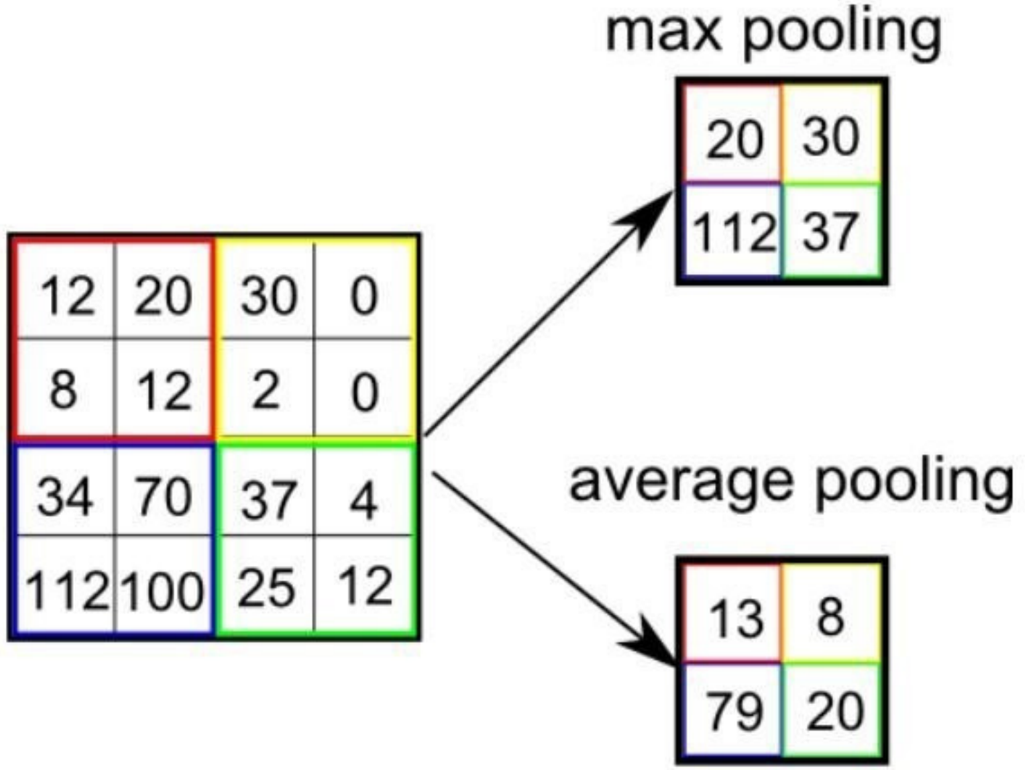
a. Max Pooling:

Max pooling is a pooling operation that selects the maximum element from the region of the feature map covered by the filter.

Pooling Layer

b. Average Pooling:

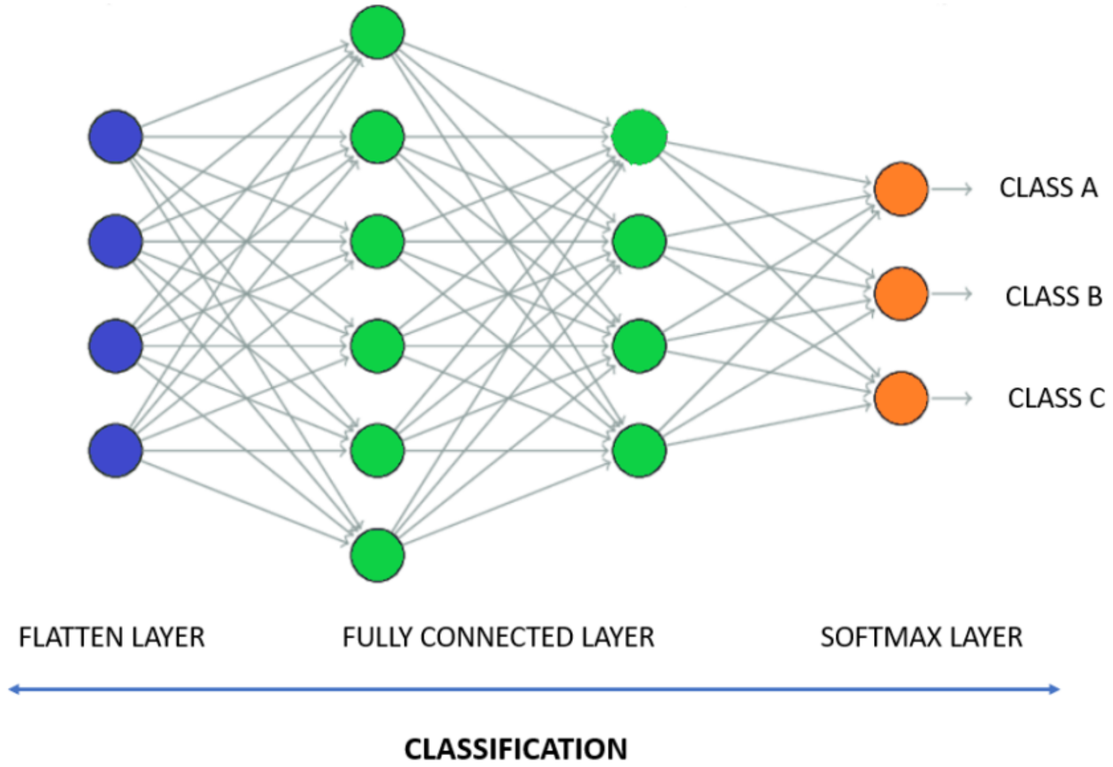
Average pooling computes the average of the elements present in the region of feature map covered by the filter.



https://poojamahajan5131.medium.com/max-pooling-210fc94c4f11

Fully Connected Layer

Fully Connected layer takes input from Flatten Layer which is a one-dimensional layer (1D Layer). It classifies the output.



https://indiantechwarrior.com/fully-connected-layers-in-convolutional-neural-networks/

Text To Speech Translation

we have used pyttsx3 library to convert the recognized words into the appropriate speech. The text-to-speech output is a simple workaround, but it's a useful feature because it simulates a real-life dialogue.

Libraries involved

1. Open CV
2. Mediapipe
3. CVzone
4. Traceback
5. Keras
6. Pyenchant
7. PIL
8. Pyttsx3
9. Tkinter
10. Numpy
11. OS,sys
12. Math

Libraries

1. OpenCV: OpenCV is an open-source computer vision library that provides tools for for image and video processing tasks.

2. **Mediapipe**: Mediapipe is a Google-developed open-source framework for real-time multimedia processing, providing modules for tasks like hand tracking, pose estimation, and facial recognition.

3. **CVZone**: CVZone is a library that extends the capabilities of OpenCV, providing additional functionality and tools for image and video processing tasks.

Libraries

4. **Traceback:** Traceback is a Python library that provides tools for printing error messages, for debugging purposes during program execution.

5. **Keras:** Keras is a high-level deep learning library that simplifies the process of building , training and evaluating deep learning models.

6. **PyEnchant:** PyEnchant is a library for spell checking and suggesting corrections in Python.

Libraries

7. **PIL(Python Imaging Library):** PIL is a popular library for image processing and manipulation like opening, editing, and saving images in Python.

8. **pyttsx3:** pyttsx3 is a Python library for text-to-speech conversion. It provides a simple and convenient way to generate speech from text.

9. **Tkinter:** Tkinter is a standard Python library for creating graphical user interfaces (GUIs). It provides a set of tools to build Windows, buttons, menus, and other GUI elements for desktop applications.

Libraries

10. **NumPy :**NumPy is a Python library for efficient numerical computations, providing arrays, matrices, and mathematical functions for scientific computing tasks.

11. **Math:** The math module in Python provides mathematical functions and operations for tasks like arithmetic, trigonometry and logarithms.

Libraries

12. **OS**: The OS module in Python provides functions for interacting with the operating system, allowing tasks like file and directory operations.

sys : The sys module in Python allows interaction with system-specific features and functions, such as command-line arguments and interpreter-related operations.

Dataset

In this project, we use a CNN model to extract features from an image.The CNN model is saved with a '.h5' extension, which indicates that it is stored in the Hierarchical Data Format (HDF5) file format. HDF5 is commonly used for managing and storing large numerical data sets. In a CNN model,the HDF5 file typically contains the architecture, weights, and other parameters of the model. We utilize the Keras library to work with this CNN model.

Hardware And Software Requirements

Hardware Requirement: webcam

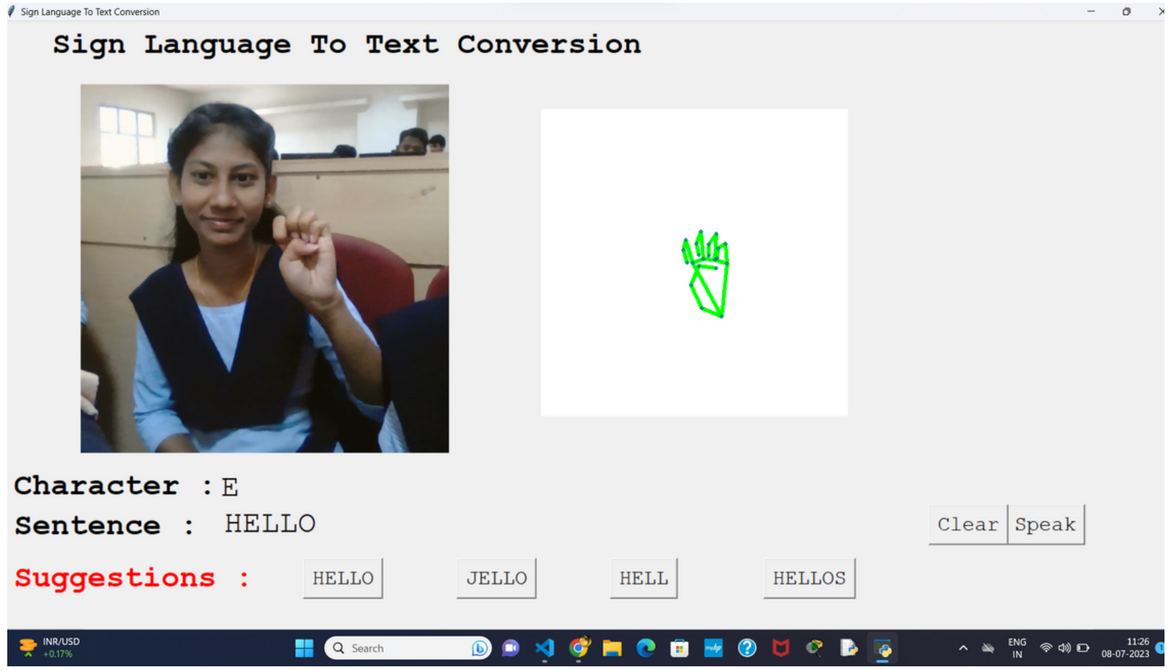
Software Requirement:

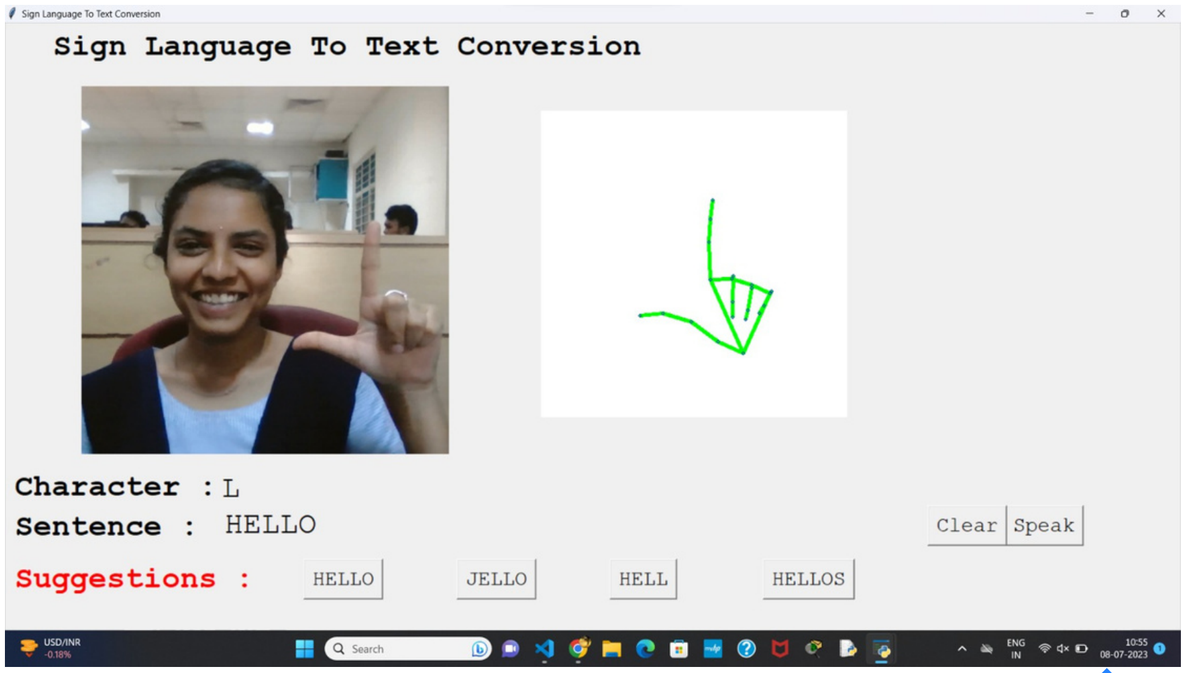
Opertaing System: Windows 8 and Above

IDE: visual studio

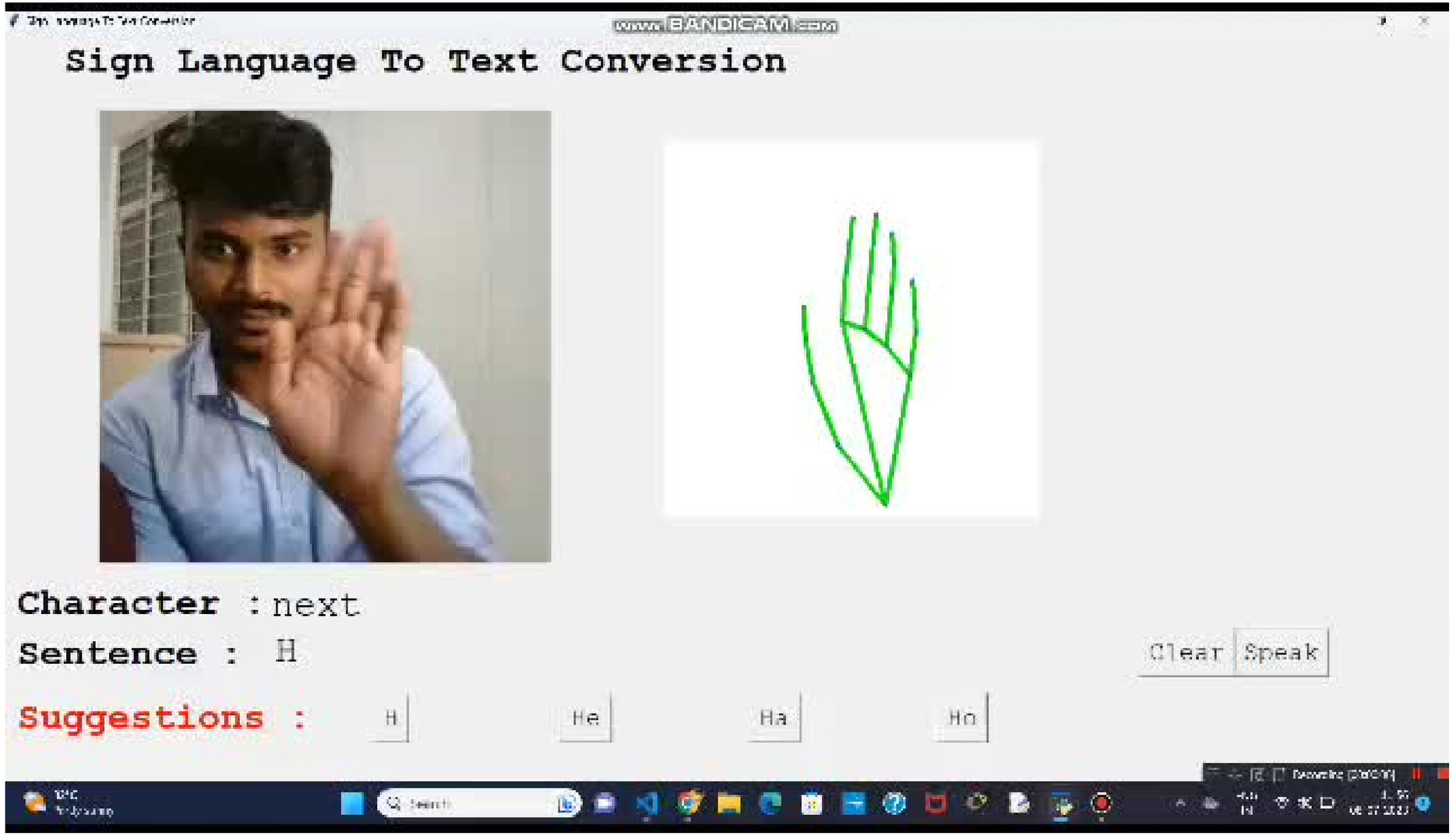
Programming Language: Python 3.9.5

Result Screenshot





Result Video



Results Comparision With Other Approaches

During the systematic literature review, we observed that the majority of research papers on sign language recognition using glove-based systems achieved an average accuracy of greater than 90% with a small number of gestures.

In our project, we have used vision based approach to recognize finger spelling in American Sign Language using convolutional neural network. Accuracy of this vision-based approach is 97% and it is also less expensive and more user-friendly when compared to glove-based approach.

Conclusion

Development of a user-friendly Human Computer Interface (HCI) that understands American Sign language holds significant potential for enhancing the lives individually who are deaf and dumb.By leveraging techniques such as CNN ,we can achieve accurate and gesture recognition and interpetion.

Future Scope

Deployment in Real-world Scenarios: Test and deploy the system in real-world scenarios, such as educational institutions, workplaces, or public spaces, to evaluate its effectiveness and impact.

Multilingual Support: Extend the system to support multiple sign languages, facilitating communication for diverse sign language communities around the world.

Accessibility Features: Introduce accessibility features like text-to-sign translation and vice versa,

References

[1] Z. R. Saeed, Z. B. Zainol, B. B. Zaidan, and A. H. Alamoodi, "A Systematic Review on Systems-Based Sensory Gloves for Sign Language Pattern Recognition: An Update From 2017 to 2022," in IEEE Access, vol. 10, pp. 17, Nov. 4, 2022. DOI: 10.1109/ACCESS.2022.3219430.

[2] A. Sultan, W. Makram, M. Kayed, and A. Amin Ali, "Sign language identification and recognition: A comparative study," Open Computer Science, May 25, 2022. [Online]. Available: https://doi.org/10.1515/comp-2022-0240.

[3] IBM, "Convolutional Neural Networks," IBM.com, [Online]. Available: https://www.ibm.com/topics/convolutional-neural-networks.

Thank You