BMI/CEN 598 (Embedded Machine Learning) Project

Generalized Hand Gesture Recognition

Dheeraj Kallakuri, Kashyap Meher, Raghavendra Dinesh,
Sai Manikanta Badiga, Yashas Puttaswamy

College of Health Solutions
Arizona State University

Instructor: Hassan Ghasemzadeh
Lab: Embedded Machine Intelligence Lab (EMIL)
Web: https://ghasemzadeh.com

Problem Statement

Generalized Hand Gesture Recognition for Multiple Applications

Significance:

1. Video conference and communication systems: control features and express themselves through gestures.
    a. Impacts:
        i. Users: allowing users to convey emotions and intentions non-verbally, improving engagement
        ii. Businesses: more effective remote collaboration, positively impacting productivity
2. ASL Communication:
    a. Impacts:
        i. A Sign Language detector empowers individuals who are deaf or mute by bridging communication gaps, enabling effective interaction with a broader audience.
        ii. This technology enhances education, independence, employment opportunities, and social inclusion while preserving and promoting sign languages and culture.
3. Home Automation: Control smart home devices, more convenient and efficient.
    a. Impacts:
        i. Homeowners:simplifies the interaction with smart home devices, more accessible and user-friendly
        ii. Energy efficient: adjust lighting and temperature, contributing to energy conservation
        iii. Disabled individuals.

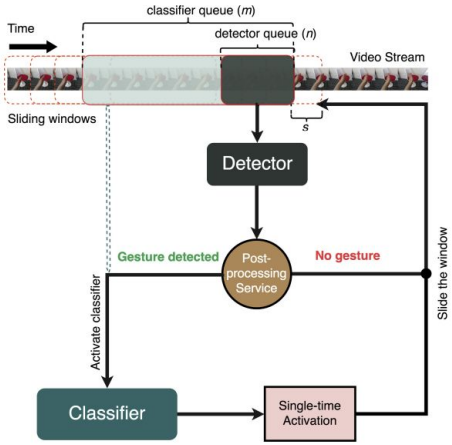
State of the art

Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks - Intel labs, 2019

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8756576

- ResNeXt-101 based architecture
- achieves the state-of-the-art accuracy of 94.04% and 83.82% for depth modality on EgoGesture and NVIDIA Dynamic Hand Gesture Dataset benchmarks, respectively

ResNeXT 101:
~7.8 billion FLOPs, ~44 million parameters.



System Design - A Novel Approach

1. Single frame Classifier

2. Generality Across Applications:

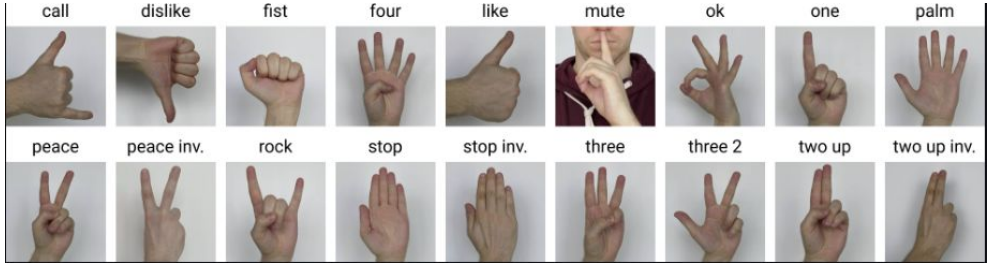
designed to be versatile and applicable to a wide range of use cases. aim to create a framework where each gesture can be customized to specific features based on the intended application.

3. Reducing Model Size:

existing solutions utilize large and resource-intensive models, we prioritize efficiency and scalability. Our approach employs small, lightweight models that are well-suited for resource-constrained environments.

Dataset

Hagrid dataset:



Extensive Data: HaGRID is a massive dataset with 553,991 FullHD RGB images, totaling 723GB in size.

Gesture Classes: It includes 18 distinct classes for recognizing various hand gestures.

No_Gesture Class: The dataset also features a "no_gesture" class for scenarios where no specific gesture is performed, with 108,056 samples.

Varied Conditions: HaGRID incorporates diverse lighting conditions, including artificial and natural light, as well as challenging scenarios like facing and backing to a window.

Diverse Subjects: The dataset covers 37,563 unique individuals aged 18 to 65, with gestures performed at distances ranging from 0.5 to 4 meters from the camera.

At the moment we have tried training with a sample of this dataset from Kaggle's HaGRID Sample 30k 384p 31,833 images.

Dataset

ASL Alphabet:

The data set is a collection of images of alphabets from the American Sign Language, separated in 29 folders which represent the various classes.

The training data set contains 87,000 images which are 200x200 pixels.
There are 29 classes, of which 26 are for the letters A-Z and 3 classes for SPACE, DELETE and NOTHING.

These 3 classes are very helpful in real-time applications, and classification.
The test data set contains a mere 29 images, to encourage the use of real-world test images.

We have used only 26 classes of the alphabets for training and testing the model.

Hardware/Software

Hardware

- Rpi
- Webcam
- Screen
- Keyboard and mouse
- HDMI and power supply

Software

- Rpi OS: Raspbian
- Python library
  - Mediapipe
  - Open CV
  - Scikit Learn
  - PySimpleGUI

Architecture

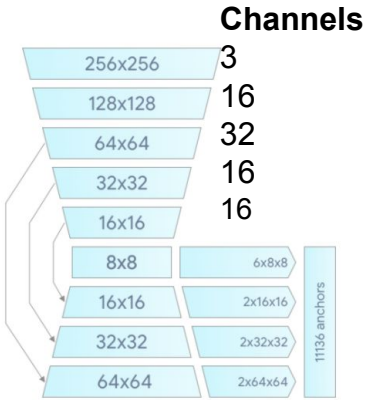
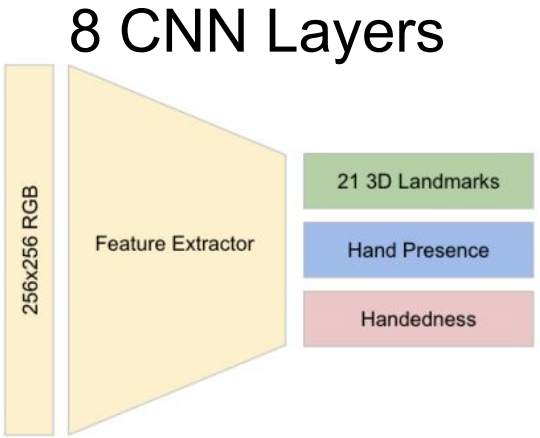
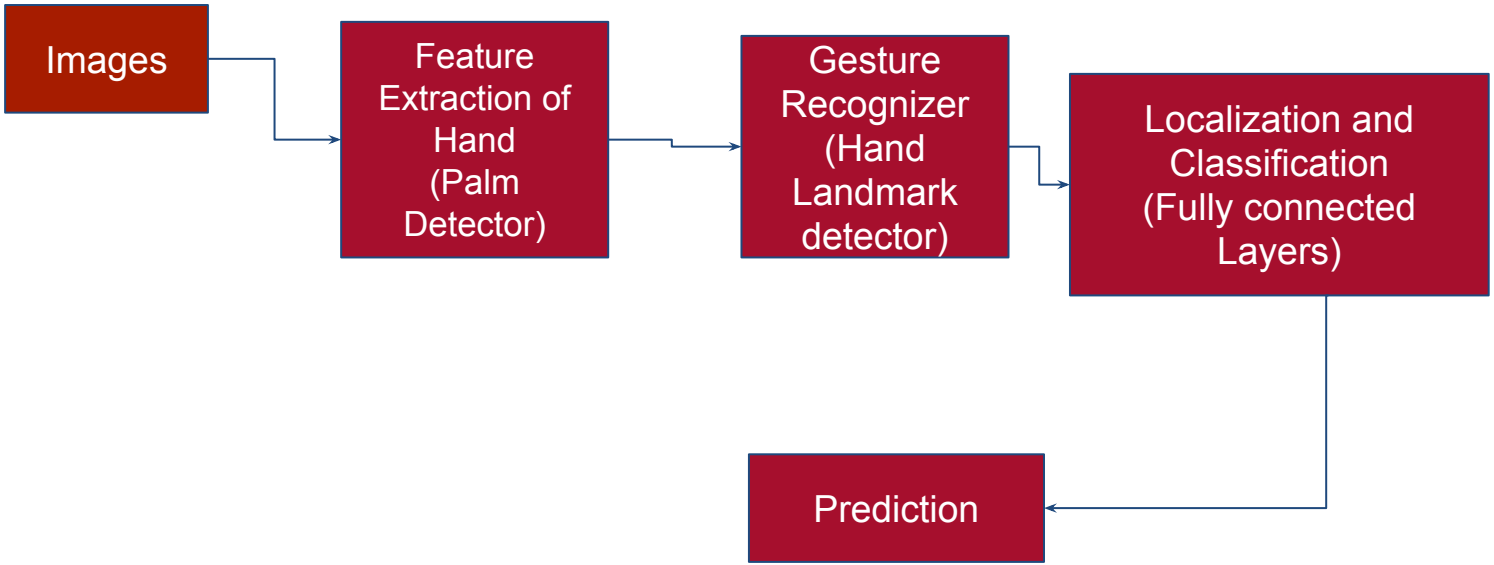


Figure 2: Palm detector model architecture.

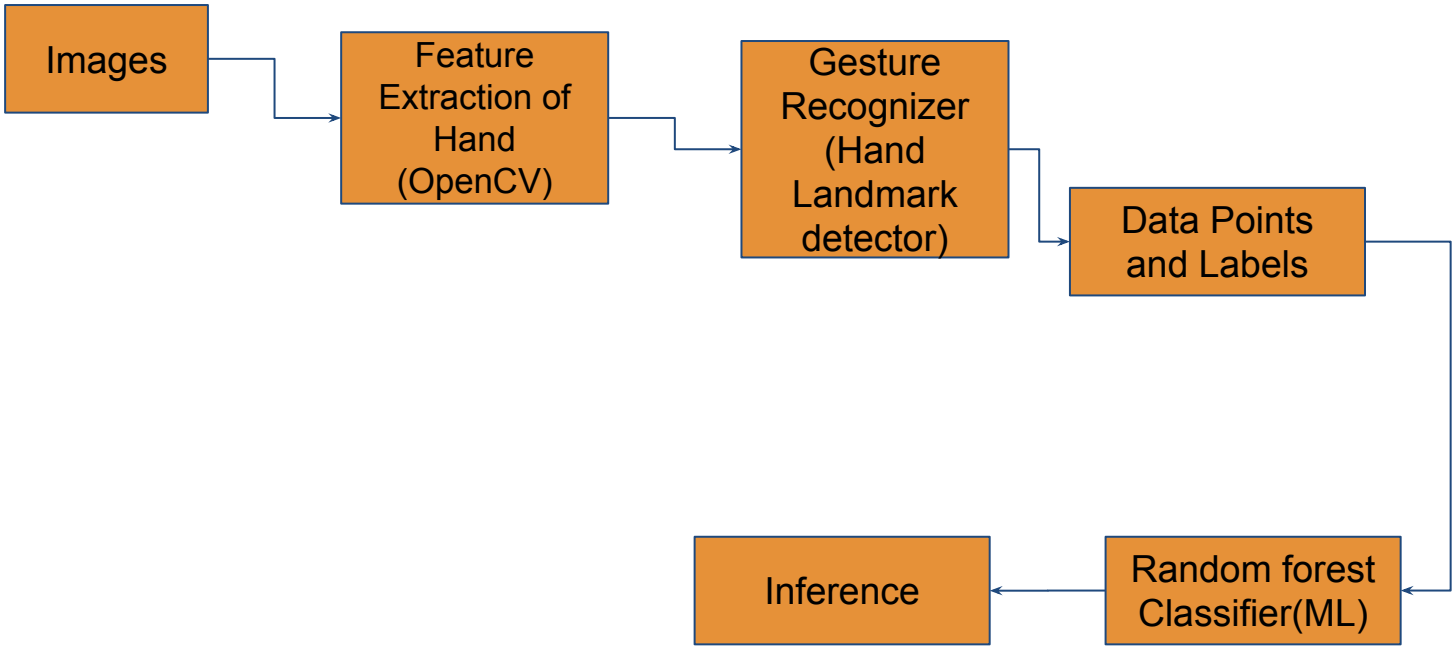


1. 21 hand landmarks consisting of x, y, and relative depth.
2. A hand flag indicating the probability of hand presence in the input image.
3. A binary classification of handedness, *e.g.* left or right hand.

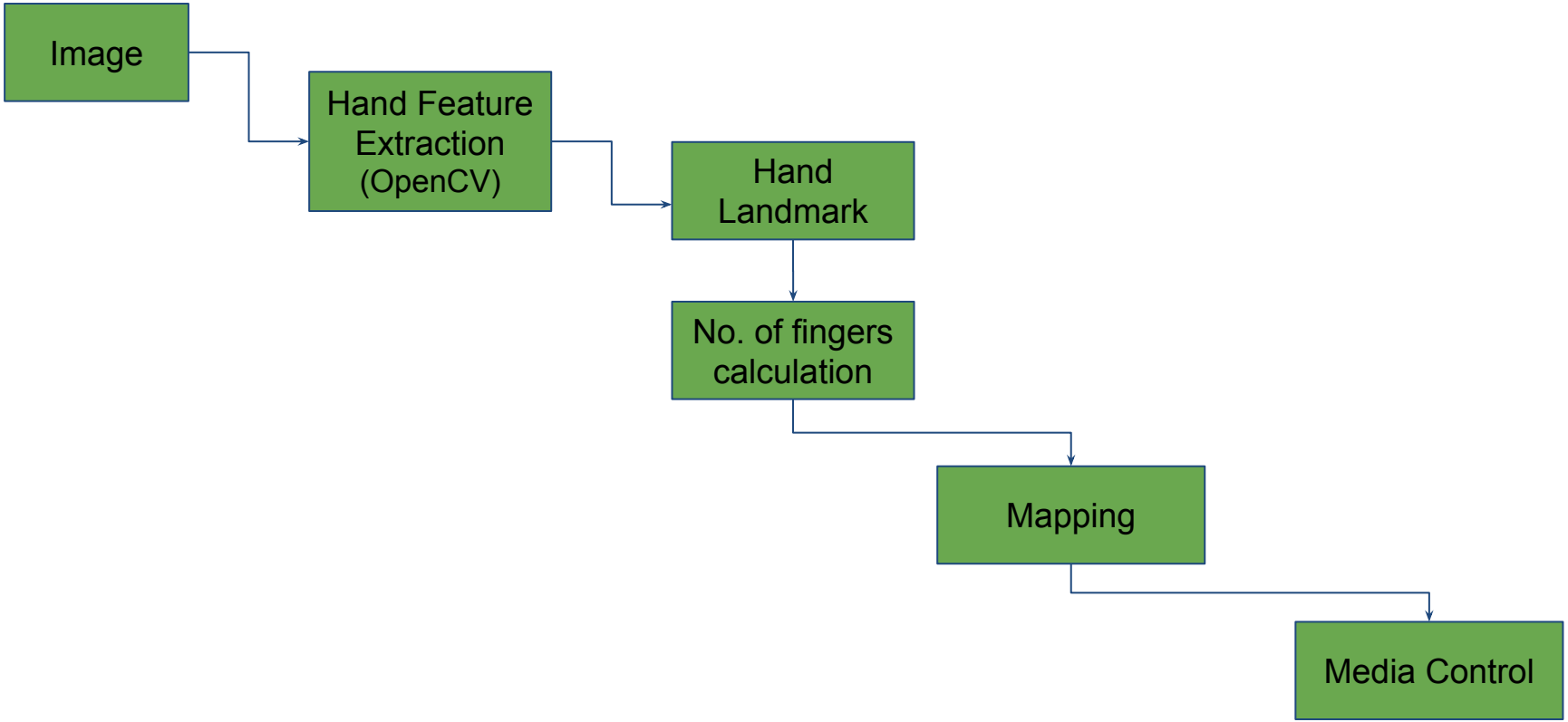
Video Conference Application



Sign Language Application



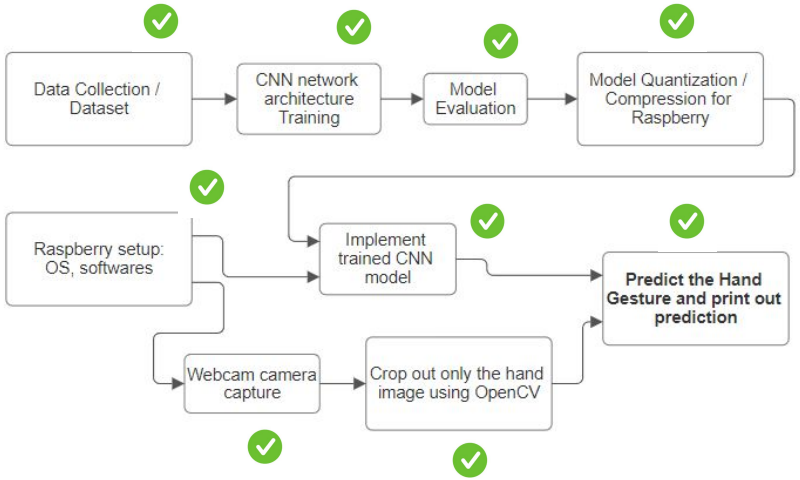
Media Control Application



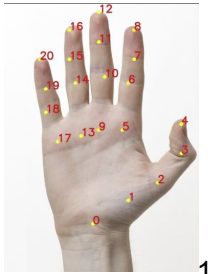
Summary of project

Application	Dataset	Model
Video Conference	Hagrid Dataset	MediaPipe Gesture Recognizer Palm detector + Hand Landmark FPS: 8
Sign Language	ASL Alphabet	OpenCV palm detector + Hand Landmark detection + Random Forest Classifier machine learning model. FPS: 31
Media Control	None	Pretrained Media Pipe Hand Landmark detection + Theoretical number of fingers Calculations FPS: 50

Progress



- **Changes:**
  - We initially developed a CNN network that yielded results. However, we introduced an innovative twist by extracting keypoints using the MediaPipe Gesture Recognizer and assembling a distinct dataset exclusively for these keypoints. This specialized dataset was then employed as input for a Random Forest Classifier machine learning model.
- **Steps:**
  - 1st method: Use Custom model to predict the gesture - too big
  - 2nd method: Extract keypoints using mediapipe, create a new dataset and pass it to a random forest classifier

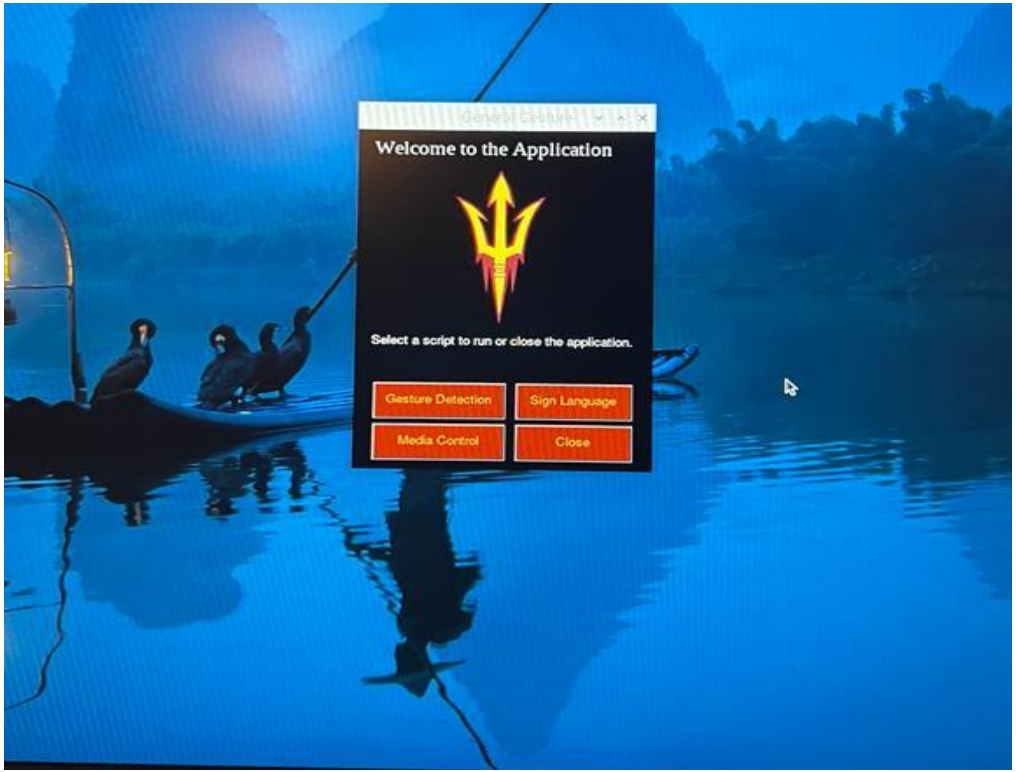


Previous Results:

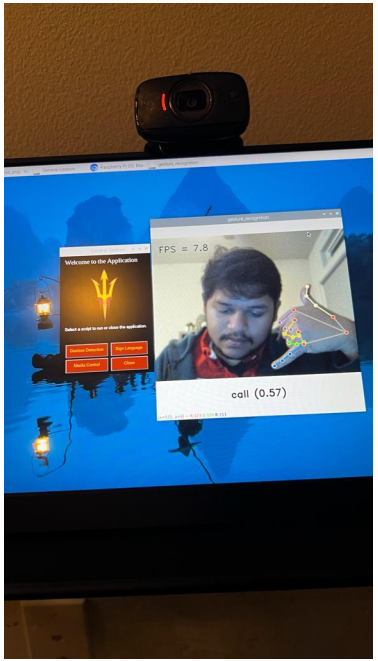
Model 1:

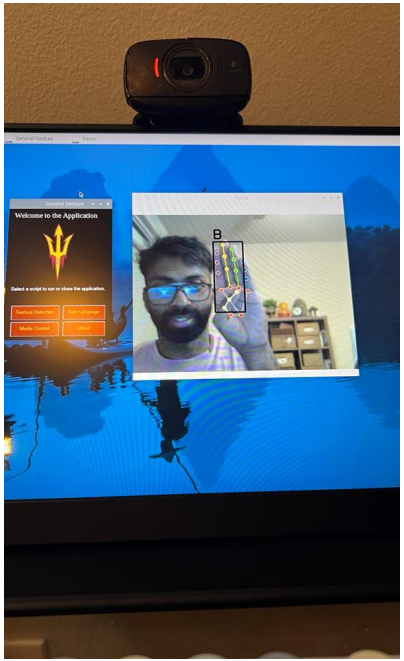
Model 2:

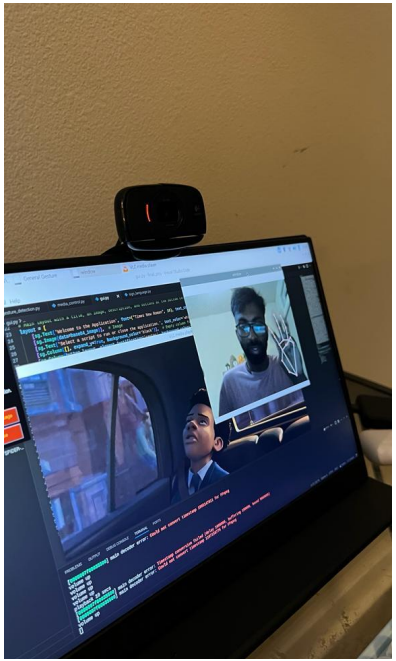
GUI


Welcome to the Application
Select a script to run or close the application.
Gesture Detection
Sign Language
Media Control
Close

Applications GUI







Gesture Control	Sign Language	Media Control

Activities

- Data collection of different gestures - **Done**
- Data preprocessing and labeling - **Done**
- Machine learning Model building - **Done**
- Training, testing and Validation - **Done**
- Model Tuning and Optimisation as per hyper parameters - **Done**
- Implementing on Raspberry pi with User interface - **Done**
- Results related to Accuracy- **Done**
- Implementation - **Done**

Demo Link

https://youtu.be/tW3jCIIxz8g

Future Scope

1. User Friendly GUI
2. Scaling home automation application
3. Making it as extensions in web browsers
4. Addition of multiple application as per use case
5. Custom gestures training using pretrained model
6. Gesture to speech/text module
7. Compact portable device

References

- Real-time Hand Gesture Detection and Classification Using Convolutional Neural Networks - Intel labs, 2019
- HaGRID Sample 30k 384p
- Media Pipe Hands:On-device Real-time Hand Tracking
- https://www.kaggle.com/datasets/grassknoted/asl-alphabet
- https://developers.google.com/mediapipe/

Live Demo


THANK YOU!