**GESTURE RECOGNITION MODEL FOR CONTROLLING VLC PLAYER**

**ABSTRACT:**

Build a Neural Network based solution to recognise the different human gestures, and based on that control a VLC Media player playing a movie.

**Keywords:** PUI, KUI, Gesture, CNN, Transfer Learning, Feature Vector, One Hot Vector, Softmax vector, Categorical Crossentropy.

**INTRODUCTION**

Gesture Recognition tasks are one of the hot topics in research communities today. Gesture recognition is a topic in computer science and language technology with the goal of interpreting human gestures via mathematical algorithms. Gestures can originate from any bodily motion or state but commonly originate from the face or hand.

**WHAT ARE GESTURES**:

Gestures are communicative, meaningful body motions – i.e., physical movements of the fingers, hands, arms, head, face, or body with the objective to convey information or interact with the environment. Waving goodbye, showing full palm for stop, rotating your head to say no, etc are examples of some common gestures. Gestures are like words, i.e. they are understood and exist within the minds of individuals in a society. Just like words they have meaning or concepts attached to them. Nowadays the technology is pursuing this niche to enable computers understand the human body language , i.e. gestures. With the advent of Machine Learning capabilities, this dream seems not so far away.

**THE PUIs :**

Gesture recognition, along with facial recognition, eye tracking and lip movement recognition are components of what developers refer to as a perceptual user interface (PUI), where the input commands to a computer are not direct but are to be guessed by the computer itself.

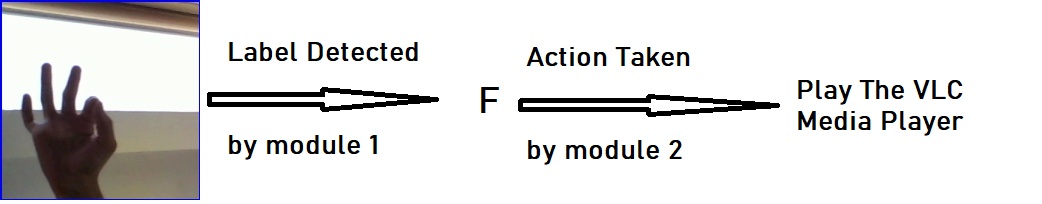
In addition to the technical challenges of implementing gesture recognition, there are also social challenges. Gestures must be simple, intuitive and universally acceptable. The study of gestures and other nonverbal types of communication is known as kinesics

**PROBLEM STATEMENT**

**Module 1**. -> To build and train a Neural network model for inputting a image, and outputting a label for the image corresponding to a particular hand gesture in the image.

**Module 2.** -> To build a python script to control the VLC Media player using the label from 1.

**EXPECTED OUTPUT**

****

# IMPLEMENTATION AND UNIT TESTING METHODOLOGY

# MODULE 1.

**DATA:**

We have used a Kaggle Dataset for ASL (American Sign Language), It contained around 3000 images for each english Alphabet in American Sign Language. This is the link for the dataset- <https://www.kaggle.com/grassknoted/asl-alphabet> .

**MODEL BUILDING:**

I have built and trained the model in Google Colaboratory, an online free Cloud Computing platform from Google, having Support for GPU and TPU and is quite suitable for creating and training small and moderate sized neural networks. The link for the same is as follows, It is advised to open the following link and follow the code along with upcoming explaination.

https://colab.research.google.com/drive/1vVhauCRU9hYFdDWwIM3enZB2aWU48Z6B?usp=sharing

**ARCHITECTURE:**

The model is a simple CNN architecture. Here we have used the VGG16 pretrained neural network as a feature vector generator, and have applied transfer learning to tune the model to our problem of gesture recognition. The model inputs a image of size (64,64,3) and outputs a 7-dimensional vector having probability for each class i.e. softmax vector, namely

F, G, K, L, P, Y, or nothing.

The label having highest probability is assigned to the image. The following table gives you the gesture and its corresponding meaning in ASL.

|  |  |
| --- | --- |
| **IMAGE** | **CORRESPONDING ALPHABET IN ASL** |
|  | **F** |
|  | **G** |
|  | **K** |
|  | **L** |
|  | **P** |
|  | **Y** |
|  | **NOTHING** |

The steps involved in building and training the encoder are as follows:

1. Import Required Libraries. Namely cv2 (OpenCV),keras, tensorflow, numpy, os and sklearn.
2. Prepare the data. We resize the images to size(64,64, 3). And create a numpy variable X containing images of size (no. of images, 64,64,3), and output ,Y , as one hot vector of 7 classes.

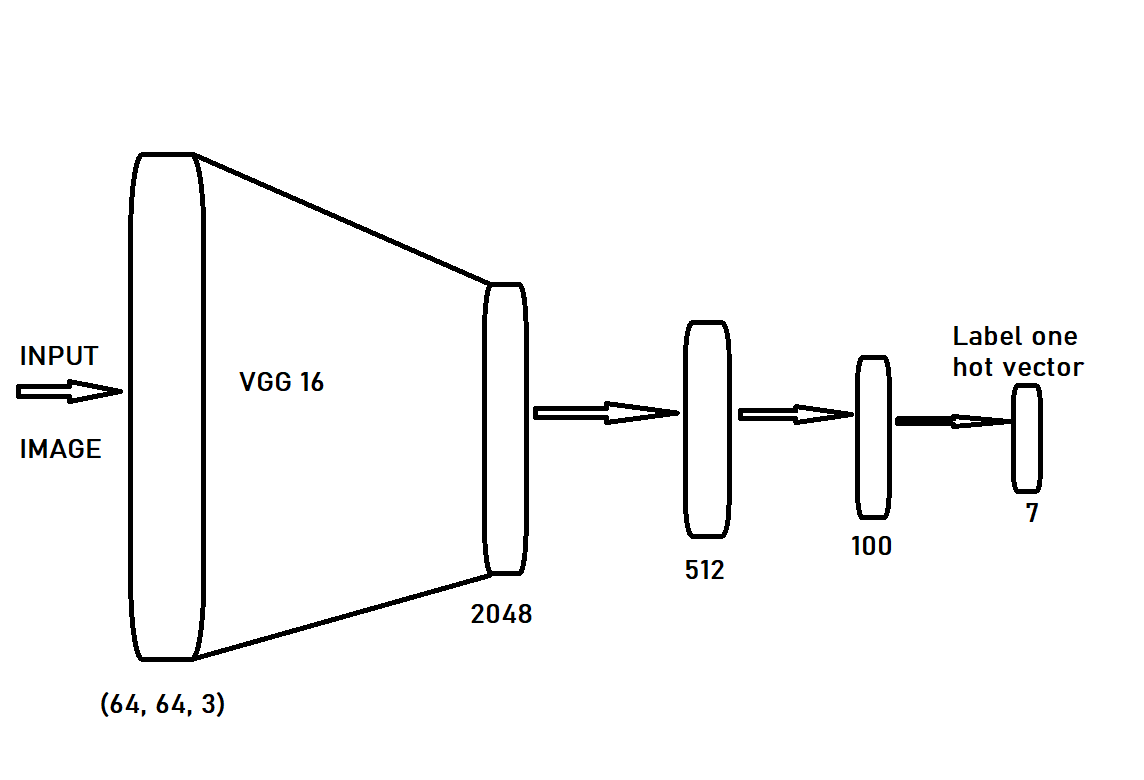
The following dictionary maps the ASL Alphabets to the index in one Hot vector.

dct = {'P' : 0, 'F' : 1, 'G' : 2, 'K': 3, 'L':4, 'Y':5, 'nothing':6}

1. We Split the X, Y in 20-80 ratio for test and train set.
2. We Use the Categorical Cross Entropy loss function for calculating the loss
3. We compile it using Adam optimizer.
4. Then we train the model.

**THE ML MODEL:**

This model takes one Image of size (64,64,3) and outputs a 7-dimensional vector having probability for each class i.e. softmax vector. We import pretrained VGG16 model for input of size(64,64,3), and it outputs a feature vector of size 2048. Then this is followed by Densely Connected layers of size 512->100->7.



**INTERPRETING THE OUTPUT:**

The model outputs a 7-dimensional vector. We pick the index of highest value in the vector. Then we map the index to the label using dct.

dct = {'P' : 0, 'F' : 1, 'G' : 2, 'K': 3, 'L':4, 'Y':5, 'nothing':6}

**TESTING**

We test the model on training set.

We get an accuracy of 99.29 percent accuracy on train set and 99.39 percent accuracy on test set.

The live demonstration of this module can be viewed [here](https://drive.google.com/file/d/1nlYWkXFHXNnxpID0KjXXBHEEaCoUz4GU/view?usp=sharing) .

# MODULE 2.

In this module, the target is to control the vlc media player using python script. For this we use the python-vlc package. The instructions for installing the same are-

sudo apt-get update

sudo apt-get upgrade

pip install python-vlc

Note: This package is for Ubuntu only.

After installing the package, test if it is correctly installed by importing it

import vlc

Some of the commands for using the package are-

player = vlc.MediaPlayer(‘path of file’’)

player.play()

player.pause()

player.audio\_get\_volume()

player.audio\_set\_volume()

player.audio\_mute()

# THE FINAL MODULE BY COMBINING MODULE1. AND MODULE 2.

The final output of our project is integration of module 1. and module 2. allowing us to control the media player from hand gestures. the following table maps the ASL alphabet to the action taken on VC Media player.

|  |  |
| --- | --- |
| **ASL ALPHABET** | **ACTION TAKEN** |
| P | PLAY |
| G | PAUSE |
| F | SEEK FORWARD |
| Y | SEEK BACKWARD |
| K | VOL DOWN |
| L | VOL UP |
| NOTHING | NO ACTION |

A demonstration of our final working module is [here](https://drive.google.com/file/d/1AqCHJ0TsBhxS4MlsvOZNBTebczzDCDOF/view?usp=sharing) .

**CONCLUSION**

The conclusion is that using Simple CNN model and Transfer Learning, we can achieve a high performance Gesture recognition ML model.

**REFERENCES**

Hand gesture recognition Literature review-

<https://pdfs.semanticscholar.org/7785/7eef494b30663e0c17270486f35851246e6d.pdf>

<http://ijarcet.org/wp-content/uploads/IJARCET-VOL-2-ISSUE-1-86-90.pdf>

Working with Google Colab :

<https://towardsdatascience.com/getting-started-with-google-colab-f2fff97f594c#:~:text=You%20can%20create%20notebooks%20in,upload%20Kaggle%20files%2C%20download%20your>