## **Deep Learning Project- Gesture Recognition**

* **Dharmendra Kumar – Group Facilitator**
* **Ashwin Kumar h**

# Problem Statement

In this group project, you are going to build a 3D Conv model that will be able to predict the 5 gestures correctly. Please import the following libraries to get started. Problem Statement

Imagine you are working as a data scientist at a home electronics company which manufactures state of the art smart televisions. You want to develop a cool feature in the smart-TV that can recognise five different gestures performed by the user which will help users control the TV without using a remote.

The gestures are continuously monitored by the webcam mounted on the TV. Each gesture corresponds to a specific command:

| **Gesture** | **Corresponding Action** |
| --- | --- |
| Thumbs Up | Increase the volume. |
| Thumbs Down | Decrease the volume. |
| Left Swipe | 'Jump' backwards 10 seconds. |
| Right Swipe | 'Jump' forward 10 seconds. |
| Stop | Pause the movie. |

The data is in a zip file. The zip file contains a 'train' and a 'val' folder with two CSV files for the two folders. These folders are in turn divided into subfolders where each subfolder represents a video of a particular gesture. Each subfolder, i.e. a video, contains 30 frames (or images). Note that all images in a particular video subfolder have the same dimensions but different videos may have different dimensions. Specifically, videos have two types of dimensions - either 360x360 or 120x160 (depending on the webcam used to record the videos). Hence, you will need to do some pre-processing to standardise the videos.

# Understanding the Dataset

**Each row of the CSV file represents one video and contains three main pieces of information - thename of the subfolder containing the 30 images of the video, the name of the gesture and the**

Your task is to train a model on the 'train' folder which performs well on the 'val' folder as well (as usually done in ML projects). We have withheld the test folder for evaluation purposes - your final model's performance will be tested on the 'test' set. **Numeric label (between 0-4) of the video.**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Exp** | **Model** | **Hyper- Parameters** | **Result** | **Decision + Explanation** | **Parameters** | |
| **1** | **Conv3D** | **n\_frames=16**  **Epoch = 20**  **Batch size =64** | **Train Accuracy: .90**  **Val Accuracy: 0.62**  **val\_loss improved from 1.73372 to 1.07627.**  **Reducing learning rate to 0.00020000000949949026.** | **As we seen Train Accuracy is .90 and Val accuracy 0.62 in first model, to improve further we increase the number of frames and Epoch to improve Accuracy.** | | **2,067,621** |
| **2** | **Conv3D** | **n\_frames = 20**  **Epoch = 25**  **batch size = 64** | **Train Accuracy: .82**  **Val Accuracy: 0.62**  **val\_loss did not improve from 1.07627** | **After increasing hyper parameter Train accuracy is not improving and also val\_loss is not improving.** | | **4,050,085** |
| **3** | **Conv3D** | **n\_frames = 25**  **Epoch = 20**  **batch size = 64** | **Train Accuracy: .82**  **Val Accuracy: 0.75**  **val\_loss did not improve from 0.79441**  **Reducing learning rate to 8.000000525498762e-06.** | **Further we increase the Hyper Parameters n\_frames=25 and reduce the Epoch we see Accuracy is not improving but Val Accuracy is improving. Total Parameters are also increased.** | | **7,023,781** |
| **4** | **Conv3D** **(Same pooling in Conv3D layer** **)** | **n\_frames = 20**  **Epoch = 20**  **batch size = 64** | **Train Accuracy: .81**  **Val Accuracy: 0.50**  **val\_loss did not improve from 0.79441** | **After Decresing n\_frames we can see Train Accuracy and Val Accuracy is decreasing. And also Val\_loss is not improving.** | | **1,907,909** |
| **5** | **Conv3D** **(Reduced kernel to (2,2,2), Batch Normalization before MaxPooling** **)** | **n\_frames = 20**  **Epoch = 20**  **batch size = 64** | **Train Accuracy: .97**  **Val Accuracy: 0.50**  **val\_loss did not improve from 0.79441**  **reducing learning rate to 4.0000001899898055e-05** | **In MaxPooling model we can see Train Accuracy is Up to .97 from previous model. And Val\_loss is not changing. Also learning rate is reducing, and number of Parameter is also decreasing.** | | **1,301,045** |
| **6** | **Conv2D+LSTM** **(Switching BatchNormalization before MaxPooling** **)** | **n\_frames = 20**  **Epoch = 30**  **batch size = 32** | **Train Accuracy: 89**  **Val Accuracy: 0.31**  **val\_loss did not improve from 0.79441** | **In Conv2D+LSTM**  **Accuracy is not improving and also number of parameter is increasing. So we implement other model.** | | **3,084,133** |
| **7** | **Transfer Learning (MobileNet) with LSTM** | **n\_frames = 20**  **Epoch = 30**  **batch size = 64** | **Train Accuracy:1.0**  **Val Accuracy: 1.0**  **val\_loss improved from 0.35556 to 0.24042** | **In LSTM Model we can see Train Accuracy is: 1.0 and Val Accuracy is 1.0 which is best among all the model. So we consider to this Model. Also val\_loss is improving from 0.35556 to 0.24042 in this model.** | | **4,611,781** |