**Problem Statement:**

Imagine we are working as a data scientist at a home electronics company which manufactures state of the art smart televisions. We want to develop a cool feature in the smart-TV that can recognize 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:

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

Each video is a sequence of 30 frames (or images)

**Approach:**

In the generator function, the images have been cropped and resized as the images given are of 2 different dimensions. Also, the batches have been normalized.

Used Conv3D first which gave us an accuracy of 0.79. The training parameters were above 9 million and it took a long time to finish. Hence, tried another way of achieving this i.e., Conv3D + RNN. For RNN, we chose GRU, as GRU gives a good accuracy with less training parameters. It uses lesser memory and is faster than LSTM.

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Result** | **Decision + Explanation** |
| **1** | **Conv3D** | **Out of memory error** | **Batch size:64; need to reduce batch size so as to use maximum GPU usage** |
| **2** | **Conv3D** | **Accuracy: 0.79** | **9 Million+ trainable parameters; too much training time** |
| **3** | **Conv + RNN** | **Accuracy: 0.28** | **Batchsize:64; Epochs:70; Accuracy too less** |
| **4** | **Conv + RNN** | **Accuracy: 0.80** | **Batch size:16; Epochs:70** |
| **Final Model** | **Conv + RNN** | **Accuracy: 0.80** | **Batch size:16; Epochs:70** |