## **Gesture Recognition:** Deep Learning Assignment

* **Jayshree Singh**
* **Manisha Modak**

# Problem Statement

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 recognise five different gestures performed by the user which will help users control the TV without using a remote.

🡪 Each gesture corresponds to a specific command:

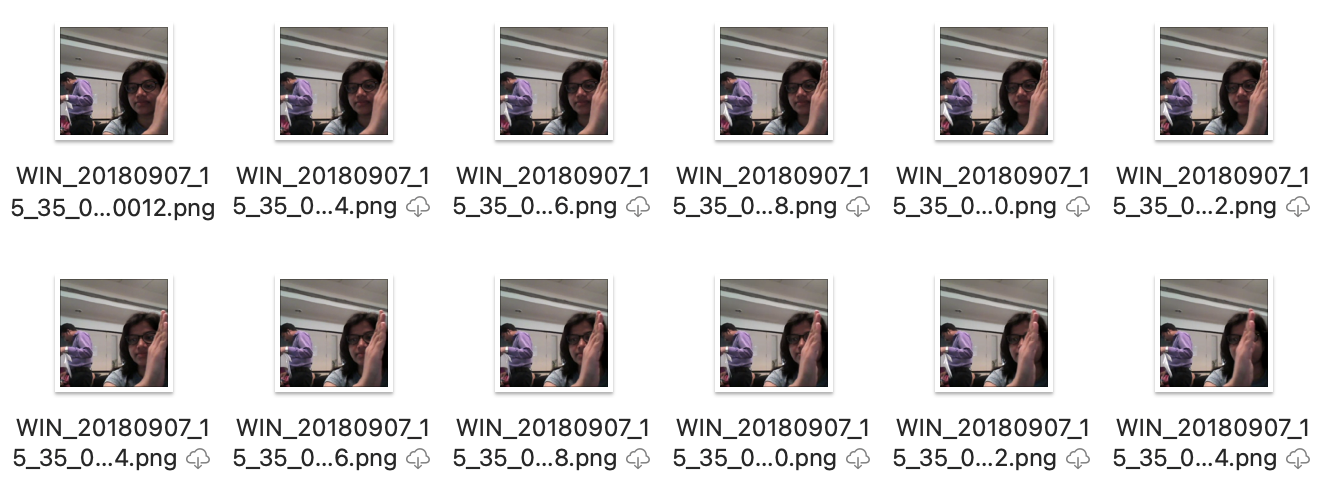
* Thumbs up:  Increase the volume
* Thumbs down: Decrease the volume
* Left swipe: 'Jump' backwards 10 seconds
* Right swipe: 'Jump' forward 10 seconds
* Stop: Pause the movie

**ß**

**DATA SET :** <https://drive.google.com/uc?id=1ehyrYBQ5rbQQe6yL4XbLWe3FMvuVUGiL>

# Understanding the Dataset

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 folder again has 663 folders, each folder represent a video(contains 30 frames/images).

****

# Objective

Our task is to train different models on the 'train' folder to predict the action performed in each sequence or video and which performs well on the 'val' folder as well.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Combination of epoch,batch size and layers** | **Result** | **Decision + Explanation** |
| **Initial** | **Conv3D** | epoch=20  num\_of\_images=15,  MaxPooling3D | **Throws Generator error** | **Crop the images correctly, try to overfit on less amount of data** |
| **1** | **Conv3D** | epoch=1,  num\_of\_images=30,  kernel\_size=3,  MaxPooling3D(k=2) | **Accuracy: 0.18** | **Simple Conv3D model** |
| **2** | **Conv3D** | epoch=1  num\_of\_images=30,  kernel\_size=5,  MaxPooling3D(k=2) | **Accuracy: 0.28** | **3DConv Model with kernel size 5, Improvement in accuracy** |
| **3** | **Conv3D** | epoch=1  num\_of\_images=30,  kernel\_size=3,  MaxPooling3D(k=2) | **Accuracy: 0.10** | **3DConv Model with kernel size 3, accuracy dropped** |
| **4** | **Conv3D** | epoch=1  num\_of\_images=30,  kernel\_size=5,  MaxPooling3D | **Accuracy: 0.26** | **3DConv Model with kernel size 5 and different max pooling kernel size, accuracy increased.** |
| **5** | **Conv3D** | epoch=1  num\_of\_images=30,  kernel\_size=5,  MaxPooling3D | **Accuracy: 0.18** | **3DConv Model with kernel size 5 and increased layer**  **-> accuracy dropped to 0.18 again** |
| **6** | **Conv3D** | epoch=1  num\_of\_images=30,  kernel\_size=5,  MaxPooling3D | **Accuracy: 0.26** | **3DConv Model with kernel size 5 and different dropouts ->accuracy restored to 0.26 again** |
| **7** | **Conv2D +GRU** | epoch=2 num\_of\_images:15,  GlobalAveragePooling3D | **Accuracy:** 0.6062 | **loss: 1.0006 - categorical\_accuracy: 0.6062**  **Better than previous.** |
| **8** | **Conv2D +GRU** | epoch=2 num\_of\_images:15,  GlobalAveragePooling2D | **Accuracy**  **.5241**  **Generator finished before completion** | **loss: 1.0827 - categorical\_accuracy: 0.5241** |
| **9** | **Conv2D+LSTM** | epoch=2 num\_of\_images:30, | **Accuracy .46. and with some variation acc is .43** | **Model accuracy is dropped, model is overfitting. could be because of less number of epochs.** |
| **Final Model** | **Conv2D +GRU** | epoch=2 num\_of\_images:15,  GlobalAveragePooling3D | **Accuracy: 0.6062** | **This is the best accuracy we could achieve with the limitation of number of epochs. loss: 1.0006 - categorical\_accuracy: 0.6062.** |

**Limitation:** We had faced the limitation with number of Epochs, could not run more than 2 epochs. It is failing with Error: Failed to format this callback filepath: "<ModelName>/model2-{epoch:05d}-{loss:.5f}-{categorical\_accuracy:.5f}-{val\_loss:.5f}-{val\_categorical\_accuracy:.5f}.h5". Reason: \'val\_loss\''

**Other Experiments**

1. **Conv2D +GRU**

Ran with epoch=20 num\_of\_images:15, GlobalAveragePooling3D

Error: Throws Generator error. With high number of high epoch

1. **Connv2D+LSTM: with 2 Epoch, batch\_size=20 anf image size=15. (even numbers only)**

