Deep Learning Course Project- Gesture Recognition

Created By: Piyush Sukhija

# Problem Statement

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

• 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.

# Experiment Logs

**Image size fixed:**160\*160

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | EXPERIMENT | RESULT | DECISION + EXPLANATION | PARAMETERS |
| Conv3D | 1 | Incompatibility with imageio | Redesigning the data generator | - |
| Conv3D | 2 | Training Accuracy : 0.9600  Validation Accuracy: 0.8100  *Achieved at 12 epochs* | Base model without dropouts and normalization  (epoch = 20, batch size=5) | 1,734,661 |
| Conv3D | 3 | OOM Error | Increasing epoch to 30 and batch size to 30 and adding batch normalization with changes to filters  (epoch = 20, batch size=30) | 2,420,837 |
| Conv3D | 4 | Training Accuracy : 0.9900  Validation Accuracy : 0.8800  *Achieved at 20 epoch* | Reducing batch size =5 and increasing epoch to 30 | 2,420,837 |
| Conv3D | 5 | Training Accuracy : 0.9975  Validation Accuracy : 0.8700  *Achieved at 35 epoch* | *Adding dropouts and epochs is increased to 40* | 2,420,933 |
| Conv3D | 6 | Training Accuracy : 0.9900  Validation Accuracy : 0.8600  *Achieved at 21 epoch* | Dropout reduced a lot of accuracy, reducing dropout limit (0.15) and using asymmetrical filters with reduced batch size (5) and increased epoch to 50 | 1,940,133 |
| Conv3D | 7 | Training Accuracy : 0.9950  Validation Accuracy : 0.9400  *Achieved at 32 epoch* | Reverting to symmetric filters with more dense layers with GlobalAveragePooling3D in place of flatten | 1,704,069 |
| Conv3D | 8 | Training Accuracy : 0.99  Validation Accuracy : 0.98 | Adding more layers and removing maxpooling to stack convolution layers for better learning and reducing batch size to 5 | 1,397,253 |
| Conv2D-GRU | 9 | Training Accuracy : 0.80  Validation Accuracy : 0.83 | Moving to different structure as further addition only provide diminishing returns, with expectation of lower parameters | 99,269 |
| ConvLSTM2D- Dense | 10 | Training Accuracy : 0.84  Validation Accuracy : 0.90 | Moving to a different layer unit of ConvLSTM2D | 13,589 |
| ConvLSTM2D- GRU-Dense | 11 | Training Accuracy : 0.90  Validation Accuracy : 0.89 | Adding more layers with dropouts for better learning capacity | 434,501 |
| Transfer learning-RNN | 12 | Training Accuracy : 0.89  Validation Accuracy : 0.67 | EfficientNetV2S used as feature extractor | 177,905 |