**GESTURE\_RECOGNITION PROJECT WRITE UP**

The project is based on developing a model for a smart-TV that can accurately identify the five gestures of the user to control the TV settings.

The gestures given below refer to these corresponding commands:

* 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 consists of a sequence of 30 frames.

**Given below are the models attempted:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| Model\_Name | No of Parameters | Augment Data | Highest  Validation Accuracy | Corres-ponding  Training Accuracy | Observation | Some important info |
| Con\_3d1 | **3,509,189** | no | 78% | 98.7% | Clear case of overfitting | Img res: 120  Epoch : 20  Frames : 16  Batch : 10  Dropout L: None  Dense N : 512  Filter Size: 3 |
| Con\_3d2 | 1,900,997 | no | 82% | 95% | Overfitting has reduced to some extent with the addition of drop out rate. | Img res: 120  Epoch : 20  Frames : 16  Batch : 10  Dropout L: 0.5  Dense N : 256  Filter Size: 3 |
| conv\_3d3 | **1,695,797** | yes | 82% | 88.6% | Reduced the filter size to see if the model improves.  The training accuracy has fallen. However, overfitting has reduced even more. This shows that filter size 2 is better for the model. | Img res: 120  Epoch : 20  Frames : 16  Batch : 10  Dropout L: 0.5  Dense N : 256  Filter Size: 2 |
| conv\_3d4 | 3,303,989 | yes | 91% | 95% | Increased the Dense Neurons to 512 to make the training model learn better .  This model has solved the overfitting problem with very high Validation and training accuracy | Img res: 120  Epoch : 20  Frames : 16  Batch : 10  Dropout L: 0.5  Dense N : 512  Filter Size: 2 |
| conv\_3d5 | 9,517,461 | yes | 84% | 94% | Reduced a layer to check if the layer is redundant:  Trying to reduce the img res, batch size and frames did not help much.  Overfitting can be seen again. Also, Model needs the con3d 32 layer. | Img res: 100  Epoch : 20  Frames : 10  Batch : 8  Dropout L: 0.5  Dense N : 512  Filter Size: 2  Removed Con3d 32 layer |
| conv\_3d6 | 7,399,349 | yes | 80 % | 91% | Reduced a different layer to check if the layer is redundant: Without the Con3d 128 layer, the model did worse. Clearly this layer is much needed. | Img res: 120  Epoch : 20  Frames : 10  Batch : 8  Dropout L: 0.5  Dense N : 512  Filter Size: 2  Removed: Con 128 layer |
| conv\_3d7 | 14,798,581 | yes | 81% | 94% | Reduced a different layer to check if the layer is redundant: Without the Con3d 64 layer the performance is not that great. So. this shows that this layer is much needed.  Conclusion: All the layers are crucial parts of the model and are utilized well. | Img res: 120  Epoch : 20  Frames : 10  Batch : 8  Dropout L: 0.5  Dense N : 512  Filter Size: 2  Removed: Con 64 layer |
| rnn\_cnn1  CNN\_LSTM | 3,338,869 | yes | 86% | 80% | Testing CNN + LSTM model  This model is an improvement from the earlier one.  Here the validation accuracy is higher than the training one. | Img res: 120  Epoch : 20  Frames : 16  Batch : 10  Dropout L: 0.5  Dense N : 128  LSTM cell: 128  Filter Size: 2 |
| conv\_3d9  With data augmentation | 1,096,901 | yes | 75% | 87% | The model performs poorly with data augmentation. Data. Overfitting is strongly present.  Conclusion: Data Augmentation causes our model to overfit. | Img res: 120  Epoch : 20  Frames : 16  Batch : 8  Dropout L: 0.5  Dense N : 128  LSTM cell: 128  Filter Size: 3 |
| rnn\_cnn\_tl  Transfer Learning | 3,840,453 | yes | 84% | 85% | Trying out mobilenet transfer (LSTM) model  This model is not that great as it starts overfitting from 8th epoch onwards.  The validation and training accuracy could be better. | Img res: 120  Epoch : 20  Frames : 16  Batch : 8  Dropout L: 0.5  Dense N : 128  LSTM cell: 128 |
| rnn\_cnn\_tl2  Transfer Learning with GRU | 3,693,253 | yes | 99% | 99% | Trying out mobilenet transfer (GRU) model  This model is the best in terms of validation and training accuracy. | Img res: 120  Epoch : 20  Frames : 16  Batch : 10  Dropout L: 0.5  Dense N : 128  GRU\_cell: 128 |

Conclusion: The best model is **conv\_3d4.**

\* Training Accuracy : 95%, Validation Accuracy : 91%

\* Number of Parameters: 3,303,989 and is much less than model 11 which has the best performance but the model 11 has 389,264 parameters more than model 4.