**Neural Networks Project - Gesture Recognition**

# Problem Statement:

A home electronics company which manufactures state of the art smart televisions, 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 following table consists of the experiments done to build a model to predict the gestures from the given data set.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Exp.No** | **Architecture** | **Hyper Parameter** | **Total number of Parameters** | **Resullts** | **decision + Explanation** |
| 1 | Cov3d | image\_height = 160 image\_width = 160 channels=3 num\_classes=5 total\_frames=30 frames\_to\_sample=15 batch\_size=20 num\_epochs=2 | Total params: 3,308,613 Trainable params: 3,308,005 Non-trainable params: 608 | loss: 1.7057 categorical\_accuracy: 0.4118 val\_loss: 4.5219 val\_categorical\_accuracy: 0.2100 | This is an ablation experiment to test the working of the model for 2 epochs |
| 2 | Cov3d | image\_height = 160 image\_width = 160 channels=3 num\_classes=5 total\_frames=30 frames\_to\_sample=20 batch\_size=40 num\_epochs=15 | Total params: 1,117,061 Trainable params: 1,116,325 Non-trainable params: 736 | loss: 1.6054 categorical\_accuracy: 0.4329 val\_loss: 1.5445 val\_categorical\_accuracy: 0.3200 | Increased the number of epochs to 40 and the validation lows improved compared to previous model. But got resource exhaused error. |
| 3 | Cov3d | image\_height = 160 image\_width = 160 channels=3 num\_classes=5 total\_frames=30 frames\_to\_sample=20 batch\_size=20 num\_epochs=15 | Total params: 1,117,061 Trainable params: 1,116,325 Non-trainable params: 736 | loss: 0.3284 categorical\_accuracy: 0.8914 val\_loss: 1.0289 val\_categorical\_accuracy: 0.6600 | Enabled data augmentation with batch size of 20 and the validation accuracy is increased from 0.32 to 0.66 from previous model |
| 4 | Cov3d | image\_height = 120 image\_width = 120 channels=3 num\_classes=5 total\_frames=30 frames\_to\_sample=16 batch\_size=20 num\_epochs=20 | Trainable params: 698,533 Non-trainable params: 736 | loss: 1.6513 categorical\_accuracy: 0.3906 val\_loss: 1.7845 val\_categorical\_accuracy: 0.2100 | Model did not improve on validation accuracy and earlystopping occurred at 11th epoch |
| 5 | Cov3d | image\_height = 160 image\_width = 160 channels=3 num\_classes=5 total\_frames=30 frames\_to\_sample=20 batch\_size=20 num\_epochs=20 | Total params: 3,638,981 Trainable params: 3,637,477 Non-trainable params: 1,504 | loss: 1.8765 categorical\_accuracy: 0.4005 val\_loss: 2.7586 val\_categorical\_accuracy: 0.2100 | Model did not improve on validation accuracy and earlystopping occurred at 11th epoch |
| 6 | Cov3d | image\_height = 120 image\_width = 120 channels=3 num\_classes=5 total\_frames=30 frames\_to\_sample=16 batch\_size=30 num\_epochs=25 | Total params: 1,762,613 Trainable params: 1,761,109 Non-trainable params: 1,504 | loss: 2.1730 categorical\_accuracy: 0.3243 val\_loss: 1.7409 val\_categorical\_accuracy: 0.2100 | Model did not improve on validation accuracy and earlystopping occurred at 11th epoch |
| 7 | Conv2d+LSTM | image\_height = 120 image\_width = 120 channels=3 num\_classes=5 total\_frames=30 frames\_to\_sample=15 batch\_size=20 num\_epochs=20 lstm\_cells=64 dense\_neurons=64 dropout=0.25 | Total params: 3,392,869 Trainable params: 3,392,389 Non-trainable params: 480 | loss: 0.2386 categorical\_accuracy: 0.9465 val\_loss: 0.6688 val\_categorical\_accuracy: 0.6900 | Best Validation loss and accuracy so far |
| 8 | Conv2d+GRU | image\_height = 120 image\_width = 120 channels=3 num\_classes=5 total\_frames=30 frames\_to\_sample=15 batch\_size=20 num\_epochs=20 gru\_cells=128 dense\_neurons=128 dropout=0.25 | Total params: 2,573,925 Trainable params: 2,573,445 Non-trainable params: 480 | loss: 1.4059 categorical\_accuracy: 0.4133 val\_loss: 1.7165 val\_categorical\_accuracy: 0.1800 | GRU model did not give much accuracy compared to LSTM model |
| 9 | Mobile\_net+LSTM | image\_height = 120 image\_width = 120 channels=3 num\_classes=5 total\_frames=30 frames\_to\_sample=15 batch\_size=20 num\_epochs=20 lstm\_cells=128 dense\_neurons=128 dropout=0.25 | Total params: 3,840,453 Trainable params: 609,541 Non-trainable params: 3,230,912 | loss: 0.2892 categorical\_accuracy: 0.9035 val\_loss: 0.5169 val\_categorical\_accuracy: 0.7500 | Applied transfer learning - used mobile net without training weights and got best validation accuracy so far |
| 10 | Mobile\_net+GRU | image\_height = 120 image\_width = 120 channels=3 num\_classes=5 total\_frames=30 frames\_to\_sample=15 batch\_size=10 num\_epochs=20 gru\_cells=128 dense\_neurons=128 dropout=0.25 | Total params: 3,693,253 Trainable params: 3,669,317 Non-trainable params: 23,936 | loss: 0.0164 categorical\_accuracy: 0.9970 val\_loss: 0.0263 val\_categorical\_accuracy: 1.0000 | Applied transfer learning - used mobile net with training weights and got best validation accuracy so far |