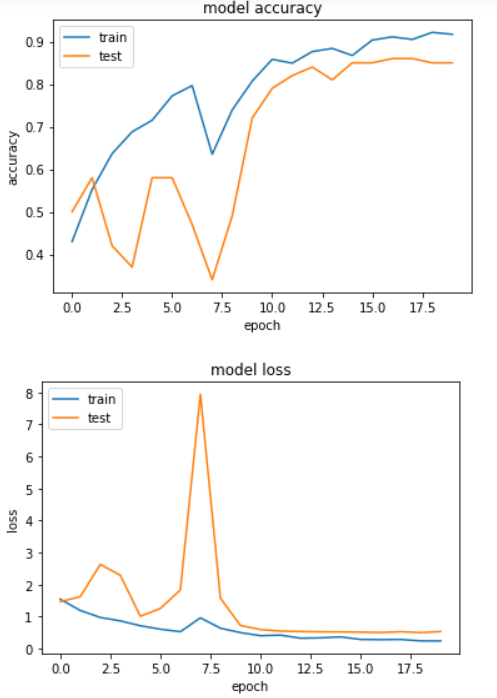
1. Dropouts (usually it was set to 0.25 in dense connections) and batch normalization play an important role for regularization but not explicitly mentioned them in the model column.
2. We zeroed down to batch size of 10 for all the models as it was not giving any memory issues and models were training well with it.
3. We zeroed down to experimenting with 20 images per video.
4. We zeroed down to experimenting with either 120 by 120 image size OR 160 by 160 image size.
5. Final models are highlighted in yellow along with their accuracy and loss numbers.
6. Data transformations/augmentation didn’t seem to be helping much in improving accuracy or loss so we started avoiding them in the final constructive modeling.

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Result** | **Decision + Explanation** |
| **1** | **Conv 3D Model using 30 frames per video**  **+ 16, 32, 64, 128 filters conv 3D layers each followed by max pool 3D**  **+ 256 dense nodes**  **+ 128 dense nodes**  **+ image size 120 by 120** | **Training Accuracy: 90%**  **Validation Accuracy: 80%** | **Low validation accuracy as compared to training accuracy.** |
| **2** | **Conv 3D Model using 20 frames per video**  **+ 16, 32, 64, 128 filters conv 3D layers**  **+ 256 dense nodes**  **+ 128 dense nodes**  **+ image size 120 by 120** | **Training loss: 0.2800**  **Training Accuracy: 91%**  **Validation loss: 0.5032**  **Validation Accuracy: 86%** | **Low loss and high accuracy and comparable between training and validation data sets.**  **Parameters - ~1.9 million.** |
| **3** | **Conv 3D Model with 30 frames per video**  **+ 16, 32, 64, 128 filters conv 3D layers**  **+ 256 dense nodes**  **+ 128 dense nodes**  **+ image size 160 by 160** | **Training Accuracy: 87%**  **Validation Accuracy: 82%** | **Parameters are on higher side and even the model accuracy is not good enough.**  **Parameters-~3 million** |
| **4** | **Conv 3D Model with 20 frames per video**  **+ 16, 32, 64, 128 filters conv 3D layers**  **+ 256 dense nodes**  **+ 128 dense nodes**  **+ image size 160 by 160** | **Training Accuracy: 88%**  **Validation Accuracy: 79%** | **120 by 120 image size is better for training than 160 by 160 image size.** |
| **5** | **Conv 3D Model with 30 frames per video**  **+ 16, 32, 64, 128 filters conv 3D layers**  **+ 256 dense nodes**  **+ 128 dense nodes**  **+ Random data transformations on training data set** | **Training Accuracy: 75%**  **Validation Accuracy:59%** | **Accuracy dropped with addition of random data transformation significantly.** |
| **6** | **Conv 3D Model with 30 frames per video**  **+ 8, 16, 32, 64 filters conv 3D layers**  **+ 256 dense nodes**  **+ 128 dense nodes**  **+ image size 120 by 120** | **Training Accuracy: 89%**  **Validation Accuracy: 81%** | **Parameters are on lower side ~900K**  **But slightly low validation accuracy.** |
| **7** | **Conv 3D Model with 30 frames per video**  **+ 16, 32, 64, 128 filters conv 3D layers**  **+ 128 dense nodes**  **+ 64 dense nodes**  **+ image size 120 by 120** | **Training Accuracy: 92%**  **Validation Accuracy:84%** | **Though the no of parameters were less approx. 1 million the accuracy was not to the mark.** |
| **8** | **Conv 3D Model with 20 frames per video**  **+ 32, 64, 128, 256 filters conv 3D layers**  **+ 128 dense nodes**  **+ 128 dense nodes**  **+ image size 120 by 120** | **Training Accuracy: 88%**  **Validation Accuracy: 80%** | **Increasing number of filters increased the number of parameters to ~2.7 million and validation accuracy is lower only.** |
| **9** | **Conv 3D Model with 30 frames per video**  **+ 5 by 5 filter size (16 conv 3d filters layer)**  **+ 3 by 3 filter size (32, 64, 128) conv 3D layers**  **+ 256 dense nodes**  **+ 128 dense nodes**  **+ image size 120 by 120** | **Training Accuracy: 78%**  **Validation Accuracy: 78%** | **Model seems to be fitting pretty well but slightly low accuracy.**  **We trained for 15 epochs due to time constraint.**  **We will train it for 25 epochs and check again.** |
| **10** | **Mobilenet (retrain all weights)**  **+ GRU (128 cells)**  **+ Dense (128 nodes)**  **+ image size 120 by 120**  **+ 20 images per video** | **Training Accuracy: 99%**  **Validation Accuracy: 95%** | **Retrained all the weights of Mobilenet.**  **Batch size = 10**  **Epochs = 20** |
| **11** | **Mobilenet (retrain all weights)**  **+ GRU (128 cells)**  **+ Dense (128 nodes)**  **+ image size 120 by 120**  **+ 30 images per video**  **+ random data transformations on the images** | **Training Accuracy: 98.8**  **Validation Accuracy: 95%** | **Retrained all the weights of Mobilenet.**  **Batch size = 10**  **Epochs = 20** |
| **12** | **Mobilenet (fine tune after 50th layer)**  **+ GRU (128 cells)**  **+ Dense (128 nodes)**  **+ image size 120 by 120**  **+ 20 images per video** | **Training Accuracy: 99%**  **Validation Accuracy: 88%** | **Batch size = 10**  **Epochs = 20**  **Fine tuning not giving good validation accuracy.**  **Hence always opting to retrain the full network.** |
| **13** | **Mobilenet (retrain all weights)**  **+ LSTM (128 cells)**  **+ Dense (128 nodes)**  **+ image size 120 by 120**  **+ 20 images per video** | **Training Accuracy: 98.06%**  **Validation Accuracy: 96%** | **Retrained all weights.**  **Batch size = 10**  **Epochs = 20** |
| **14** | **Mobilenet (retrain all weights)**  **+ GRU (128 cells)**  **+ Dense (64 nodes)**  **+ image size 120 by 120**  **+ 20 images per video** | **Training Accuracy: 98.21**  **Validation Accuracy: 94%** | **Retrained all weights.**  **Batch size = 10**  **Epochs = 20** |
| **15** | **Mobilenet (retrain all weights)**  **+ LSTM (128 cells)**  **+ Dense (64 nodes)**  **+ image size 120 by 120**  **+ 20 images per video**  **+ random data transformations on the images** | **Training Accuracy: 98%**  **Validation Accuracy: 93%** | **Retrained all weights.**  **Batch size = 10**  **Epochs = 20** |
|  |  |  |  |
|  |  |  |  |
| **Final Models** | **Conv 3D Model using 20 frames per video**  **+ 16, 32, 64, 128 filters conv 3D layers**  **+ 256 dense nodes**  **+ 128 dense nodes**  **+ image size 120 by 120** | **Training Accuracy: 91%**  **Training loss: 0.2800**  **Validation Accuracy: 86%**  **Validation loss: 0.5032** | **Low loss and high accuracy and comparable between training and validation data sets.**  **Parameters - ~1.9 million.** |
|  | **Mobilenet (retrain all weights)**  **+ LSTM (128 cells)**  **+ Dense (128 nodes)**  **+ image size 120 by 120**  **+ 20 images per video** | **Training Accuracy: 98.06%**  **Training loss: 0.0696**  **Validation Accuracy: 96%**  **Val loss: 0.1266** | **High accuracy and low loss in both training and validation data sets.**  **Total params: ~3.8 million** |

**Conv 3D best model plot as below:**

**Mobilenet + LSTM best model plot as below:**

