#### Gesture Recognition Case study IIITB Assignment 2022

By Ashwini Abhang

The following is a summary of the project's experiments and observations.

* I conducted many more experiments, but these are the important ones, along with explanations.
* I attempted to summarize the model in the model column itself.
* It should be noted that dropouts (which are commonly set to 0.25 in dense connections) and batch normalization play a vital role in regularization but are not explicitly specified in the model column.
* We reduced the batch size to 10 for all models because it did not cause memory concerns and the models trained effectively with it.
* We decided to experiment with either 20 or 30 photos per video.
* We decided to experiment with either a 120 by 120 or a 160 by 160 image size.
* The accuracy and loss values for final models are indicated in yellow.
* Because data transformations/augmentation did not appear to be significantly enhancing accuracy or loss, we began eliminating them in the final constructive model building.
* Here I have used Mobilenet for transfer learning as it is one of the most lightweight architectures in the CNN models and has one of the best accuracies when used for image classification.

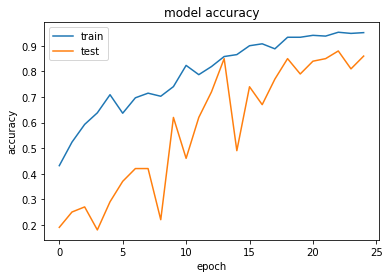
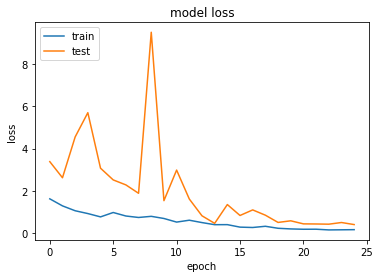
|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Result**  **(in %)** | **Decision + Explanation** |
| **1** | **Conv 3D Model with 30 frames per video**  **16, 32, 64, 128 filters conv 3D layers**  **+ 128 dense nodes + 64 dense nodes + image size 120 by 120** | **Training Accuracy: 96.38**  **Validation Accuracy: 80.00** | **Low validation accuracy as compared to training accuracy.**  **Parameters - 1,933,765** |
| **2** | **Conv 3D Model with 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: 95.32**  **Validation**  **Accuracy: 88.00** | **Great training and validation accuracy with a smaller number of parameters.**  **Parameters - 1,933,765** |
| **3** | **Conv 3D Model with 30 frames per video**  **16, 32, 64, 128 filters conv 3D layers**  **+ 256 nodes + 128 dense nodes +image size 160 by 160** | **Training Accuracy: 93.06**  **Validation Accuracy: 85.00** | **Parameters are on higher side and even though the model accuracy is good, its validation accuracy is less compared to training.**  **Parameters-** **3,604,933** |
| **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: 79.19**  **Validation Accuracy: 75.00** | **The training and validation accuracy are not up to the mark.**  **120 by 120 image size is better for training than 160 by 160 image size. Even with higher number of parameters.**  **Parameters- 3,604,933** |
| **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: 82.81**  **Validation Accuracy: 67.00** | **Accuracy dropped with addition of random data transformation significantly. Also, the validation accuracy is very low.**  **Parameters - 1,933,765** |
| **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: 95.17**  **Validation Accuracy: 81.00** | **Parameters are on lower side and the**  **validation accuracy is very**  **low as compared to training**  **accuracy.**  **Parameters - 911,973** |
| **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: 89.59**  **Validation Accuracy: 87.00** | **The training accuracy and validation accuracy are close but not enough for a good model.**  **Parameters - 1,105,093** |
| **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: 85.37**  **Validation Accuracy: 81.00** | **Increasing number of filters increased the number of parameters and validation accuracy is low only.**  **Parameters - 2,790,149** |
| **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: 88.08**  **Validation Accuracy: 77.00** | **Model seems to be fitting pretty well but very low accuracy.**  **We trained for 15 epochs.**  **Parameters - 1,938,469** |
| **10** | **Mobilenet (retrain all weights)**  **+ GRU (128 cells)**  **+ Dense (128 nodes)**  **+ image size 120 by 120**  **+ 20 images per video** | **Training Accuracy: 99.70**  **Validation Accuracy: 95.00** | **Retrained all the weights of Mobilenet.**  **Batch size = 10**  **Epochs = 20**  **Parameters - 3,693,253** |
| **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: 99.70**  **Validation Accuracy: 95.00** | **Retrained all the weights of Mobilenet. No change from previous model.**  **Batch size = 10**  **Epochs = 20**  **Parameters - 3,693,253** |
| **12** | **Mobilenet (fine tune after 50th layer)**  **+ GRU (128 cells)**  **+ Dense (128 nodes)**  **+ image size 120 by 120**  **+ 20 images per video** | **Training Accuracy:100.00**  **Validation Accuracy:92.00** | **Fine tuning not giving good validation accuracy.**  **Hence opting to retrain the full network.**  **Batch size = 10**  **Epochs = 20**  **Parameters - 3,693,253** |
| **13** | **Mobilenet (retrain all weights)**  **+ LSTM (128 cells)**  **+ Dense (128 nodes)**  **+ image size 120 by 120**  **+ 20 images per video** | **Training Accuracy:99.10**  **Validation Accuracy:96.00** | **Retrained all weights.**  **Batch size = 10**  **Epochs = 20**  **Parameters - 3,840,453** |
| **14** | **Mobilenet (retrain all weights)**  **+ GRU (128 cells)**  **+ Dense (64 nodes)**  **+ image size 120 by 120**  **+ 20 images per video** | **Training Accuracy:99.24**  **Validation Accuracy:97.00** | **Retrained all weights. Best validation performance yet and lowest number of parameters in transfer learning.**  **Batch size = 10**  **Epochs = 20**  **Parameters - 3,684,677** |
| **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:98.64**  **Validation Accuracy:95.00** | **Retrained all weights.**  **Batch size = 10**  **Epochs = 20**  **Parameters - 3,831,877** |

#### Final Models after Observations

|  |  |  |  |
| --- | --- | --- | --- |
| **2** | **Conv 3D Model with 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: 95.32**  **Validation**  **Accuracy: 88.00** | **Great training and validation accuracy with a smaller number of parameters.**  **Parameters - 1,933,765** |

|  |  |  |  |
| --- | --- | --- | --- |
| **14** | **Mobilenet (retrain all weights)**  **+ GRU (128 cells)**  **+ Dense (64 nodes)**  **+ image size 120 by 120**  **+ 20 images per video** | **Training Accuracy:99.24**  **Validation Accuracy:97.00** | **Retrained all weights. Best validation performance yet and lowest number of parameters in transfer learning.**  **Batch size = 10**  **Epochs = 20**  **Parameters - 3,684,677** |

#### Plots for the performance of final models

Conv 3D Model with 20 frames per video 16, 32, 64, 128 filters conv 3D layers + 256 dense nodes + 128 dense nodes + image size 120 by 120

Mobilenet (retrain all weights) + GRU (128 cells) + Dense (64 nodes) + image size 120 by 120 + 20 images per video

