# Gesture Recognition - CNN +RNN

**Approach**

We attempted the following variations for CNN + RNN Architecture

1. Custom Conv 2D + RNN
2. Custom Conv 2D Embedded within the Time Distributed layers + RNN
3. VGGNet16 + RNN
4. Resnet + RNN

Since implementing options 1,2 and 3 turned out challenging, we only confined to the 4th approach which is detailed below

**Generator Function**

Reused the generator function for CNN (2D) + RNN approach from the one used for CNN (3D). Increased the image size according to the requirements of Resnet (requires a minimum of 197 x 197) that’s used for the CNN part to take advantage of transfer learning.

In the CNN (3D) approach, using a small subset of images (12) from the 30 frame video sufficed for the network to learn. However, using the same for CNN (2D) + RNN approach, the validation accuracies turned out to be extremely low. By choosing all 30 frames in the video, the accuracies improved.

**Neural Network Implementation**

Following experiments were carried out:

Resnet (Used the first 140 layers as is and trained the rest of them using SGD with learning rate of 0.001 applied)

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Experiment No | Network Configuration | Optimizer - Learning Rate | Number of Epochs | Batch size | Image details | Sample Size (Train, val) | Parameters (Normalization etc) | Train/ Validation Accuracy | Decision and Explanation |
| 1 | Resnet + LSTM – 1 layer (12 neurons) | SGD - None | 5 | 30 | 197x197, 12 frames | 100, 20 | (Image – Min)/ (Max – Min)  Resnet  LSTM(12) |  | Out of memory issue, reduce batch size |
| 2 | Resnet + LSTM – 1 layer (12 neurons) | SGD - None | 5 | 15 | 197x197, 12 frames | 100, 20 | (Image – Min)/ (Max – Min)  Resnet  LSTM(12) | 20%/ NAN | Neural Network is not learning |
| 3 | Resnet + LSTM – 1 layer (12 neurons) | SGD – 0.1, 0.01, 0.001, 0.0001, 0.00001 | 5 | 15 | 197x197, 12 frames | 100, 20 | (Image – Min)/ (Max – Min)  Resnet  LSTM(12) | 20%/ NAN | Neural Network is not learning |
| 4 | Resnet + LSTM – 1 layer (12 neurons) | Adam - 0.01, 0.001 | 5 | 15 | 197x197, 12 frames | 100, 20 | (Image – Min)/ (Max – Min)  Resnet  LSTM(12) | 20%/ NAN | Neural Network is not learning |
| 5 | Resnet + LSTM – 1 layer (12 neurons) | RMSProp - 0.1, 0.01, 0.001 | 5 | 15 | 197x197, 12 frames | 100, 20 | (Image – Min)/ (Max – Min)  Resnet  LSTM(12) | 20%/ NAN | Neural Network is not learning |
| 6 | Resnet + GRU – 3 layers (64 neurons) | SGD – 0.001 | 5 | 15 | 200x200, 30 frames | 100, 50 | Dropout added after each layer | 19%/ 22% | Very low accuracies |
| 7 | Resnet + GRU – 1 layer (64 neurons) | SGD – 0.001 | 10 | 15 | 200x200, 30 frames | 100, 50 | Dropout added after each layer | 95%/ 15% | Changed the optimizer |
| 8 | Resnet + GRU – 3 layers (128 neurons) | SGD – 0.001 | 10 | 15 | 200x200, 30 frames | 100, 50 | Dropout after last layer and dense | 45%/ 0% | Added dropout to handle overfitting |
| 9 | Resnet + LSTM – 1 layer (128 neurons) | SGD - 0.001 | 5 | 15 | 197x197, 30 frames | 100, 20 | Mean Normalization  Resnet  LSTM(128)  Used L2 (0.1) regularizer at the Dense layer | 97%/ 65% | - Image normalization was done using mean normalization  Try on higher number of epochs  - Increase the number of frames since there is no training  - Include complete data set |
| 10 | Resnet + LSTM – 1 layer (128 neurons) | SGD - 0.001 | 10 | 15 | 224x224, 30 frames | 663, 100 | Mean Normalization  Resnet  LSTM(128)  Used L2 (0.1) regularizer at the Dense layer | 65%/ 36% | - Though the parameters worked on smaller sample size, it’s not scaling for the complete data set  - Increase the learning rate, batch size |
| 11 | Resnet + LSTM – 1 layer (128 neurons) | SGD - 0.1 | 10 | 15 | 224x224, 30 frames | 663, 100 | Mean Normalization  Resnet  LSTM(128)  Used L2 (0.1) regularizer at the Dense layer | 65%/ NAN | NAN is due to the imbalance in image size, number of batches, modify them |
| 12 | Resnet + LSTM – 1 layer (128 neurons) | SGD - 0.01 | 10 | 25 | 224x224, 30 frames | 663, 100 | Mean Normalization  Resnet  LSTM(128)  Used L2 (0.1) regularizer at the Dense layer | 60%/ 35% | Learning is not good enough, further reduce the learning rate, increase batch size |
| 13 | Resnet + LSTM – 1 layer (128 neurons) | SGD - 0.001 | 10 | 27 | 224x224, 30 frames | 663, 100 | Mean Normalization  Resnet  LSTM(128)  Used L2 (0.1) regularizer at the Dense layer | 89%/ NAN | Since there is an imbalance due to increase in batch size, further reduce learning rate and the image size |
| 14 | Resnet + LSTM – 1 layer (128 neurons) | SGD - 0.0005 | 10 | 27 | 197x197, 30 frames | 663, 100 | Mean Normalization  Resnet  LSTM(128)  Used L2 (0.1) regularizer | 94%/ 62% | The best so far among all the experiments carried out |
| 15 | Resnet + LSTM – 1 layer (128 neurons) | SGD – 0.0005 | 10 | 26 | 197x197, 30 frames | 663, 100 | Mean Normalization  Resnet  LSTM(128)  Used dropouts after LSTM and FC layers | 100%/ 20% | Validation accuracy failed to improve |
| 16 | Resnet + LSTM – 1 layer (128 neurons) | SGD – 0.0005 | 10 | 26 | 197x197, 30 frames | 663, 100 | Mean Normalization  Resnet  LSTM(128)  Used Batch Normalization after LSTM without dropouts | 100%/ 20% | Validation accuracy failed to improve even after introducing batch normalization |
| 17 | Resnet + LSTM – 1 layer (128 neurons) | SGD – 0.0005 | 10 | 26 | 197x197, 30 frames | 663, 100 | Mean Normalization  Resnet  LSTM(128)  Used Batch Normalization after LSTM with dropouts | 99%/28% | Validation accuracy failed to improve even after introducing batch normalization and dropouts |
| 18 | Resnet + LSTM – 1 layer (128 neurons) | SGD – 0.0005 | 10 | 27 | 197x197, 30 frames | 663, 100 | 2 LSTM layers (128) |  | Failed with out of memory issue |
| 19 | Resnet + LSTM – 2 layers (64 neurons) | SGD – 0.0005 | 10 | 27 | 197x197, 30 frames | 663, 100 | 2 LSTM layers (64) | 80%/ 61% | Need to try on more epochs, there is more scope to improve the accuracy |
| 20 | Resnet + LSTM – 1 layer (128 neurons) | SGD – 0.0005 | 15 | 27 | 197x197, 30 frames | 663, 100 | 2 LSTM layers (128) | 94%/ 49% |  |
| 21 | Resnet + LSTM – 1 layer (128 neurons) | SGD – 0.0005 | 12 | 27 | 197x197, 30 frames | 663, 100 | 1 LSTM layer (128) | 99%/ 71% | With 12 epochs, the training and validation accuracies were obtained |