**Deep Learning – Gesture Recognition Writeup**

The necessary image data pre-processing steps like image cropping, resizing and normalizing images, etc. are implemented in the generator function.

* Input images are resized to 80x80 size
* Batch size of 32 sequences is considered, taking into account the neural network configuration, number of parameters, performance and GPU load.
* 15 odd frames were considered for each sequence
* Epochs: 30 and validation losses are monitored

Multiple Experiments were performed on the given train and validation dataset. The following table includes the details of all the 16 Experiments performed:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Expt. No.** | **Model  Name** | **Model** | **Result** | **Parameters** | **Decision + Explanation (Observations)** |
| 1 | - | - | OOM Error | - | Reduce the batch size and reduce the number of neurons in layers (reducing overall model parameters) to avoid GPU overload |
| 2 | Model1 | Conv3D | Training Accuracy: 91.88%  Validation Accuracy: 87.5%  *(Epoch 27)* | 1,663,829 | **Model1:** 3 Conv3D (16,16,32) layers, kernel size= (3,3,3), Batch Normalization (BN), Maxpooling3D and include Dropouts at Fully Connected (FC) layer group  **Decision:** Increase the number of neurons to check improvement in model performance  **Observations:** Model overfits. Parameters are substantially less as dropouts at FC's and BN are implemented for Conv3D layers |
| 3 | Model2 | Conv3D | Training Accuracy: 87.68%  Validation Accuracy: 87.5%  *(Epoch 19)* | 1,684,645 | **Model2:** 3 Conv3D (16,32,32) layers, kernel size=(3,3,3), Batch Normalization(BN), Maxpooling3D, and include Dropouts at Fully Connected (FC) layer group  **Decision:** Increase neurons in layer 2  **Observations:** Increasing the number of neurons in layer 2 to 32 has stabilized the model and the model shows improved performance. Regularization effect is seen because of dropouts at FCs and BN at Conv3D layers. Parameters are slightly more than Model 1 but acceptable. |
| 4 | Model3 | Conv3D | Training Accuracy: 81.23%  Validation Accuracy: 81.25%  *(Epoch 27)* | 3,350,853 | **Model3:** 3 Conv3D (16,32,64) layers, kernel size=(3,3,3), Batch Normalization(BN), Maxpooling3D and include Dropouts at Fully Connected (FC) layer group  **Decision:** Check for increased neurons in layer 2  **Observations:** Increase in the neurons in the 3 layer (64) resulted in increased number of parameters. The Performance is not up to the mark in terms of accuracy Incorporated Dropouts at FC's and BN for Conv3d layers |
| 5 | Model4 | Conv3D | Training Accuracy: 88.52%  Validation Accuracy: 87.5%  *(Epoch 30)* | 3,420,133 | **Model4:** 3 Conv3D (16,64,64) layers, kernel size=(3,3,3), Batch Normalization(BN), Maxpooling3D and include Dropouts at Fully Connected (FC) layer group  **Decision:** Decrease neurons in layer 2 and 3 to 32. Change kernel size to (2,2,2)  **Observations:** Increase in the neurons in the 2nd layer (64) resulted in increased number of parameters. The model performance doesn't seem to change much w.r.t. Model3. Incorporated Dropouts at FC's and BN for Conv3d layers. |
| 6 | Model5 | Conv3D | Training Accuracy: 86.83%  Validation Accuracy: 87.5%  (Epoch 21) | 1,718,293 | **Model5:** 3 Conv3D (16,32,32) layers, kernel size=(2,2,2), Batch Normalization(BN), Maxpooling3D and include Dropouts at Fully Connected (FC) layer group (2 Dense FCs)  **Decision:** Check for a larger kernel size of (5,5,5)  **Observations:** Reduction in the number of overall parameters. Also, the model not as stable as Model2 |
| 7 | Model6 | Conv3D | Training Accuracy: 62.75%  Validation Accuracy: 62.5%  *(Epoch 19)* | 1,903,621 | **Model6:** 3 Conv3D (16,32,32) layers, kernel size=(5,5,5), Batch Normalization(BN), Maxpooling3D, and include Dropouts at Fully Connected (FC) layer group (2 Dense FCs)  **Decision:** Check for a larger neural network (32,64,64) Conv3D layers and kernel size of (3,3,3)  **Observations:** Model performance deteriorates. Comparatively lower train and validation accuracies. Losses are also quite high. |
| 8 | Model7 | Conv3D | Training Accuracy: 77.56%  Validation Accuracy: 75%  *(Epoch 30)* | 3,512,901 | **Model7:** 3 Conv3D (32,64,64) layers, kernel size=(3,3,3), Batch Normalization(BN), Maxpooling3D, and include Dropouts at Fully Connected (FC) layer group (2 Dense FCs)  **Decision:** Check for performance of Conv2D+RNN models  **Observations:** Doubled the number of neurons as compared to those in model6 (in all layers) and reduction in kernel size to (3,3,3) have almost doubled the number of total parameters. Model accuracy shows a spike yet not that significant. Model is not stable across epochs. |
| 9 | Model8 | Conv2D  +(LSTM) | Training Accuracy: 85.15% Validation Accuracy: 81.25%  *(Epoch 30)* | 6,917,029 | **Model8:** 3 Conv2D (32,64,128) pairs of layers, kernel size=(3,3), Batch Normalization(BN), Maxpooling2D, and include Dropouts at Fully Connected (FC) layer group (with LSTM)  **Decision:** Check for the performance of a smaller Conv2D+LSTM neural network (16,32,64) Conv2D pairs layers with LSTM  **Observations:** Validation losses are higher as we see the model performance over epochs. Model parameters are very high in number. |
| 10 | Model9 | Conv2D  +(LSTM) | Training Accuracy: 87.96%  Validation Accuracy: 87.5%  *(Epoch 24)* | 3,424,405 | **Model9:** 3 Conv2D (16,32,64) pairs of layers, kernel size=(3,3), Batch Normalization(BN), Maxpooling2D, and include Dropouts at Fully Connected (FC) layer group (with LSTM)  **Decision:** Check for the performance of a larger Conv2D+GRU neural network (32,64,128) Conv2D pairs layers with GRU  **Observations:** Validation losses are higher as we see the model performance over epochs. Model parameters are almost double in comparison to Model2. Also, accuracies are almost the same as 2nd Model but Model2 seems to more stable. |
| 11 | Model10 | Conv2D  +(GRU) | Training Accuracy: 83.47%  Validation Accuracy: 81.25%  *(Epoch 21)* | 5,262,117 | **Model10:** 3 Conv2D (32,64,128) pairs of layers, kernel size=(3,3), Batch Normalization(BN), Maxpooling2D, and include Dropouts at Fully Connected (FC) layer group (with GRU)  **Decision:** Check for the performance of a smaller Conv2D+GRU neural network (16,32,64) Conv2D pairs layers with GRU  **Observations:** Validation losses are higher as we see the model performance over epochs. Model parameters are lesser than similar Conv2D+LSTM Model 8 |
| 12 | Model11 | Conv2D  +(GRU) | Training Accuracy: 82.35%  Validation Accuracy: 87.5%  *(Epoch 25)* | 2,588,693 | **Model11:** 3 Conv2D (16,32,64) pairs of layers, kernel size=(3,3), Batch Normalization(BN), Maxpooling2D, and include Dropouts at Fully Connected (FC) layer group (with LSTM)  **Decision:** Check model performance using Transfer Learning (optional)  **Observations:** Validation losses are slightly higher as we see the model performance over epochs. parameters are lesser than similar Conv2D+LSTM Model9. Overall model is satisfactory yet not the best. |
| 13 | Model12 | Transfer Learning (optional) | Training Accuracy: <65%  Validation Accuracy: <55% | 3,831,877 | **Model12:** MobileNet Transfer Learning with LSTM (All Layers Non-Trainable)  **Decision:** Check the same for GRU  **Observations:** Model overfits. Highly unstable model with poor performance and fluctuating accuracies and losses. |
| 14 | Model13 | Transfer Learning (optional) | Training Accuracy: <55%  Validation Accuracy: <45% | 3,684,293 | **Model13:** MobileNet Transfer Learning with GRU (All Layers Non-Trainable)  **Decision:** Check model performance for Trainable Layers  **Observations:** Model overfits. Highly unstable model with poor performance and fluctuating accuracies and losses. |
| 15 | Model 14 | Transfer Learning (optional) | Training Accuracy: 95.24%  Validation Accuracy: 93.75%  *(Epoch 27)* | 3,831,877 | **Model14:** MobileNet Transfer Learning with LSTM (All Layers Trainable)  **Decision:** Check the same for GRU  **Observations:** Improved Model Performance. High train and validation accuracies achieved. |
| 16 | Model 15 | Transfer Learning (optional) | Training Accuracy: 99.16%  Validation Accuracy: 93.75%  *(Epoch 24)* | 3,684,293 | **Model15:** MobileNet Transfer Learning with GRU(All Layers Trainable)  **Decision:** Best Stable Model: Model 2  **Observations:** Improved Model Performance but not as stable as Model14. The variation in validation losses is quite high over epochs. The model is slightly unstable. High train and validation accuracies achieved at Epoch 24. |
| **\*\*\*** | **(Final Model)  Model2** | **Conv3D** | **Training Accuracy: 87.68%**  **Validation Accuracy: 87.5%**  ***(Epoch 19)*** | **1,684,645** | **Model2: 3 Conv3D (16,32,32) layers, kernel size=(3,3,3), Batch Normalization(BN), Maxpooling3D, and include Dropouts at Fully Connected (FC) layer group**  **Decision:** Best Stable Model: Model 2  **Observations: Out of all the Conv3D and Conv2D+RNN models, Model 2 seems to be the best and stable model for the following reasons: (1) High training and validation accuracies. Validation loss seems to be lesser than the training loss. (2) Losses substantially decrease over epochs resulting in better and stable model. (3) Efficient simple neural network with a smaller number of parameters** |