Gesture Recognition – Deep learning

Problem Statement:We need to develop a cool feature in the smart-TV that can recognise 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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Hyper-Parameters** | **Result** | **Decision + Explanation** |
| **1** | **Conv3D** | Batch size =25,  Epoch = 10,Seq Length =15,Dim = 120x120  LR = 0.001 | Train Accuracy: 0.50,Validation Accuracy: 0.45 | **Model is average, it is not learning much** |
| **2** | **Conv3D** | Batch size =30,  Epoch = 15,Seq Length =15,Dim = 120x120  LR = 0.001 | Train Accuracy: 0.58,Validation Accuracy: 0.39 | **After increasing epoch size, model is not improving but becomes more overfit** |
| **3** | **Conv3D** | Batch size =30,  Epoch = 15,Seq Length =15,Dim = 120x120  LR = 0.001  BatchNormalisation | Train Accuracy: 0.55,Validation Accuracy: 0.38 | **No impact after batch normalisation** |
| **4** | **Conv3D** | Batch size =30,  Epoch = 15,Seq Length =15,Dim = 120x120  LR = 0.001  BatchNormalisation  Dropout | Train Accuracy: 0.80,Validation Accuracy: 0.30 | **Training accuracy increased but validation accuracy decreased after using dropout and become more overfit** |
| **5** | **Conv3D** | Batch size =30,  Epoch = 15,Seq Length =15,Dim = 120x120  LR = 0.001  BatchNormalisation  Dropout  Increase layers | Train Accuracy: 0.61,Validation Accuracy: 0.38 | **Not much improvement after increasing number of layers** |
| **6** | **CNN+LSTM** | Batch size =25,  Epoch = 15,Seq Length =15,Dim = 120x120 | Train Accuracy: 1.00,Validation Accuracy: 0.73 | **Our model improved but over fitting** |
| **7** | **CNN+LSTM** | Batch size =25,  Epoch = 25,Seq Length =15,Dim = 120x120  Increase Layers | Train Accuracy: 0.92,Validation Accuracy: 0.75 | **Not getting much difference after increasing number of layers** |
| **8** | **CNN+LSTM** | Batch size =25,  Epoch = 25,Seq Length =15,Dim = 160x160 | Train Accuracy: 0.83,Validation Accuracy: 0.69 | **No improvement after increasing dimension of input image from 120\*120 to 160\*160** |
| **9** | **CNN+GRU** | Batch size =30,  Epoch = 25,Seq Length =15,Dim = 120x120 | Train Accuracy: 1.00,Validation Accuracy: 0.78 | **Using GRU and getting good results rather than conv3d and cnn+lstm** |
| **10** | **CNN+GRU** | Batch size =30,  Epoch = 35,Seq Length =15,Dim = 120x120  BatchNormalisation | Train Accuracy: 0.98,Validation Accuracy: 0.74 | **Validation accuracy is not increasing after using batch normalization also** |

**Conclusion:**The Model built with Time distributed Conv2D and GRU ( Model number 9) gave better results  
compared to all the other models and also able to achieve a validation accuracy of 78%