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.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Exp. #** | **Model** | **Hyper- parameters** | **Result** | **Decision + Explanation** |
| **1** | **Conv3D** | **Batch size = 30,**  **Filter Size = (5,5,5)**  **No of Frames = 30,**  **Epoch = 20,**  **Dim = 120x120** | **Train Accuracy: 0.277,**  **Validation Accuracy: 0.525** | **The Model is not learning anything and it is a highly underfit model.**  **Dropout layer was added in the convolution layer. Reducing the batch size to further 20.** |
| **2** | **Conv3D** | **Batch size = 20,**  **Filter Size = (5,5,5)**  **No of Frames = 20,**  **Epoch = 30,**  **Dim = 120x120** | **Train Accuracy: 0.3337,**  **Validation Accuracy: 0.3939** | **Model was early stop at 18th epochs as validation loss did not improve.**  **Dropout layer was removed from the convolution layer as model was underfitting.** |
| **3** | **Conv3D** | **Batch size = 30,**  **Filter Size = (3,3,3)**  **No of Frames = 20,**  **Epoch = 30,   Dim = 120x120** | **Train Accuracy: 0.7395,**  **Validation Accuracy: 0.3250** | **The model is overfitting even after reducing the parameters.  Therefore we will try experimenting with CNN RNN Architecture.** |
| **4** | **CNN-RNN**  **Time Distributed (LSTM)** | **Batch size = 20,**  **Filter Size = (3,3)**  **No of Frames = 20,**  **Epoch = 30,   Dim = 120x120** | **Train Accuracy: 0.7486,**  **Validation Accuracy: 0.8** | **We can observe that the val loss has decreased for 2.00046 in the Conv3D model to**  **0.6711 in the CNN-LSTM Model.**  **SO the model will perform better compare to Conv3D model.** |
| **5** | **CNN-RNN**  **Time Distributed**  **(GRU)** | **Batch size = 20,**  **Filter Size = (3,3)**  **No of Frames = 20,**  **Epoch = 25,   Dim = 120x120** | **Train Accuracy: 0.7854,**  **Validation Accuracy: 0.72** | **Since we are using GRU Less no of parameters are being used and the overfitting is also reduced due to addition of dropout in a layer.**  **Now we will be using Pretrained model from transfer learning to further improve the accuracy.** |
| **6** | **CNN-RNN**  **Transferred learning with mobilenet (GRU)** | **Batch size = 10,**  **No of Frames = 20,**  **Epoch = 20,  Dim = 120x120** | **Train Accuracy: 0.97,**  **Validation Accuracy: 0.99** | **Accuracy has improved to a significant level and Val loss has reduced.** |

# Conclusion:

The Model built with Time distributed Conv2D and ConvLSTM2D gave better results compared to all the other Conv3D models and also the model has very least number of parameters compared to other models.

As per the above experiment we have decided to go with CNN-RNN Transferred Learning (GRU) model because it gives the best accuracy as compared to any other model that have been trained.