|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Exp. #** | **Model** | **Hyper- parameters** | **Result** | **Decision + Explanation** |
| **1** | **Conv3D** | **Batch size = 128,**  **Ablation = 20, Augmentation = False,**  **LR = 0.01,**  **Seq Length = 10,**  **Epoch = 20, Dim = 120x120** | **Train Accuracy: 0.15,**  **Validation Accuracy: 0.15** | **The Model is not learning anything throughout the epochs, the loss is not decreasing. Reducing the batch size further.** |
| **2** | **Conv3D** | **Batch size = 32** | **Train Accuracy: 0.15,**  **Validation Accuracy: 0.20** | **No improvement in the model, lets add more layers to the model so that it can learn from data.** |
| **3** | **Conv3D** |  | **Negative Dimension Error.** | **The new CNN kernel sizes are not compatible with the output of previous layers. Let’s reduce the kernel size of new layers.** |
| **4** | **Conv3D** |  | **Train Accuracy: 0.20,**  **Validation Accuracy: 0.20** | **Still there is no improvement in the model. Let’s add Batch normalization layers after every CNN and dense layers.** |
| **5** | **Conv3D** |  | **Train Accuracy: 0.9062,**  **Validation Accuracy: 0.2708** | **Model is able to over-fit on less data (Ablation data set), Let’s Training on full data and increasing epochs to 50.** |
| **6** | **Conv3D** | **Ablation = None, Epoch = 50** | **Train Accuracy: 0.9062,**  **Validation Accuracy: 0.70** | **Mode is having over-fitting as there is huge gap between training and validation accuracies. Let’s add some dropouts that the model can be generalized.** |
| **7** | **Conv3D** | **Dropout = 0.2** | **Train Accuracy: 0.9896,**  **Validation Accuracy: 0.7734** | **There is a bit of increase in the model validation accuracy and training accuracy also. Lets increase the drop out values from**  **0.2 to 0.5** |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **8** | **Conv3D** | **Dropout = 0.5** | **Train Accuracy: 0.9777,**  **Validation Accuracy: 0.5391** | **After increase the dropout the model validation score further reduced and the model is over-fitted. Let’s use 0.2 only remove a CNN layer to reduce the complexity**  **of the model.** |
| **9** | **Conv3D** | **Dropout = 0.2** | **Train Accuracy: 1.00,**  **Validation Accuracy: 0.77** | **Still the model is over-fitting. Let’s use a Global Average Pooling instead of Flatten Layer.** |
| **10** | **Conv3D** |  | **Train Accuracy: 0.9509,**  **Validation Accuracy: 0. 9062** | **The model is wonderful and the training and validation scores are good. The model has 710,533 trainable parameters. Let’s try architectures too.** |
| **11** | **Time Distributed**  **+ GRU** |  | **Train Accuracy: 0.9554,**  **Validation Accuracy: 0. 8203** | **The model is working quite well on validation dataset with less trainable parameters(98,885), Lets add some drop outs after each layer, so that both train and**  **validation accuracies will be closure.** |
| **12** | **Re-trained MobileNet + GRU.** |  | **Train Accuracy: 0.9881,**  **Validation Accuracy: 0.7188** | **The Model has high accuracy , the most accurate model for the least parameter footprint (model size < 50MB) was found to be Re-trained MobileNet + GRU.** |

**Conclusion**

* The problem of recognising gestures has been treated as a video classification problem.
* After thorough experimentation with Conv3d and CNN-RNN models, the most accurate model for the least parameter footprint (model size < 50MB) was found to be Re-trained MobileNet + GRU.
* The raw videos of size 120x160 and 320x320 have been cropped and resized to 160x160 to meet the requirements of MobileNet architecture. Since its a 15fps video, all 30 frames have been used for training.
* Experiments with image augmentation using 'skin detection' have been performed but found the model trained on raw data gives decent results.
* In particular, training the model for 30 epochs gives validation accuracy of 71% with training accuracy of 99%.
* In the future, image augmentation like translation, resizing, color transformations could be performed to create a more generalised model with higher val accuracy.