**Gesture Recognition :**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Problem/Experiment** | **Decision + Explanation** | **Result** |
| **1** | **Conv3D** | **Observed images are of different sizes** | **Resized all images to 100\*100** |  |
| **2** | **Conv3D** | **Model training was hanging due to limited 8GB RAM on Mac** | **Reduced batch size from 64 to 32 and started with ablation experiment by overfitting on subset data** |  |
| **3** | **Conv3D** | **Model architectures** | **Final Arch: Conv3D 32 relu and 64 relu stacked with maxpolling3d after each conv. Flatten to connect to dense layer of 128 with tanh. Followed by softmax of 5 categories.** | **Train Acc : 89% Val Acc: 65%** |
| **4** | **Conv3D** | **Filter size Optimzation** | **Used filter of size 2\*5\*5 gave best results giving more importance to depth dimension. Gave regularisation effect due to 5\*5 filter** | **Train Acc : 87% Val Acc: 76%** |
| **5** | **Conv3D** | **Image resize vs cropping** | **Image resize gave better results.** | **Train Acc : 87% Val Acc: 76%** |
| **6** | **Conv3D** | **Adding layers/changing activations** | **Tanh in dense layer gave better results. Adding more layers caused overfitting and not helped.** | **Train Acc : 87% Val Acc: 76%** |
| **7** | **Conv3D** | **Learning rate optimzation** | **Tried LR of 0.001. LR of 0.0005 gave best result for selected arch. Applied decay of ReduceLROnPlateau on valloss for patience of 2** | **Train Acc : 87% Val Acc: 76%** |
| **8** | **Conv3D** | **Changing image samples in sequence** | **Used samples of index 2,4.. 28 in sequence considering probable redundancy in consecutive frames** | **Train Acc : 87% Val Acc: 76%** |
| **9** | **Conv3D** | **Change batch size** | **Changed from 64 to 32. Helped faster training on local GPU and also gave better accuracy results** | **Train Acc : 87% Val Acc: 76%** |
| **10** | **Conv3D** | **Increase Epoch** | **Started with 10 epoch and checked whether increasing epochs further helps better training. Eventually stick to 10 as val accuracy doppedpost that for current arch** | **Train Acc : 87% Val Acc: 76%** |
| **11** | **Conv3D** | **Drop outs** | **Filter size 2,5,5 already acted as regulariser due to 5,5 filter so without dropouts gave better result.** | **Train Acc : 87% Val Acc: 76%** |
| **12** | **Conv3D** | **Batch Normalisation Experiment** | **Performance degrades after batch norm with current arch/various settings so not used.** | **Train Acc : 60% Val Acc: 25%** |
| **13** | **Conv3D** | **Created new data using OpenCV video capture.** | **Able to detect gestures real time.** |  |
| **14** | **Convolution (Xception transfer learning) + LSTM** | **Model architecture** | **Tried Convolution +LSTM. Tried adding dropouts etc. but could not gave better results than Conv3D** | **Train acc : 96% Val Acc : 60%** |
| **Final Model** | **Conv3D** | **Conv3D 32 Relu**  **MaxPolling3D**  **Conv3D 64 Relu**  **MaxPolling3D**  **Dense tanh 128**  **Dense softmax 5** | **Number of parameters : 255,525**  **Training time : 50 sec per epoch. Total : 500 sec** | **Train Acc : 87% Val Acc: 76%** |