**Gesture Recognition**

This Document contains a detailed explanation of how and why the different decisions are taken to obtain a final model with a decent accuracy.

* **Different Experiments are conducted to find an optimal model and it’s accuracy:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Result** | **Decision + Explanation** |
| **1** | **Conv3D** | Model not trainable as a lot of parameters | Reduce the number of layers or parameters or no. of images per video. |
| **2** | **Conv3D** | Model not trainable as a lot of parameters (*ResourceExhaustedError*) | Reduced no. of parameters |
| **3** | **Conv3D** | Epochs: 30  Accuracy: 0.43182 | Reduce no. of parameters and no. of Epochs. |
| **4** | **Conv3D** | Epochs: 20  Accuracy: 0.50994 | Max accuracy |
|  |  |  |  |
| **5** | **Conv-GRU(16)** | Accuracy: 0.3045 | Try Conv-LSTM as Conv-GRU not giving very less accuracy |
| **6** | **Conv-LSTM(16)** | Accuracy: 0.3817 | Increase no. of parameters as not giving desired accuracy |
| **Final Model** | **Conv-LSTM(64)** | Accuracy: 0.6094 | Optimal Model with good Accuracy. |

1. Experiment 1 is an initial attempt in model building. Ideally, two types of model can be used for Gesture Recognition, Conv3D and CNN+RNN.

1st attempt includes a Convolution-3D Model with input as (64,20,84,84,3) where batch size is 64, No. of images per video are 20, Images size is 84x84 and 3 channels (RBG).

The Model with 3 layers gave 300k+ parameters which was not trainable. So reduced the images per videos from 20 to 15.

1. Experiment 2 contains model with input as (64,15,84,84,3), but the parameters were still more than 175k which was not trainable and gave error as “*ResourceExhaustedError”.So* Reduced no. of parameters per layer.
2. Experiment 3 was trainable with accuracy of 43% and 30 epochs which was less. So decreased the epochs from 30 to 20.
3. Experiment 4 contains a decent Conv3D model with approx. 51% accuracy with 20 epochs. So trying out combination of CNN+RNN.
4. Experiment 5 is a low accuracy model with very less model parameters and used GRU(16). So jumped to LSTM(16) as it has more gates.
5. Experiment 6 got more accuracy than GRU but not desirable. No. of parameters were less. So increased parameters.
6. Final model uses LSTM(64) which gave a good accuracy of approx. 61%.

Hence it comes out to be the best model for Gesture Recognition .