# Gesture Recognition

The problem statement is to recognition the gesture in a video sequence. The 2 types of RNN architecture options that we have are – CNN 3D and CNN+RNN. The task given was to identify a good model with minimum parameters.

Models Built

As part of this assignment, we have tried various models including Conv3D and CNN+RNN LSTM architecture. Have tried models by trying multiple architectures, various values for epochs, batch size number of frames etc. During the model training, we noticed overfitting at various points and below are few techniques that were followed to fix the problem are Data Augmentation, Dropout, Regularization, Batch Normalization and Reduce Model Complexity.

We built approx. 25 models to explore (under Model Exploration section in python notebook) and based on the various trials done and the learnings, we had built few more final models (under Final Model section in python notebook). Among the final models, we shortlisted 2 for the final comparison.

The models are -

|  |  |  |
| --- | --- | --- |
|  | Model 1 | Model 2 |
| **Type** | Conv 3D | Conv 3D |
| **Model Name** | Finalmodel1a | Finalmodel2 |
| **Architecture** | conv\_filters=[4, 8, 16, 32],  kernel\_size=(3,3,3),  dense\_nodes=[32],  dropout=0.1,  augment=yes  crop=yes  regularization\_strength=0.01 | conv\_filters=[8, 16, 32,64],  kernel\_size=(3,3,3), dense\_nodes=[64],  dropout=0.1  augment=yes  crop=yes  regularization\_strength=0.01 |
| **epochs** | 25 | 20 |
| **Batch Size** | 10 | 10 |
| **No. of frames** | 20 | 20 |
| **Total Parameters** | 70,325 | 279,333 |
| **Training Score** | Accuracy = 0.8084  Loss = 0.6926 | Accuracy = 0.9065  Loss = 0.5720 |
| **Validation Score** | Accuracy = 0.76  Loss = 0.7427 | Accuracy = 0.87  Loss = 0.7135 |
| **Accuracy Graph** |  | A graph of a graph  Description automatically generated |
| **Loss Function Graph** |  |  |
| **Comments** | In this model, we can notice that the validation accuracy has zig zags which may mean its unstable. | Validation accuracy curve seems more gradual |

Final Model

We have chosen Model 2 for the submission although it has slightly more parameters compared to model 1. Among the models we have explored this model seems to give good accuracy with minimum parameters.

Detailed Summary

Few factors affecting the number of hyperparameters that have been explored as part of this assignment:

1. CNN Architecture Parameters:
   1. Number of Convolutional Layers
   2. Filter Parameters
   3. Kernel Size
   4. Pooling
2. *RNN Architecture Parameters:*
3. RNN Type (LSTM or GRU): Chose LSTM for exploration.
4. Number of RNN Layers
5. RNN Units
6. *Input Data Shape and Sequence Length*
7. *Batch Size*
8. *Number of Frames in a Video*

Below table summarizes the various models that were built for this exercise. Please note Mid way through the model exploration we realized we didn’t crop the images post which it was added. The initial models were built with just resizing.

Conv3D

A screenshot of a computer

Description automatically generated

CNN + RNN

A screenshot of a computer

Description automatically generated

Can refer to the below sheet for better viewing.

