**Gesture Recognition Case Study Write Up**

**Problem Statement**

Imagine you are working as a data scientist at a home electronics company that manufactures state-of-the-art smart televisions. You want to develop a cool feature in the smart TV that can recognize five different gestures performed by the user which will help users control the TV without using a remote.

The gestures are continuously monitored by the webcam mounted on the TV. Each gesture corresponds to a specific command:

| **Gesture** | **Corresponding Action** |
| --- | --- |
| Thumbs Up | Increase the volume. |
| Thumbs Down | Decrease the volume. |
| Left Swipe | 'Jump' backward 10 seconds. |
| Right Swipe | 'Jump' forward 10 seconds. |
| Stop | Pause the movie. |

Each video is a sequence of 30 frames (or images).

|  |  |  |  |
| --- | --- | --- | --- |
| Experiment/ Model Number | Model | Result | Decision + Explanation |
| 1 Batch size= 40  no. of Epochs= 15  frames to sample= 20  image dim= 160X160 | **Conv3D** | **Early stopping**  **Parameters=1,117,061**  **Cat\_Accuracy=98%**  **Val\_Accuracy=23%** | **The model is highly overfitting. So for the next model, we reduce the batch size.**  **Also, reduce the filter size.** |
| 2 Batch size= 30  no. of Epochs= 25  frames to sample= 16  image dim= 120 X 120  reduced the filter size to 2,2,2. | **Conv3D** | **Early stopping**  **Parameters=1,762613**  **Cat\_Accuracy=72%**  **Val\_Accuracy=25%** | **There is an increase in the number of parameters. Also, the accuracy was reduced. The model is still overfitting. The number of parameters increases which we need to reduce.** |
| 3 Batch size= 15  no. of Epochs= 15  frames to sample= 16  image dim= 120 X 120 | **Conv3D** | **Cat\_Accuracy=72%**  **Val\_Accuracy=38%**  **( till Epoch 9/15 )** | **We increased the number of layers. reduced the batch size. In order to improve the accuracy and reduce the number of parameters. Due to this, there is an increase in overall accuracy.** |
| 4 Batch size= 10  no. of Epochs= 15  frames to sample= 16  image dim= 120X120 | **Conv3D** | **Cat\_Accuracy=%**  **Val\_Accuracy=%** | **In this model, we have reduced the batch size so as to increase the accuracy. Also, we have more layers to it.** |
| 5 Batch size= 15  no. of Epochs= 10  frames to sample= 15  image dim= 120X120 | **CNN + LSTM** | **Cat\_Accuracy=97%**  **Val\_Accuracy=49%** | **This model took less time in comparison to other models. there is a significant increase in the accuracy but the problem of overfitting exists. Let’s do some augmentation in the next model.** |
| 5A Batch size= 15  no. of Epochs= 10  frames to sample= 15  image dim= 120X120 | **CNN + LSTM**  **augmentation** | **Cat\_Accuracy=85%**  **Val\_Accuracy=30%** | **There has been no major improvement in the performance of the model. Let’s try the model with GRU** |
| 6 Batch size= 10  no. of Epochs= 15  frames to sample= 16  image dim= 120X120 | **CNN + LSTM**  **With GRU** | **Cat\_Accuracy=95%**  **Val\_Accuracy=84%** | **Using the GRU reduces the parameters and also accuracy increases. So we will be using Model 6 as our final model as it is the only model which gives better performance than the others.** |

**Observations:**

The Model built with CNN+LSTM with GRU ie. (Model - 6) gave better results compared to all the other models and also the model has less number of parameters compared to other models.

**Submitted by:**

* **Lalkrishna Goswami**
* **Rakesh Marathe**
* **Manuj Tyagi**