# **Gesture Recognition**

**Objective** of the project is to experiment with different CNN & RNN models to predict hand gestures and record the observation. Input provided for this assignment has video image frames which is categorized into 5 different classes.

1. Thumbs Up: Increase the volume
2. Thumbs Down: Lower the volume
3. Left Swipe: Fast Forward 10 seconds
4. Right Swipe: Fast Backward 10 seconds
5. Stop: Pause the movie

### **Generator Code**

* Tried Batch size of 60, 80 and 100 for this experiment. The same model when batch size was set at 100 performed more efficiently than other lower values.
* Image size was another parameter which was available to be tunable. Tried image size of 128\*128, 100\*100 ,80\*80 and 64\*64.
* Results of the model was not appreciable when image size was at 100\*100, 80\*80

Epoch 00030: saving model to model\_init\_2022-07-0816\_42\_42.124330/model-00030-1.19621-0.52187-1.72166-0.16000.h5

7/7 [==============================] - 96s 16s/step - loss: 1.1962 - categorical\_accuracy: 0.5219 - val\_loss: 1.7217 - val\_categorical\_accuracy: 0.1600 - lr: 1.0000e-05

Image Size: 80\*80

9/9 [==============================] - 90s 11s/step - loss: 1.0042 - categorical\_accuracy: 0.6139 - val\_loss: 1.5609 - val\_categorical\_accuracy: 0.3000 - lr: 1.0000e-05

**Image size : 64\* 64** 🡺 Best Result was obtained with this image size

9/9 [==============================] - ETA: 0s - loss: 0.1060 - categorical\_accuracy: 0.9608

Epoch 00028: saving model to model\_init\_2022-07-0819\_22\_18.066927/model-00028-0.10603-0.96078-0.88097-0.83000.h5

9/9 [==============================] - 90s 11s/step - loss: 0.1060 - categorical\_accuracy: 0.9608 - val\_loss: 0.8810 - val\_categorical\_accuracy: 0.8300 - lr: 5.0000e-04

### **Model**

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Result** | **Decision + Explanation** |
| **1** | **Conv3D** | **Train Accuracy: 99%**  **Validation Accuracy: 78%**  **Validation Loss: 0.79**  **Time Taken: 2958sec** | **Model accuracy was the best compared to other models**  **Increasing the number of trainable parameters did not yield better result** |
| **2** | **CNN + RNN (GRU)** | **Train Accuracy: 99%**  **Validation Accuracy: 63%**  **Validation Loss: 1.137**  **Time Taken: 2939sec** | **Model failed to respond when 2 layers of GRU added.** |
| **3** | **CNN + RNN (GRU)** | **Train Accuracy: 99%**  **Validation Accuracy: 58%**  **Validation Loss: 1.13** | **Increase the amount of trainable data, reduce the filter size**  **Model was clearly overfitting** |
| **4.** | **CNN + RNN (LSTM)** | **Train Accuracy:99%**  **Validation Accuracy: 58%**  **Validation Loss: 1.15**  **Time Taken: 2973sec** | **Results were no better than GRU model** |

### **Conclusion:**

With all the different combinations experimented find that CONV3D model performed much better when processing gesture recognition and classification problem.