Explanation steps for modeling.

* Data Generator object was created that could pass the data in batch without consuming much memory.
* Initially the batch size was selected as 512 as the sequence size for training data was 663.
* As we started reducing the batch size we could see the batch size of 64 was a perfect match it used maximum memory for the training an entire epochs.
* Preprocessing of images was done before training the data as the size was 360X360 and 120X160, all images were cropped that were uneven size into square (1:1 ratio). And it was resized to 80x80 dimension
* Multiple deep learning models were created based on 2 architectures ( Conv3D & Con2D+RNN)

Below are the results:

|  |  |  |  |
| --- | --- | --- | --- |
| **Experiment Number** | **Model** | **Result** | **Decision + Explanation** |
| 1 | Conv3D | Resource Exhausted Error | Used huge batch size of 512, 7layers, 3 epochs for training |
| 2 | Conv3D | Resource Exhausted Error | Used huge batch size of 256 7layers, 3 epochs for training |
| 3 | Conv3D | Resource Exhausted Error | Used huge batch size of 128 7layers, 3 epochs for training |
| 4 | Conv3D | Perfect fit for batch size and ran successfully | Used batch size of 64 7layers, 3 epochs for training |
| 5 | Conv3D | Train acc = 99.25, Val acc = 74 | Trained previous model more with 20 epochs to gain some more insights) |
| 6 | Conv2D + GRU | Train acc = 52.54, Val acc = 45 | Batch size of 64 8layers (Con2D) + 4layers(RNN), 10 epochs (added Batch Normalization & Dropout) |
| 7 | Conv2D + GRU | Train acc = 52.54, Val acc = 46 | Batch size of 64 9layers (Con2D) + 4layers(RNN), 10 epochs (removed Batch Normalization & Dropout to reduce Underfitting)) |

Conclusions:

* We are able to achieve a good fit model using Conv2D+RNN with best accuracy of training & validation reaching nearly 91.7% & 79%, with minimal trainable parameters.
* The possible reason for Conv2D+RNN performing better than Conv3D, as it gives more importance to the sequence of the frames.
* The trainable parameters can be further decreased by reducing the processed image size from 80x80 to 60x60 and also the training time could be reduced by using only 15 frames for 1 sequence instead of 30 frames.