Project Final Report

## Recognition of Stretching Body Through Video

Muhammad Hussain Anwaar

Graduate Student

Industrial and Systems Engineering

Texas A&M University

**Improvement Strategies:**

One of the areas where the model from last submission lagged was predicting in more diverse conditions, e.g. real time videos where there is low light or the action is being performed from a certain angle. The model had difficulty classifying the actions in the above-mentioned conditions. Also, I realized that previously submitted model was a bit under-fitted, as the validation and training accuracy were pretty close and were off by just 2%.

Last time as we experienced that the Ensemble architecture gave superior performance than using the simple architecture. So, the improvement strategies included enhancing the dataset by adding more diverse videos to the training and test datasets to make the model more robust.

To address the second issue of under-fitting, I thought of expanding on the idea of Model Ensemble using different inception modules. This time I decided on using Transfer Learning as well, to see if we can improve upon the previous test accuracy.

Thus, I tried two model architectures. One is a novel architecture that I designed, which is based on Model Ensemble of three different architectures one of which includes the Inception modules for better learning and the other includes deep convolutional network and the 3D convolutional neural network architectures.

The other architecture is a deep and expansive architecture which uses the concept of 2D and 3D convolution and ensemble them together in the end and then use a dense layer for classification purpose.

**Data Set Enhancement:**

Previously the dataset had 400 videos out of which 224 were for “Stretching” and about 176 were for the “No-Stretching” class. I felt the need to make the dataset more robust as was mentioned in the feedback videos regarding the project that videos in diverse conditions should be added to confuse the model.

Thus, I recorded videos on my own for Stretching and No-Stretching in diverse conditions ranging from Low Light, Shadowed Frame and video captured from different angles. I recorded about 50 new videos out of which 30 were for “Stretching” and about 20 were for “No-Stretching”.

Thus, now the updated dataset of 450 videos had a class distribution of:

1. Stretching: 254 Videos
2. No-Stretching: 196 Videos

Test Dataset was also created out of these 450 videos. About 69 videos were randomly selected from these 400 videos to be used as test dataset. The class proportion in test dataset is as follows:

1. Stretching: 39 Videos
2. No-Stretching: 30 Videos

**Model Architectures:**

**Model Ensemble Architecture:**

Model Ensembling is a powerful tool, where we can combine the two or more models together and where they would try to learn through the data in parallel and then combined at the output level. Motivation to join different models is that they are of different architectures and the hope is that they would learn different features that might help with the predictions.

For the Inception leg model, I used three types of Inception modules as they would be explained below, the basic inception and the inception and reduction module from Inception v4 model.

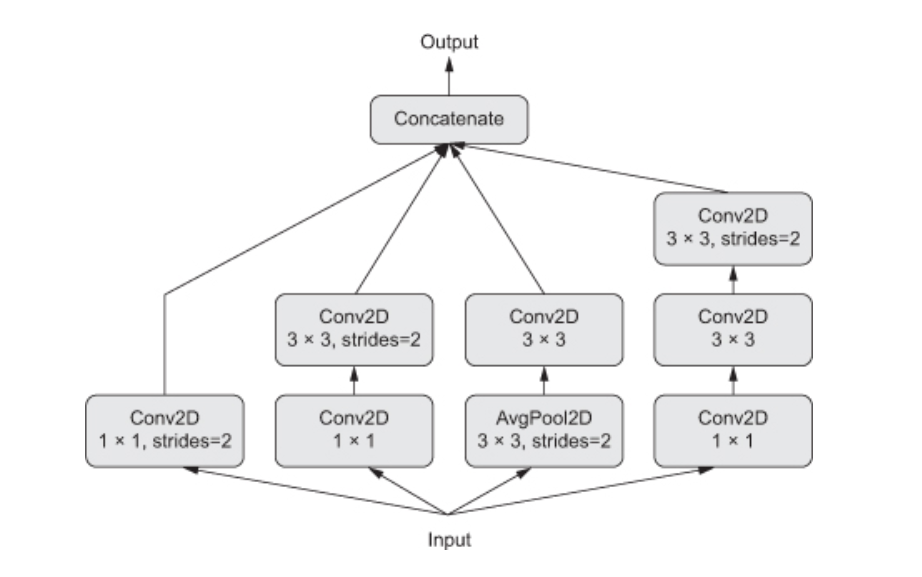


Figure 1: Basic Inception Module

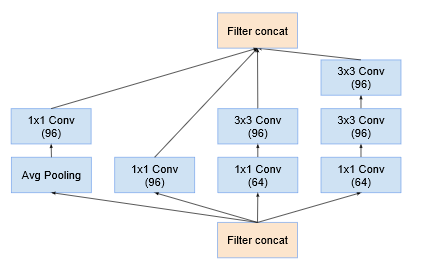


Figure : Inception Module A

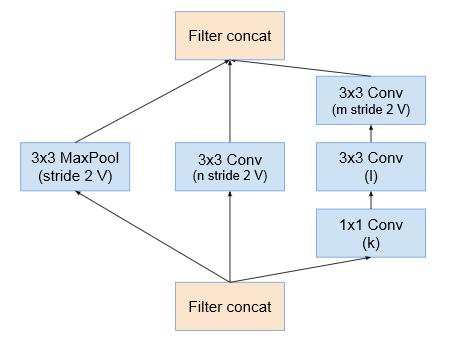


Figure : Reduction Module

Thus, I combined the three architectures, one is a deep 2D-CNN and LSTM architecture, while the other comprises of Time Distributed CNN with Inception Blocks embedded, this model contains inception block (simple), inception block A from Inception v4 and the reduction block from Inception-V4 as well. I had used the same kind of architecture previously, so what different in this architecture is that it is deeper and more diverse than the previous one.

Previously, the 2D-Conv leg of the model contained 3 blocks of the Convolutional Block which is explained below.

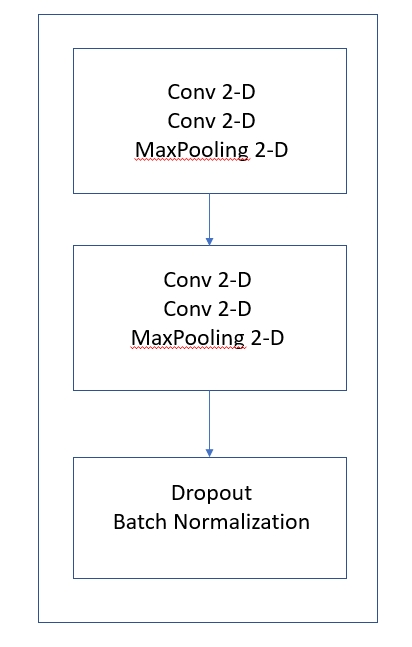


Figure 4: Base Convolutional Block

The new architecture would contain four of these blocks making it a deeper architecture. Whereas the First Leg of the ensemble model is designed using various inception blocks multiple times for a deeper and robust model architecture. The motivation for this architecture is being taken from the Inception-V4 and Ensemble architectures. Below is the figure defining the complete architecture.

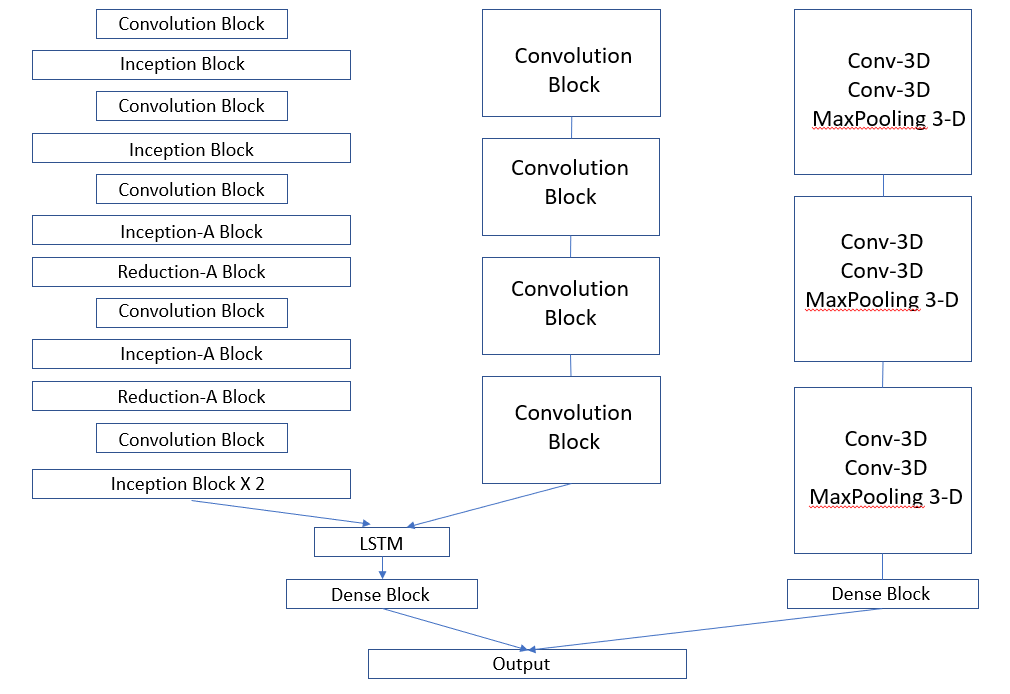


Figure 5: Ensemble Model Architecture

The above-mentioned model was trained on 30, 50 and 60 epochs, with batch size of 10, 15 and 20. I used “Adam” and “RMSprop” as the optimizers and the finest results were obtained from epochs=50 and batch size = 15. The test accuracy from this ensemble architecture is 90% which is an improvement of about 2% from the previously defined architecture. This seems to be a decent improvement, considering the dataset has remained the same. The figure below shows the model’s history when trained on the train dataset.

