Tennis Action Recognition Based on a Deep Learning Model

Sawyer Nelson
Computer Science Department
Advisor: Ming Ma

WINONA STATE UNIVERSITY
Computer Science Department
http://cs.winona.edu

Introduction

Action recognition in various sports is becoming increasingly more popular mostly due to its uses for analysis and training by coaches. Using action recognition specifically for tennis can be a very helpful tool to put together a collection of different precise strokes and actions that can be analyzed by coaches for training.

The aim of this research is to determine whether the addition of LSTM layers is beneficial to the Inception V3 CNN network + LSTM deep learning action recognition approach. Action recognition is a technique in which we train a model using our dataset of different tennis actions and evaluate them based off the accuracy. Theoretically, by adding LSTM layers we should see an increase in the evaluations due to the LSTM layers decreasing the sequence prediction problems.

Method

The deep learning model used was pretrained with ImageNet which uses over a million images. All parameters used on this model will stay the same while adding additional LSTM layers to have the evaluations be fair and only focused on the difference the layers make. We extract features using the Inception V3 CNN network, these features are then given to the LSTM layers whose SoftMax outputs are averaged across frames to obtain a final stroke prediction.

The dataset used was organized into 6 different tennis strokes. The strokes that were used to be recognized were forehand, backhand, forehand volley, backhand volley, service and smash. The dataset is composed of a total of 1980 RGB videos at a size of 640x480.

For each LSTM layer that was added to our model we compared the outputs by looking at the F1-Score, precision and recall. By looking at the results of each layer that was added, we can determine if adding layers increase the outputs.

Table 1. Evaluation Results

F1-Scores and Evaluation Results							
Number of Layers	F1-Score Accuracy	F1-Score Macro Avg	F1-Score Weighted Avg	Precision Macro Avg	Precision Weighted Avg	Recall Macro Avg	Recall Weighted Avg
1	83	76	82	86	85	74	83
2	84	79	83	86	85	75	84
3	84	77	83	85	84	74	84

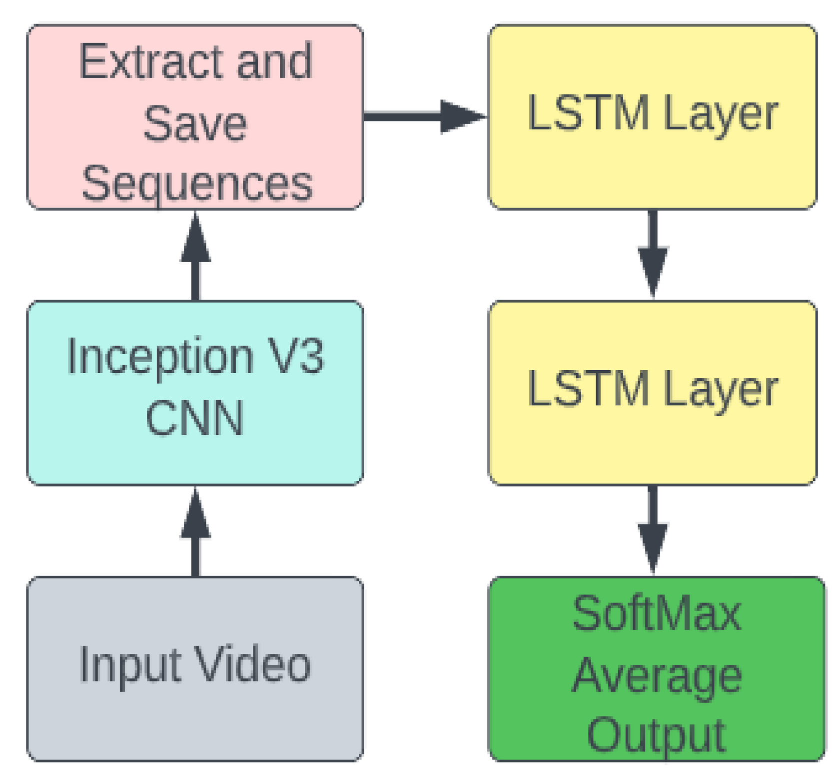

Fig1. The deep learning model using Inception V3 CNN + two LSTM Layers.

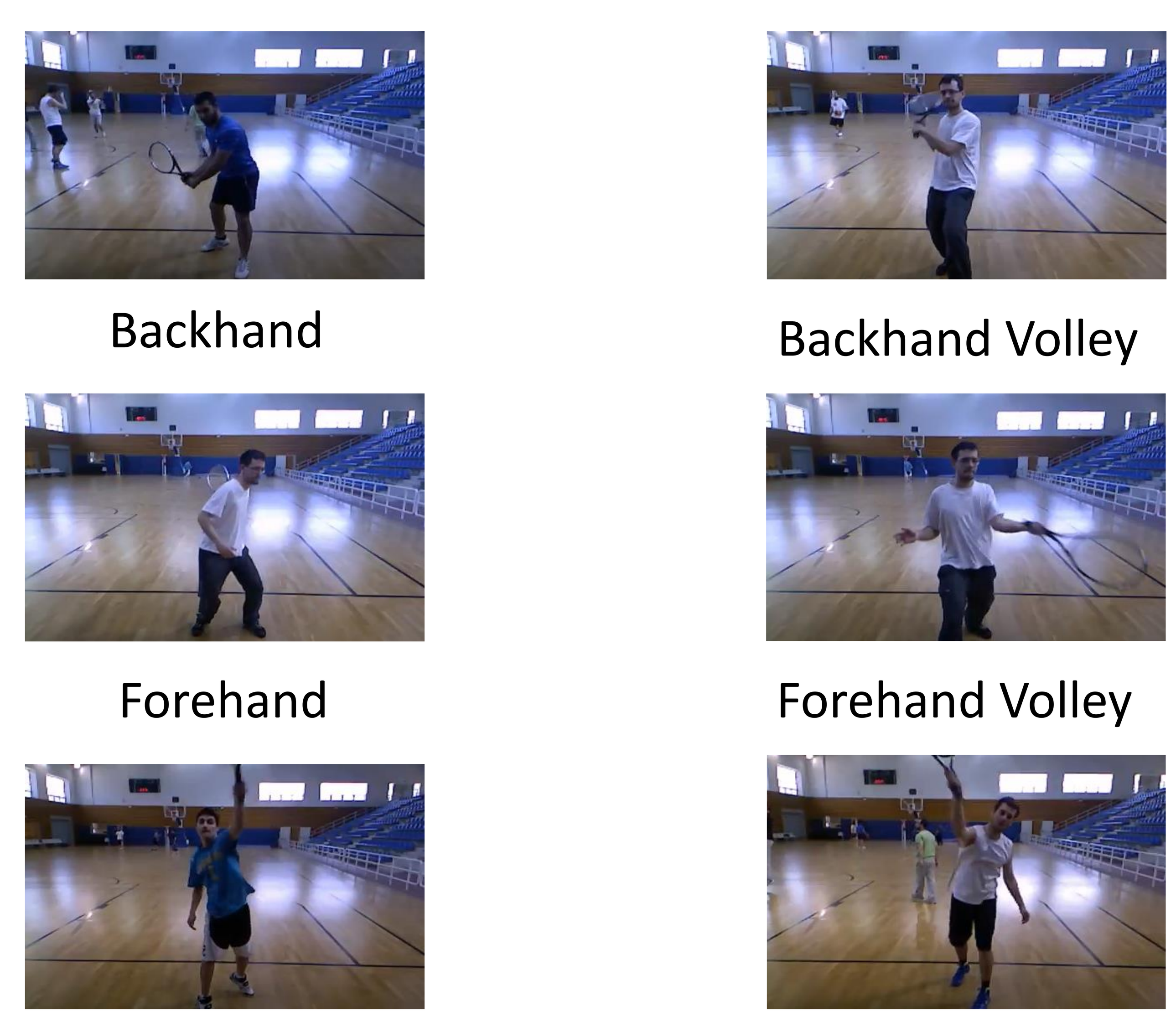

Fig2. Illustration of six different tennis actions in the dataset.

Results

First, we can see that by adding a second LSTM layer to the model already containing a single LSTM layer, the F1-Score, precision and recall had improved. We kept the parameters the same for training when adding the additional LSTM layers. We took this further and looked at the results of adding two LSTM layers to the model already containing a single LSTM layer as well. There was not much of an increase in training time when additional layers were added, each training took approximately 2.5 hours.

The table is a visual representation of the F1-Score from each layer that was added. We can see an increase in the results when the second LSTM layers is added, we can also see that with a third LSTM layer that we achieved relatively the same higher results than with a single LSTM layer.

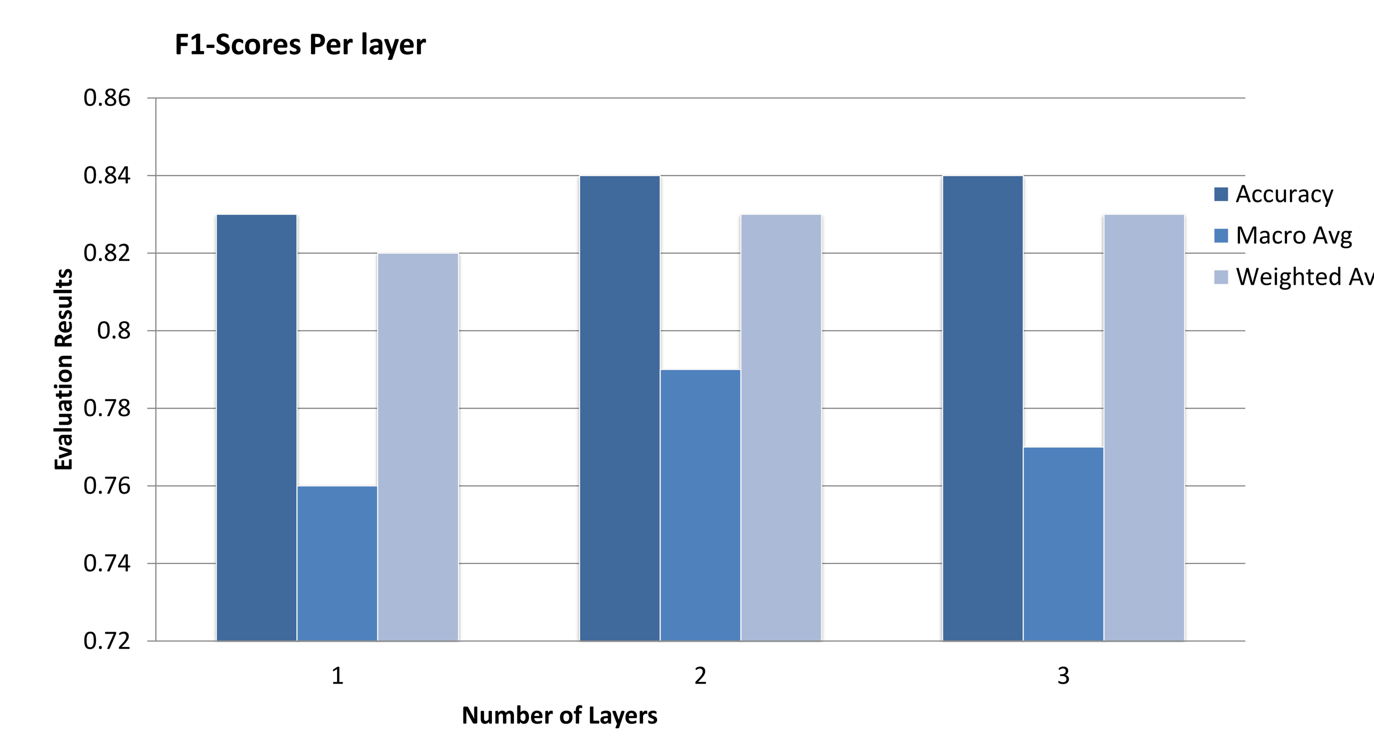

Fig3. Results in terms of F1-score for models using different number of LSTM layers.

Conclusion

The addition of extra LSTM layers had a positive effect on the deep learning model used in this experiment. Further study could be directed towards implementing additional layers in other deep learning models. In addition, it would be beneficial to know what parameters result in the best evaluation.

Winona State University Foundation