Identifying Player Actions in Tennis from Visual Data

Harsh Malara and Shivam Utreja

College of Information and Computer Science, University of Massachusetts, Amherst.

Motivation

Sports is a field with a **huge abundance of high quality visual data.** This data holds immense potential to be utilized with the state of the art techniques from computer vision, for a variety of tasks:-

- **Players and coaches** can immensely benefit from action-labelled video data to improve their performance and **cut down time on footage analysis.**
- People undergoing **physical therapy** can use action quality assessment feedback to **speed up the recovery process.**
- **Outcomes of games can be predicted**, by combining the video analysis of the current game, along with the past statistics.

Therefore our project aims to first, replicate the results in [3], and then experiment with the architecture of the model in order to improve results.

THETIS Dataset

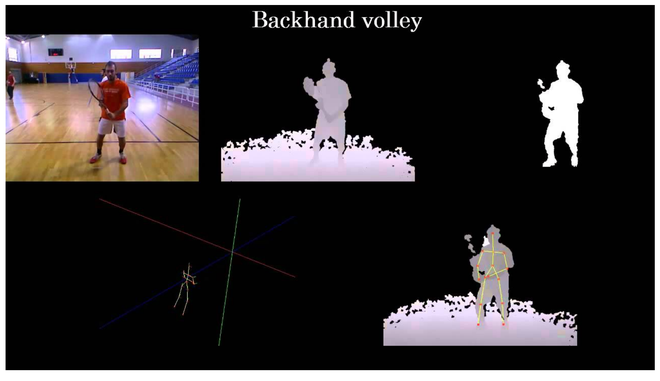


Fig. 1: THree Dimensional Tennis Shots Dataset.

- Contains monocular RGB videos, depth, skeleton 2D and 3D and silhouettes.
- 1980 monocular RGB videos of 12 tennis actions (fine-grained) performed three times by 55 different players (31 amateurs and 24 experienced).
- We use the RGB videos exclusively to keep the model usable with real-world data, and expandable to other sport videos.

Methodology

The task at hand is performed using the following key steps:

- Using a pre-trained CNN (Alexnet/Inception/LeNet) to extract features for every frame of every video.
- Training a many-to-one recurrent architecture using the features extracted above, and the labels from the THETIS dataset in order to learn action detection from video. The architecture used in [1] is shown below.

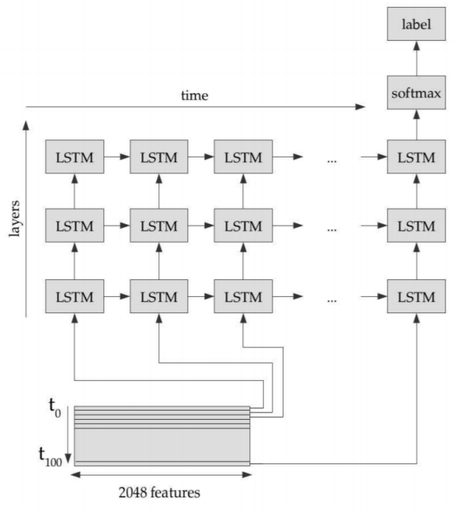


Fig. 2: Architecture used by Mora et al.[3]

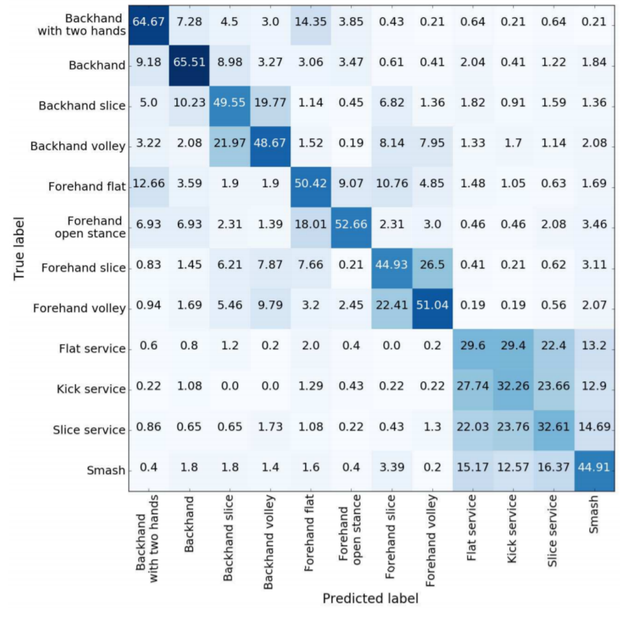
Experiments

The properties of the dataset and the network architectures used, proved several experimentation possibilities:

- Performing general stroke classification as opposed to fine-grained.
- Trying out different pretrained CNNs to extract frame features.
- Seeing the effects of changing LSTM depth & size of hidden layer.
- Seeing effect of player expertise on network performance.

Results & Conclusions

We are currently in the process of performing hyper-parameter tuning over the LSTM architecture of **depth 3** and with **hidden/cell-state dimensionality 90**.
*The results presented here are from [3], for completeness.



Acknowledgements

We thank Professor Subhransu Maji and the TAs (Aruni RoyChowdhury and Archan Ray) of COMPSCI670 for this project opportunity, and for guiding us throughout the course.

References

[1] Sofia Gourgari et al. "Thetis: Three dimensional tennis shots a human action dataset". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2013, pp. 676–681.

[2] Sepp Hochreiter and Jürgen Schmidhuber. "Long short-term memory". In: *Neural computation* 9.8 (1997), pp. 1735–1780.

[3] Silvia Vinyes Mora and William J Knottenbelt. "Deep learning for domain-