

# 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

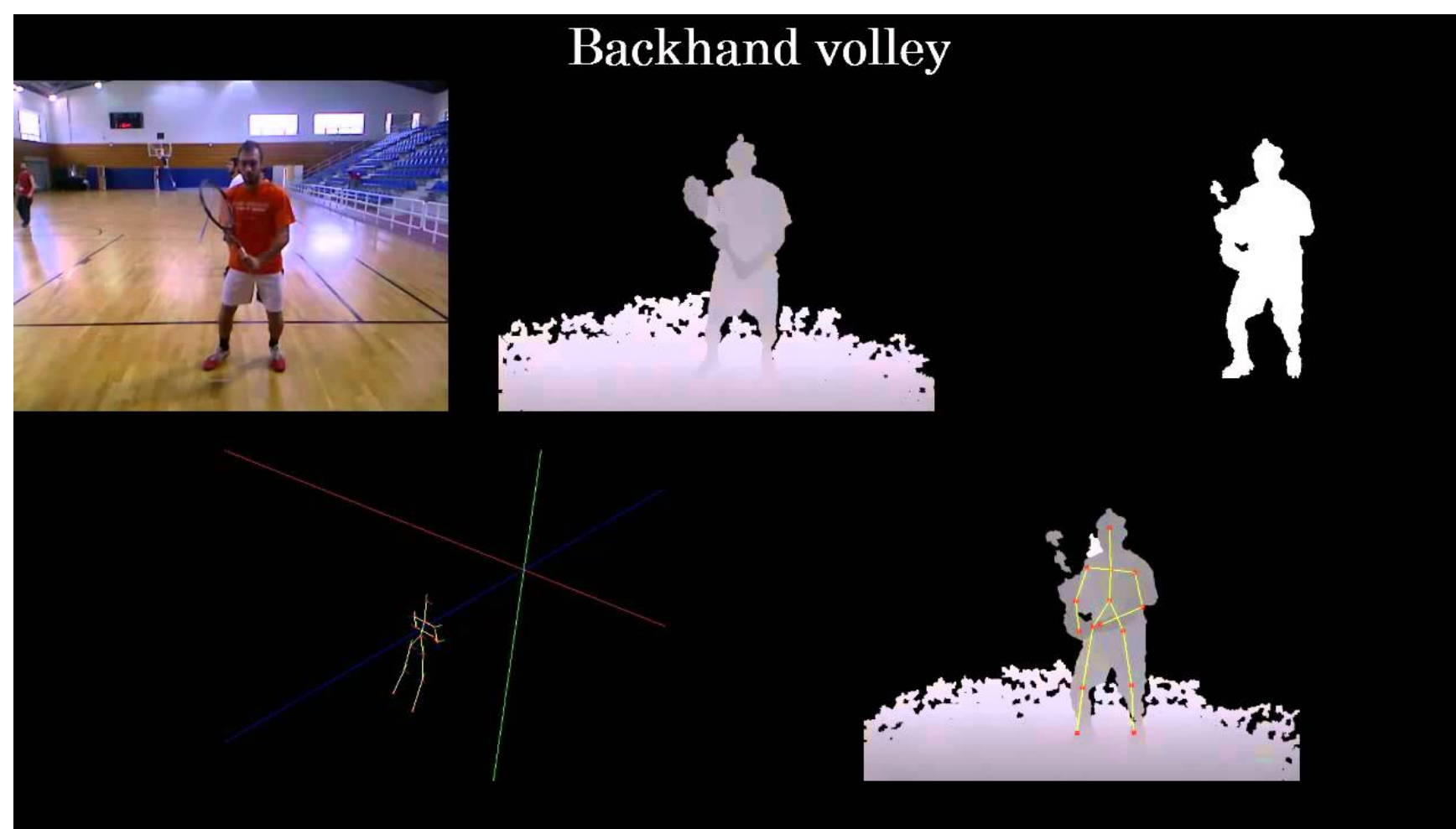


Fig. 1: THree Dimensional Tennis Shots Dataset.

- 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.

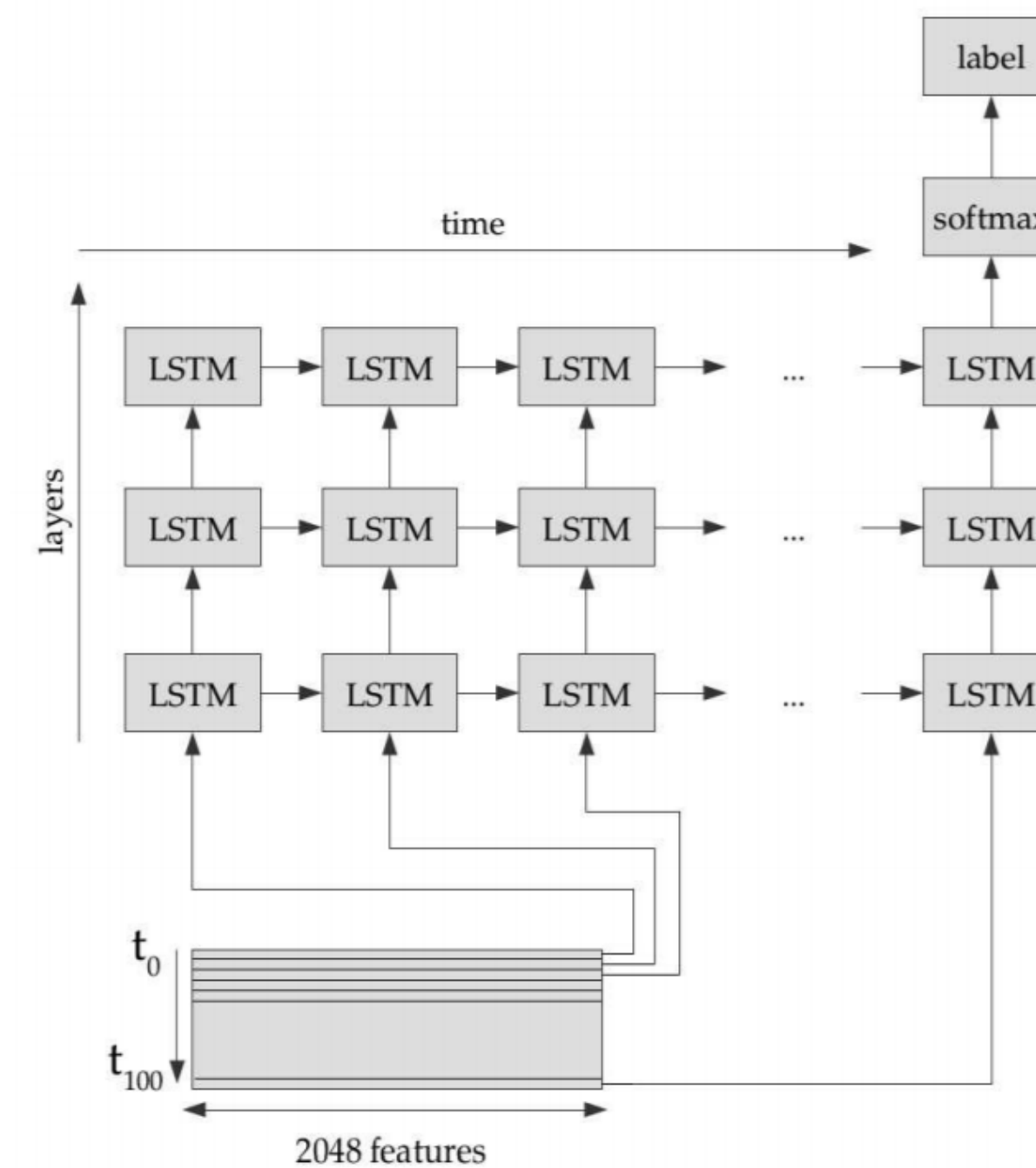


Fig. 2: Architecture used by Mora et al.[3]

## 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.

	Backhand with two hands	Backhand	Backhand slice	Backhand volley	Forehand flat	Forehand open stance	Forehand slice	Forehand volley	Flat service	Kick service	Slice service	Smash
Backhand with two hands	64.67	7.28	4.5	3.0	14.35	3.85	0.43	0.21	0.64	0.21	0.64	0.21
Backhand	9.18	65.51	8.98	3.27	3.06	3.47	0.61	0.41	2.04	0.41	1.22	1.84
Backhand slice	5.0	10.23	49.55	19.77	1.14	0.45	6.82	1.36	1.82	0.91	1.59	1.36
Backhand volley	3.22	2.08	21.97	48.67	1.52	0.19	8.14	7.95	1.33	1.7	1.14	2.08
Forehand flat	12.66	3.59	1.9	1.9	50.42	9.07	10.76	4.85	1.48	1.05	0.63	1.69
Forehand open stance	6.93	6.93	2.31	1.39	18.01	52.66	2.31	3.0	0.46	0.46	2.08	3.46
Forehand slice	0.83	1.45	6.21	7.87	7.66	0.21	44.93	26.5	0.41	0.21	0.62	3.11
Forehand volley	0.94	1.69	5.46	9.79	3.2	2.45	22.41	51.04	0.19	0.19	0.56	2.07
Flat service	0.6	0.8	1.2	0.2	2.0	0.4	0.0	0.2	29.6	29.4	22.4	13.2
Kick service	0.22	1.08	0.0	0.0	1.29	0.43	0.22	0.22	27.74	32.26	23.66	12.9
Slice service	0.86	0.65	0.65	1.73	1.08	0.22	0.43	1.3	22.03	23.76	32.61	14.69
Smash	0.4	1.8	1.8	1.4	1.6	0.4	3.39	0.2	15.17	12.57	16.37	44.91
	Backhand with two hands	Backhand	Backhand slice	Backhand volley	Forehand flat	Forehand open stance	Forehand slice	Forehand volley	Flat service	Kick service	Slice service	Smash

## Acknowledgements

We thank Professor Subhansu 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 Gourhari 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-specific action recognition". In: *2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 2018, pp. 1735–1744.