Action Recognition in Tennis Using Deep Neural Networks

Vincent Chow and Ohi Dibua

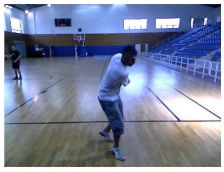
Introduction

Goal: Apply **deep learning** to **classify videos** of players performing **tennis strokes.**

Example strokes: forehand, backhand, service

Dataset

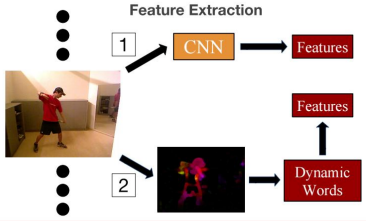
- **55 players** of varying skill levels
- **1980** RGB **videos**, sized 640 x 480
- Up to **12 classes** of tennis strokes



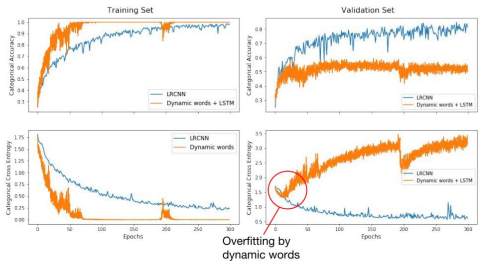
Architectures

	Architectures
1	Inception V3 CNN + LSTM (LRCNN)
2	Optical flow/dynamic words + LSTM

*Note: Dynamic words uses many-to-one LSTM



Results



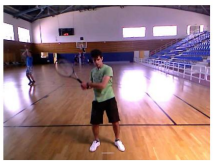
Overfitting by dynamic words

Hyperparameters			
Learning Rate	**Batch Size**	**LSTM Hidden Units**	**Dropout Rate**
1e-3	128	128	0.3

Accuracies			
Architecture	**Training**	**Validation**	**Test**
LRCNN	98.7%	77.9%	82.3%
Dynamics Words + LSTM	100%	53.2%	53.2%

Discussion

- **Small dataset** fundamentally **limits achievable accuracies**
- CNN captures **useful information lost** in dynamic words approach
- Example video **misclassified** by both architectures on the right



Misclassified video:
1. Player performs stroke incorrectly
2. Inconsistent background

Suggests better dataset will improve performance

Future Work

- Try feeding **pose keypoints** to LSTM
- **Backpropagate** through **CNN** layers
- Gather **better data** and **more data**

Citations:

[1] Sofia Gourgari, Georgios Goudelis, Konstantinos Karpouzis, and Stefanos Kollias. Thetis: Three dimensional tennis shots a human action dataset. International workshop on Behavior Analysis in Games and modern Sensing devices, 2013.

[2] Jeff Donahue, Lisa Anne Hendricks, Marcus Rohrbach, Subhashini Venugopalan, Sergio Guadarrama, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. arXiv preprint arXiv:1411.4389, 2016.

[3] Pablo Negri Claudio Delrieux Jonathan Vainstein1, Jose F. Manera and Ana Maguitman. Modeling video activity with dynamic phrases and its application to action recognition in tennis videos. CIARP 2014: Progress in Pattern Recognition, Image Analysis, Computer Vision, and Applications, 2014.