Ciências ULisboa Faculdade de Ciências da Universidade de Lisboa

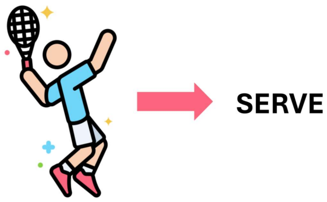
M.Sc. *Data Science*
Deep Learning class 2024/2025

Action Recognition in Tennis

Diana Santos (64478) , Leonor Fandinga (64481) and Sofia Rocha (65111)

Objectives

The main goal of this project is to establish the model ST-GCN for the analysis of video data to accurately recognize and classify various types of tennis movements (such as forehand, backhand, serve, slice, etc). The study also aims to see the model's ability to identify these movements regardless of whether or not the ball is in play.



Problem

Why this topic? Because athletes, including ourselves, constantly strive to improve. We're developing a model to predict player movement, offering a data-driven approach to enhance technique, optimize training, and gain a competitive advantage. This isn't just for tennis – it's a game-changer for all sports.

Datasets

For this project, we used NTU RGB-D[1,2] to pre-train the model on general human actions and THETIS[3] to fine-tune and evaluate tennis-specific movements.

Table 1 - THETIS and NTU RGB+D datasets used for training and testing.

DATASET	MODALITY	CLASSES	Nº OF SAMPLES	YEAR	Description
THETIS	DEPH	12	1980	2013	Tennis shots from 55 players (31 amateurs and 24 advanced players)
	MASK	12	1980	2013	
	RGB	12	1980	2013	
	SKELET2D	12	1217	2013	
	SKELET3D	12	1217	2013	
NTU RGB+D	RGB + D + IR +3D JOINTS	60	56880	2016	Daily actions, mutual actions, and medical conditions

Methodology

For training, we explored two scenarios: full network training and single-layer fine-tuning.

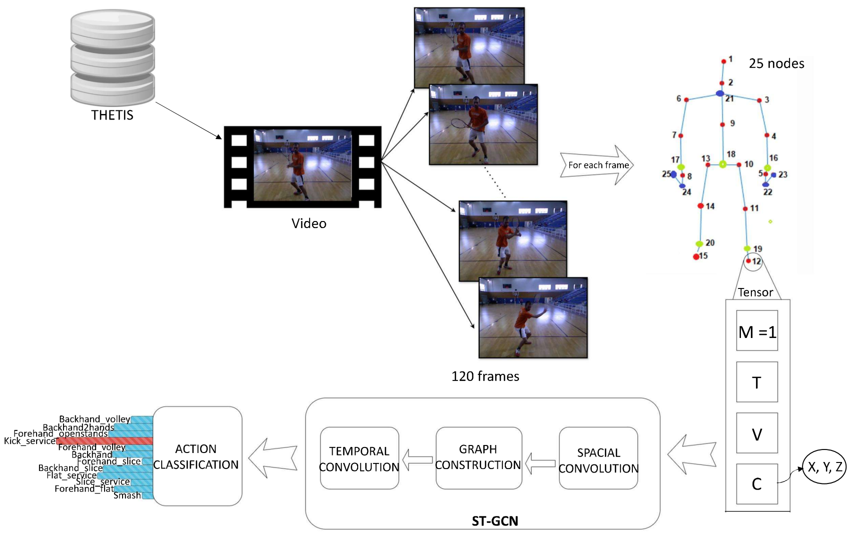


Figure 1 - Methodology overview; In the tensor M = Number of people, T = Number of frames, V = Number of joints, C = Number of coordinates

Results

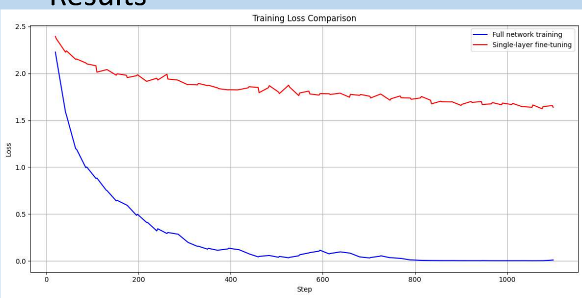


Figure 2 - Loss curves for training in full network training (Blue) and for in single-layer fine-tuning (Red).

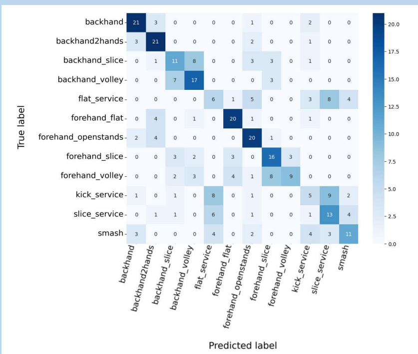


Figure 3 – Confusion Matrix with full network training.

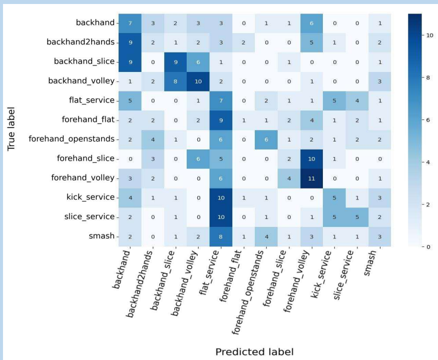


Figure 4 – Confusion Matrix with single-layer fine-tuning.

Conclusions

- Better performance was obtained for the full network training method.
- The single-layer method couldn't predict at all the forehand volley action.
- Future work includes applying the most accurate method to our own dataset and evaluating the effect of incorporating the ball.

References

[1] Mora, S. V., & Knottenbelt, W. J. (2017). Deep learning for domain-specific action recognition in tennis. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 170–178). IEEE. https://doi.org/10.1109/CVPRW.2017.27
[2] Shahroudy, A., Liu, J., Ng, T.-T., & Wang, G. (2016). NTU RGB+D: A large scale dataset for 3D human activity analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 1010–1019). IEEE. https://doi.org/10.1109/CVPR.2016.115
[3] Gourgari, S., Goudelis, G., Karpouzis, K., & Kollias, S. (2013). THETIS: Three dimensional tennis shots, a human action dataset. In *2013 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 676–681). IEEE. https://doi.org/10.1109/CVPRW.2013.102

Acknowledgements: This poster session is kindly sponsored by:

NTT DATA