### UNIVERSITATEA DIN BUCUREȘTI

**FACULTATEA DE MATEMATICĂ ȘI INFORMATICĂ**

**SPECIALIZAREA INFORMATICĂ**

**Lucrare de licență**

# TITLUL LUCRĂRII DE LICENȚĂ

### Absolvent Numele studentului

**Coordonator științific**

**Titlul și numele profesorului coordonatorului**

**București, iunie 2021**

**Rezumat**

Statisticile miscarilor dintr-un meci de tenis pot oferi un mare avantaj in pregatirea unui jucator. O cunoastere detaliata a unui oponent poate fi benefica in pregatirea unui jucator, oferind un avantaj in intelegerea punctelor forte si a celor slabe ale unui adversar, ceea ce aduce un plus pentru desfasurarea antrenamentelor.

Pentru aceasta vom crea o aplicatie ce va oferi predictii asupra miscarilor unui jucator dintr-un videoclip folosind limbajul python si biblioteca PyTorch pentru a implementa modelele.

**Abstract**

The statistics of the movements in a tennis match can offer a great advantage in the preparation of a player. A detailes knowledge of an opponent can be beneficial in the preparation of a player, offering an advantage in understanding the strengths and weakness of opponent, which brings a plus for training.

For this we will create an application that will provide predictions on the movements of a player from a video using python and the PyTorch library to implement the models.

Contents

[UNIVERSITATEA DIN BUCUREȘTI 1](#_Toc132668605)

[TITLUL LUCRĂRII DE LICENȚĂ 1](#_Toc132668606)

[Absolvent Numele studentului 1](#_Toc132668607)

[1. Introducere 4](#_Toc132668608)

[1.1. Motivatie 4](#_Toc132668609)

[1.2 Structura lucrarii 5](#_Toc132668610)

[1.2.1: Detectia jucatorului 5](#_Toc132668611)

[1.2.2: Predictia actiunii 5](#_Toc132668612)

[2. Concepte teoretice 6](#_Toc132668613)

[2.1: LSTM 6](#_Toc132668614)

[2.2 YOLO 7](#_Toc132668615)

[2.3 YOLO pose 7](#_Toc132668616)

[3. Bibliografie 8](#_Toc132668617)

# Introducere

### 1.1. Motivatie

Analiza video a sporturilor este o solutie pentru a extrage diferite informatii care pot fi folosite de profesionisti cat si de telespectatori pentru a avea o privire de ansamblu analitica a ceea ce se intampla pe parcursul jocului.

O astfel de tehnologie este sistemul „Hawk-Eye Live” care poate semnala daca mingea a fost data afara suprafetei de joc. Insa instalarea acestui sistem este dificila si costisitoare, conform [Espn, 2023] costul pentru instalarea acestui sistem pe un teren este intre $60.000 si $70.000, facand folosirea acesteia imposibila pentru antrenamente sau pentru turneele mici.

Scopul acestei lucrari este de a oferi o solutie accesibila din punct de vedere computational, fiind usor de rulat pe un GPU.

Lucrarea propune o modalitate de a vedea ce miscari precise efectueaza un jucator de tenis in cadrul unei filmari, pentru a putea fi analizata. Prin intermediul acestei abordari se poate analiza atent jocul jucatorului, antrenorii putand imbunatati strategiile, tehnica de joc, precum si pregatirea fizica inaintea meciurilor.

Abordarea aleasa va folosi limbajul pyhon pentru implementare si tehnici inovatoare pentru a detecta jucatorul si coordonatele pozitiilor membrelor. Astfel pentru detectia jucatorului vom folosi YOLOv7 care conform [Wang et al., 2022] intrece toti detectorii in privinta vitezei si a acuratetei, oferind o acuratete de 56.8% pe setul de date MS COCO. Pentru obtinerea coordonatelor membrelor vom folosi un detector de pose estimation antrenat pe YOLOv7 avand la baza tehnica prezentata in [Maji et al., 2022] care ofera o acuratete buna, de 68.5 pe setul de date COCO 2017 si are o performanta computationala foarte buna.

### 1.2 Structura lucrarii

* **Detectia jucatorului**: Selectarea jucatorului care are dimensinile cele mai mari din fiecare cadru, extragerea coordonatelor membrelor si normalizarea datelor pentru a putea fi prelucrate.
* **Predictia actiunii**: Antrenarea, proiectarea si inferenta intr-un videoclip bazandu-se pe coordonatele membrelor pentru a oferi predictia miscarii efectuate, pe un segment de video, avand numarul de date prestabilit ca hiperparametru.

### 1.2.1: Detectia jucatorului

Pentru a detecta jucatorul vom folosi modelul YOLOv7 preantrenat pentru detectia pozitiilor membrelor, pentru a detecta toate persoanele dintr-o imagine. Dupa aceasta detectie vom selecta numai persoana care are dimensiunea cea mai mare din cadru (daca ar fi feed-ul unei camere din transmisia TV, acest model va selecta numai jucatorul din partea de jos). Pentru fiecare membru din „cutia” dominanta ca si dimensiuni, coordonatele acestuia se vor normaliza in intervalul [0, 1]. Aceasta operatiune se va repeta pentru fiecare cadru dintr-o filmare

### 1.2.2: Predictia actiunii

Vom lua datele de la punctul anterior, ne vom folosi de un set de date adnotat cu videoclipuri cu lovituri din tenis pentru a antrena un model care sa poata da o predictie asupra miscarii efectuate de catre un jucator, oferind ca iesire una din 12 miscari, elementare jocului de tenis.

# Concepte teoretice

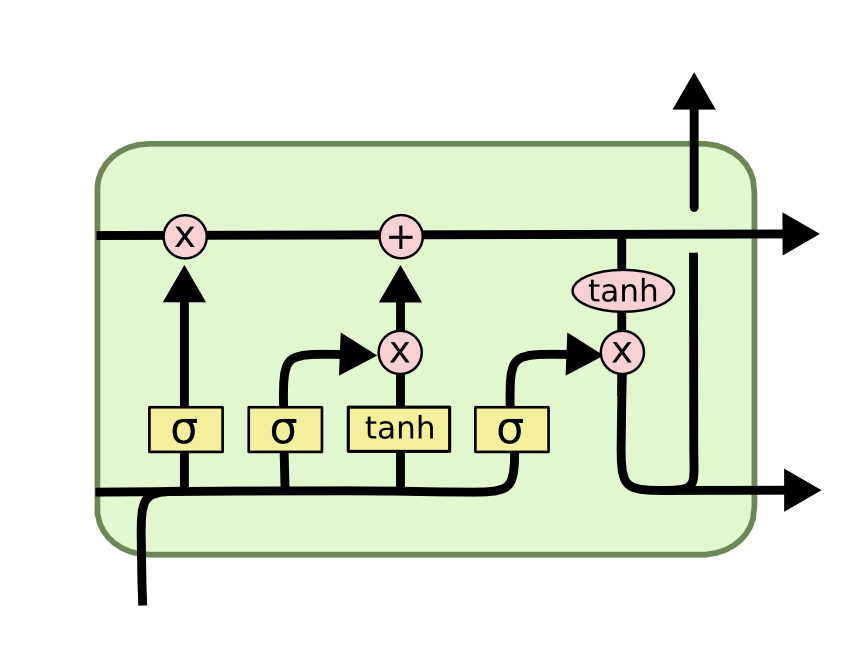
### 2.1: LSTM

Retelele de tip LSTM (Long short term memory) vin ca o evolutie a retelelor RNN (Recurrent neural networks), care conform [Staudemeyer et al., 2019.] nu pot depasi mai mult de 5-10 pasi de timp intrucat gradientii se pot micsora foarte tare (dispar) sau pot avea valori foarte mari (explodeaza). In cazul in care gradientii explodeaza, valoarea „w” oscileaza, iar in cazul in care gradientii dispar, antrenarea dureaza foarte mult timp sau chiar nu functioneaza.

Solutia acestei probleme este folosirea LSTM care poate avea conform [Staudemeyer et al., 2019.] si mai mult de 1000 de pasi de timp. Retelele LSTM exceleaza pe sarcinile care necesita ca un set de date secvential sa fie tinut minte pentru o perioada lunga de timp. Aceasta propietate este atribuita blocurilor de memorie, care pot accesa intrarea si produc o iesire, prevenind astfel ca informatiile irelevante sa intre sau sa iasa din acest bloc. De asemenea aceste blocuri mai au si o poarta de „uitare” astfel incat daca o caracteristica devine irelevanta, aceasta nu mai este luata in calcul [Staudemeyer et al., 2019.].

Aceste retele se pot suprapune pentru a forma o retea deep LSTM, astfel iesirea unei astfel de retele, reprezinta intrarea urmatoarei. Aceasta abordare a adus o imbunatatire dramatica fata de o retea cu un singur strat, in domeniul transcrierii mesajelor audio in text [Graves et al., 2013].

Pentru a accelera antrenarea putem folosi „Gradient clipping” care mentine gradientii intr-un interval stabilit ca si hiperparametru. De exemplu daca am setat valoarea de 0.25 ca hiperparametru si gradientul nostru are valoarea de 0.7, acesta va fi setat la valoarea de 0.25. Conform [Zhang et al., 2019] folosind aceasta tehnica, au testat pe o sarcina de prelucrarea limbajului natural, folosind modelul AWD-LSTM, au putut observa o crestere a vitezei cu care modelul este antrenat.



Imagine descriptiva a unui bloc LSTM [colah.github.io, 2023]

### 2.2 YOLO

YOLO (You only look once) este o retea capabila sa detecteze locatiile obiectelor in timp real din imagini. In aceasta lucrare vom folosi YOLOv7 care imbunatateste acuratetea si viteza. Pentru antreneare acesta foloseste setul de date MS COCO, antrenarea facandu-se de la 0, nefolosindu-se alte seturi de date sau retele preantrenate. Comparativ cu alte metode, de exemplu YOLOv4, acest model are cu 75% mai putini parametrii si aduce o imbunatatire a preciziei medii de 1.5%. Comparativ cu modelul state-of-the-art YOLOR-CSP, autorii ne prezinta ca modelul are cu YOLOv7 are cu 43% mai putini parametrii si o precizie medie cu 0.4% mai mare [Wang et al., 2022].

### 2.3 YOLO pose

Sarcina acestui model este de a detecta coordonatele articulatiilor tuturor persoanelor dintr-p imagine. Pentru antrenarea acestui model, autorii au folosit setul de date COCO Keypoint Detection, unde au augumentat datele folosind redimensionare, translatare, inversare, augumentari mozaic si diferite auggumentari ale culorilor. Pe setul de date COCO val2017 precizia medie este de 68.5% iar cel mai bun model este HigherHRNet cu 70.5%. Conform autorilor la rezolutii mai mici, YOLOv5s6-pose are o performanta mult mai buna decat alte modele de complexitate mica, de exemplu EfficientHRNet, pe setul de date COCO val2017 [Maji et al., 2022].

# Bibliografie

[Espn, 2023] URL: <https://www.espn.com/tennis/story/_/id/30877297/hawk-eye-live-gains-more-support-australian-open> (accesat la data 17.04.2023)

[Wang et al., 2022] Wang, Chien-Yao, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors." *arXiv preprint arXiv:2207.02696* (2022).

[Maji et al., 2022] Maji, Debapriya, Soyeb Nagori, Manu Mathew, and Deepak Poddar. "YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2637-2646. 2022.

[Staudemeyer et al., 2019.] Staudemeyer, Ralf C., and Eric Rothstein Morris. "Understanding LSTM--a tutorial into long short-term memory recurrent neural networks." *arXiv preprint arXiv:1909.09586* (2019).

[Graves et al., 2013] Graves, Alex, Abdel-rahman Mohamed, and Geoffrey Hinton. "Speech recognition with deep recurrent neural networks." In *2013 IEEE international conference on acoustics, speech and signal processing*, pp. 6645-6649. Ieee, 2013.

[Zhang et al., 2019] Zhang, Jingzhao, Tianxing He, Suvrit Sra, and Ali Jadbabaie. "Why gradient clipping accelerates training: A theoretical justification for adaptivity." *arXiv preprint arXiv:1905.11881* (2019).

[colah.github.io, 2023] URL: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/> (accesat la data 17.04.2023)