

# Google Football Research

## Solutie abordata

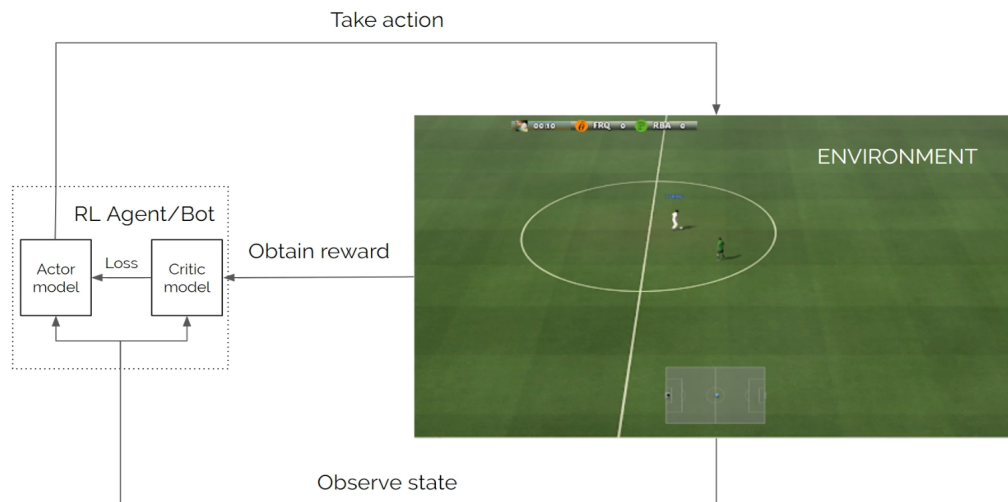
Am folosit Proximal Policy Optimization si metoda Actor-Critic.  
Se extrag feature-urile din imagini folosind MobileNetV2  
si creez o retea fully conected (tensorflow.Dense) cu 1024 de neuroni pentru a obtine probabilitatile corespunzatoare actiunilor.  
Pentru varianta de prelucrare raw data(simple115) nu mai este nevoie de extragerea feature-urilor.

Pentru a depista cea mai buna abordare long term, folosim Generalized Advantage Estimate algorithm. Urmam pasii in sens invers si atribuim rewardul corespunzator fiecarei stari.

$$\delta = r_t + \gamma \cdot V(s_{t+1}) \cdot m_t - V(s_t)$$

La crearea unei politici noi se calculeaza diferentele fata de cea veche, iar daca acestea sunt prea mari ea este refuzata (clip). Astfel, actorul nu poate avea schimbari sporadice, lipsite de sens.

Actorul alege actiunea cea mai potrivita starii curente, iar criticul analizeaza starea ce rezulta din aceasta actiune, triminand un feedback (ex: reward pozitiv daca se da un gol).



## Rezultate

- Image based
- `clipping_val = 0.3`
- `critic_discount = 0.2`
- `entropy_beta = 0.001`
- `gamma = 0.99`
- `lmbda = 0.95`
- `11_vs_11_hard_stochastic`

Discountul fiind mic agentul se focuseaza in principal pe pase. Funtia de reward oferita de enviroment este foarte restrictiva, oferinda +/- 1 pentru gol sau cate un 0.3 pentru 3 checkpointuri. Astfel, marcarea unui gol devine un obiectiv secundar, agentul preferand sa protejeze mingea si sa joace cat mai mult in jumatatea adversarilor de teren pentru a luat rewardul oferit de checkpointuri.

Observatie: De foarte multe ori agentul alege sa stea sau sa iasa cu mingea din teren, deoarece nu exista niciun fel de depunctare. Practic, strategia ce se deduce de cele mai multe ori este de a trage de timp, nu de a marca. Daca jucatorii ataca poarta sunt sanse mari sa piarda mingea sau sa fie interceptata de portar, urmand sa fie atacati.

- Image based
  - `clipping_val = 0.3`
  - `critic_discount = 1.0`
  - `entropy_beta = 0.001`
  - `gamma = 0.99`
  - `lmbda = 0.95`
  - `11_vs_11_hard_stochastic`

La inceput era foarte agresiv, singurul obiectiv fiind marcarea. Nu prea existau pase, dar intrari violente la minge si suturi largi pe poarta. In final, dupa mult timp de antrenare, a ajuns la o concluzie pe care nu pot exact sa o inteleg. Urmareste autogolul, dar intr-un ritm lent.

- Data based
  - `clipping_val = 0.115`
  - `critic_discount = 0.997`
  - `entropy_beta = 0.00155`
  - `gamma = 0.5`
  - `lmbda = 0.95`

Cel mai interesant agent. Strategia aleasa este de a faulta echipa adversa. Adversarii nu pot marca, fiind constand doborati, iar checkpointurile sunt din cand in cand atinse, obtinandu-se un reward mic.

Cel mai bun rezultat:

- Data input

- `clipping_val = 0.154`
- `critic_discount = 0.915`
- `entropy_beta = 0.0015`
- `gamma = 0.995`
- `lmbda = 0.95`

Drible foarte bune, pase doar spre poarta lor, suturi pe poarta doar ca de la o distanta prea mare.

