Paper Reading

2022.09.22

Guan Yunyi

Real-time Active Vision for a Humanoid Soccer Robot Using Deep Reinforcement Learning

Soheil Khatibi, Meisam Teimouri , Mahdi Rezaei

Key words: Deep Reinforcement Learning, Active Vision, Deep Q-Network, Humanoid Robot, RoboCup

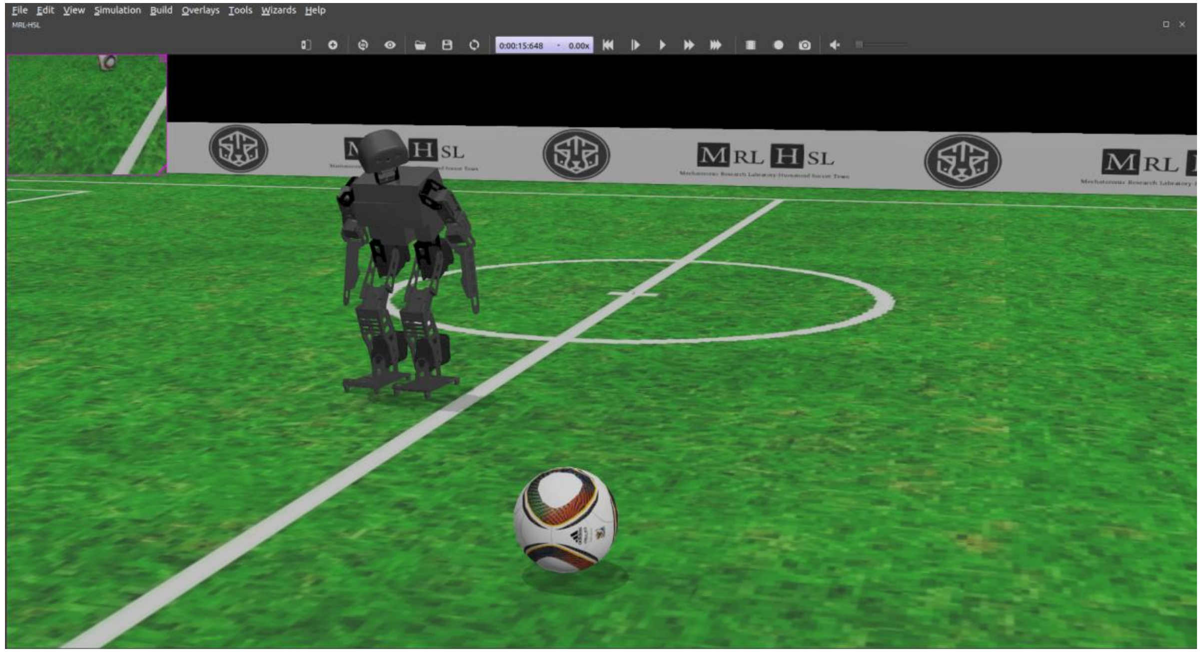
01

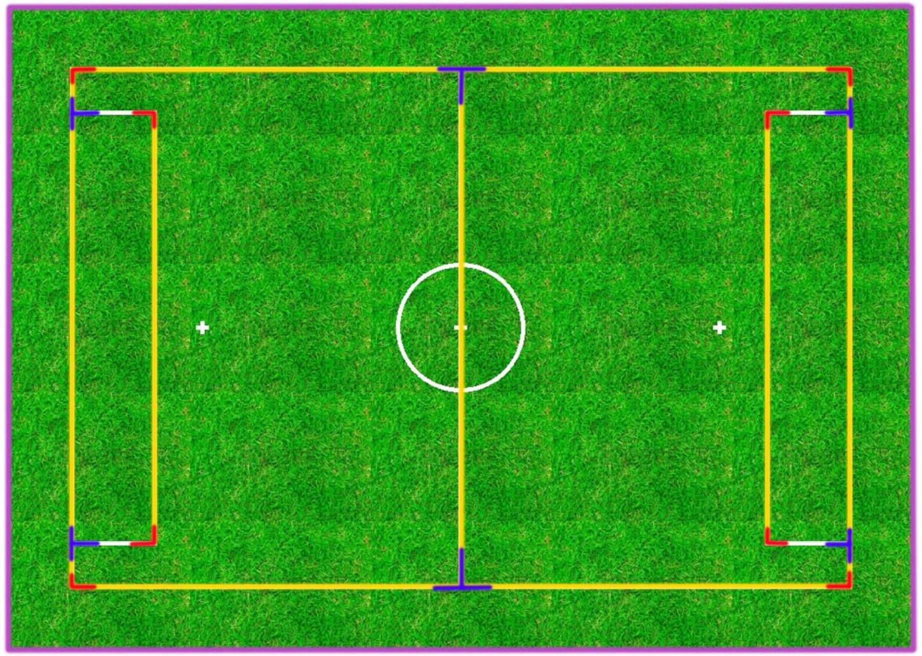
Introduction

1. Introduction

- **Main idea:** adaptively optimizes the viewpoint of the robot to acquire the most useful landmarks for self-localization while keeping the ball into its viewpoint
- **Task setting**:

- Webots simulator: viewpoint is controlled by the actuators of the robot’ s head
- Environment: a Robocup soccer field





02

Methods

2. Settings and Process

- **Predefine:**

- Camera position (viewpoint): **discrete**, 10 pan $\times$ 4 title angles $p = (\theta_{pan}, \theta_{tilt})$

$$\frac{-\pi}{2} < \theta_{pan} < \frac{\pi}{2}$$

$$\frac{\pi}{36} < \theta_{tilt} < \frac{13\pi}{36}$$

- Observations: gray-scale image
- State: a sequence of observations
- Action: **discrete**, 3 degrees rotations in a certain direction
- Reward: D/D': distance to the goal position (NBV) before/after taking the action

$$Reward = \begin{cases} -2, & \text{for missing all balls} \\ \operatorname{sign}(D' - D), & \text{elsewhere} \end{cases}$$

- **In each episode:** Randomize Robot and Soccer positions

→ Find "ground truth NBV" of current position

→ Select which action to take by DQN

2.1 Entropy-based Goal Determination

- Find "ground truth NBV" of current position

Algorithm 1 viewpoint exploration

Input: $X_t = \mathcal{N}(\mu_t, \Sigma_t)$, $ballpose = (x_t, y_t)$

$H_{min} = \infty$

for each $p \in P$ **do**

 $X' = X_t$

 $Z = get\_observations(X', p)$

 for each $z \in Z$ **do**

  $X' = apply\_UKF(X', z)$

 end for

 $H_{X'} = \frac{1}{2}\ln(|(2\pi e)\Sigma|)$

 if $H_{X'} < H_{min}$ and $ball\_is\_visible(X', ballpose)$ **then**

  $H_{min} = H_{X'}$

  $p^* = p$

 end if

end for

return p^*

represent the belief of the robot position using **multivariate normal distribution**

get all observations Z from visible landmarks at the current position, each z is related to a visible landmark and represents the distance and angle of the landmark to the robot.

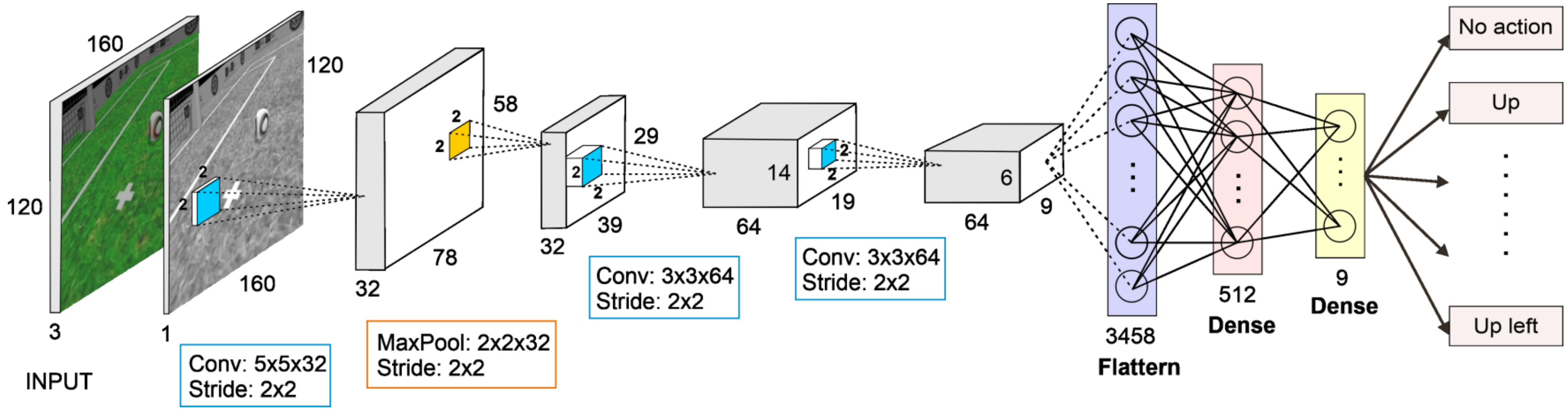
for every z, update belief X' using **Unscented Kalman filter**

Calculate the entropy of updated belief X'

The best viewpoint p* contains the ball and minimizes the entropy of the model

2.2 RL Process

- Select action by **DQN +PER:** train the neural network from a **prioritized experience replay** in which important transitions are picked more frequently
- Input: current image (independent to the localization accuracy)
- Output: a vector of Q-values whose length equals the number of possible actions



- Episode:
- Success if **reaches the goal viewpoint**
- Fail if misses the ball from its field of view
- Terminates after 20 timesteps to avoid long episodes

03

Experiments

3.1 3 criteria

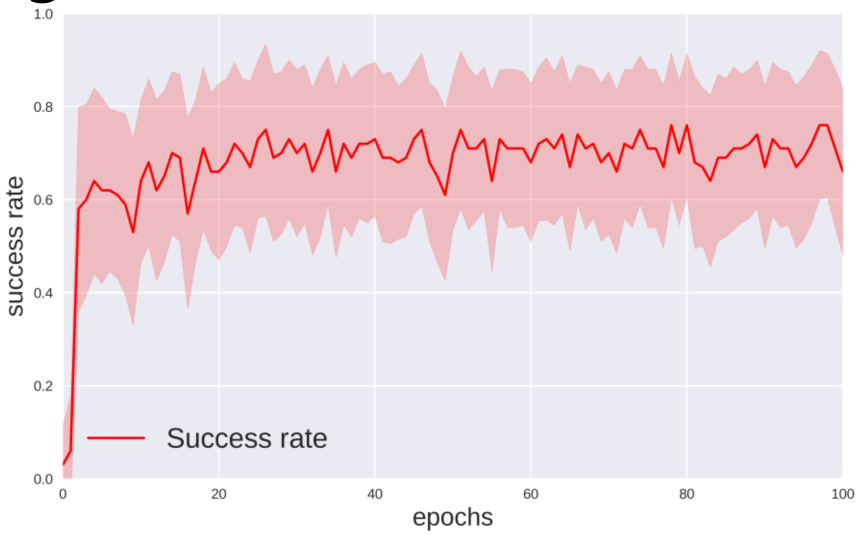
- **SuccessRate:** indicates how much of the desirable landmarks in the best viewpoint are observed

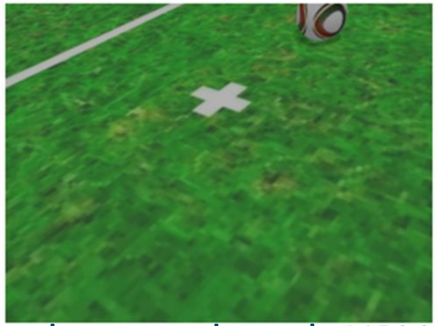
$$SuccessRate = \frac{|\{observed\ landmarks\}|}{|\{desired\ landmarks\}|}$$

- **Success duration**: timesteps taken to reach the goal viewpoint. 20 means not to reach the goal.
- **Ball loss duration**: timesteps taken to lose the ball. 20 means the robot has not lost the ball in that episode.

3.2 Results

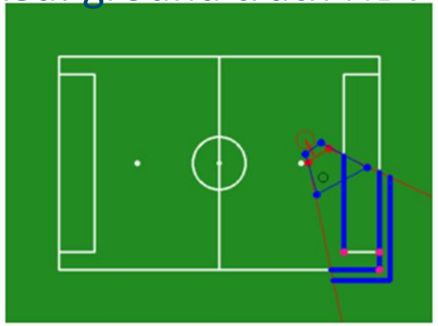
- Agent has been trained for 30000 timesteps

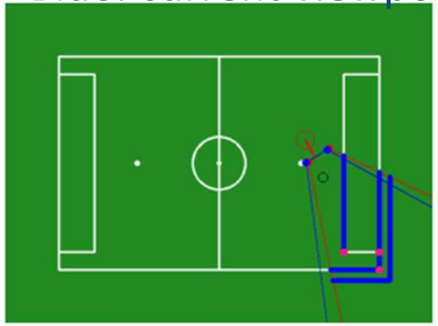


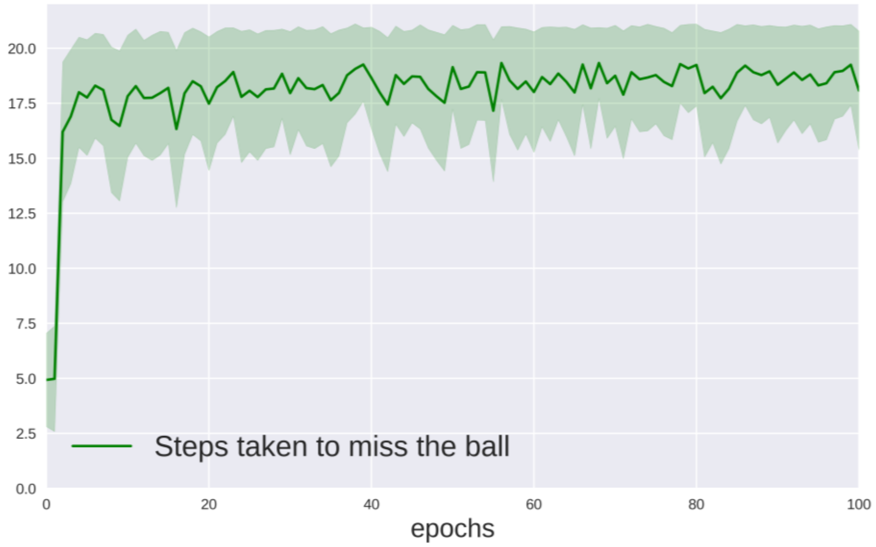


Red: ground truth NBV

Blue: current viewpoint

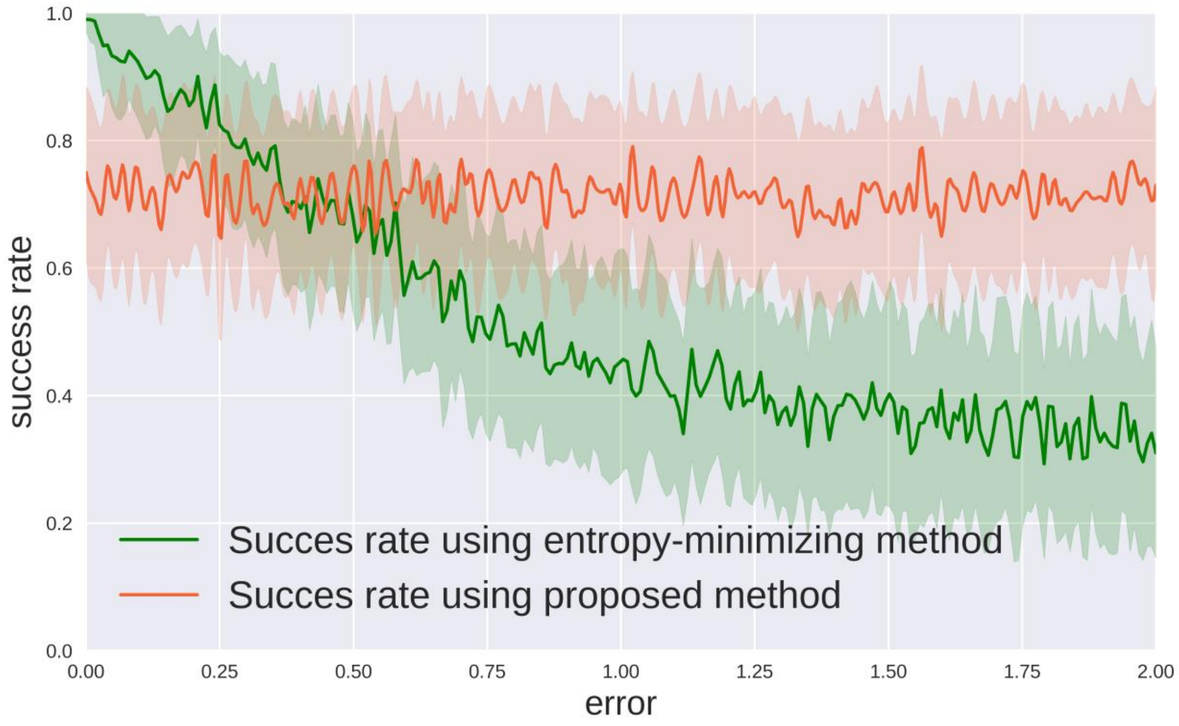






3.3 Comparison with entropy-based method

- Entropy-based methods have had the best performance in active vision tasks so far in RoboCup and similar contexts
- Different episodes start from different random positions as the self-localization error increases



- entropy-based methods operate as a function of positions → the more localization error, the lower performance, highly dependent on the accuracy of the self localization.
- proposed method works as a function of the current input image and **doesn' t rely on the localization accuracy → the performance remains steady.**

04

Limitations

4. Limitations

- The problem can be solved with newer algorithms of RL that consider the **continuous action space** such as PPO and DDPG.
- The performance of the method might be improved by passing a rough representation of the robot **position along with the image**.
- My opinion:

\- RL only for achieving the goal NBV not for selecting unknown NBV

→ it is difficult to prove that the calculated best view is indeed the best using entropy minimization alone.