Facing the ball carrier in AI World Cup soccer

Tim Velthoven

June 27, 2021

Abstract

Robot soccer competitions have been around for a while and have been a great environment to develop AI algorithms in. One of these environments is the AI world cup. The AI world cup environment is a virtual environment where two teams with five robots each play a soccer match. This paper focuses on defending the attacker that is carrying the ball in the AI world cup environment. This is achieved by comparing two approaches, a rule-based algorithm, and a Deep Deterministic Policy Gradient reinforcement learning algorithm. The two algorithms are based on a coaching framework build by the TU Delft. This framework implements the actions of the robots, such as moving, jumping, and kicking the ball, and the communication between the robots. The two approaches have been evaluated and the rule-based outperforms the reinforcement learning algorithm. Furthermore, a teamwork strategy between the goalkeeper and defender has been developed, where this teamwork is a significant improvement over the non-teamwork algorithm.

1 Introduction

Team sports games are an important testbed used for benchmarking Artificial Intelligence algorithms [1], with many different game elements and possible tactics. With all the possible game elements and tactics, such as collaboration between the players and individual skills of the players, one wants to find a tactic that ideally wins all games.

1.1 Background

There has been an extensive history of football simulation competitions, where teams play each other with their own tactics and game plan. The most well-known competition is the RoboCup, where robots play a 5 versus 5 match on a real pitch. Lately there has been a shift to simulations of the RoboCup format, where these competitions aim to focus more on the algorithmic aspects of the competition and not so much on the physics involved. One of these competitions is the AI world cup, a football simulation competition with virtual robots. This competition is created to make it easier to develop and experiment with new tactics, designs and architectures.

To help with the development of AI algorithms, a framework for the AI World Cup [2] is used. This framework has implemented a hybrid AI architecture featuring agents for the 5 robot players plus one for a coordinator coach. For all the agents, various essential skills (including accurate 2D and 3D passing) have been implemented and tested for attacking and defending purposes. Of course, a winning team needs to perform satisfactorily at all skills, not just in one or two. This research aims at directly contributing to this team's success by focusing on the development of smart defensive skills, specifically defending the attacker with the ball. Defending the opponent carrying the ball is an essential part of the defense. When the attacker with the ball is not properly opposed, it is easy for the attacker to shoot at an uncovered part of the goal or to make a pass to another attacker in a dangerous position.

1.2 Research questions

One of the most essential tasks in defense is defending the opponent who is carrying the ball. In order to research this, the main question that this research aims to answer is "What are the best positioning methods for a defender to face the attacking opponent who carries the ball?". This research question can be broken down into multiple sub-questions, which when all answered form the answer to the main research question together. The sub-questions are "On which factors should we decide which player should oppose the ball carrier?", "How could we decide where to position a defender to hinder shot access to the uncovered regions of the goal?" and "How could we decide where to position a defender to hinder passing possibilities to the opponent's team members?".

1.3 Relevant work

Recently there has been a lot of success in the development of Reinforcement Learning (RL) algorithms for multi-agent games [3]. In the case of robot soccer, we are dealing with an action space that is a discrete-continuous hybrid [4]. The MP-DQN algorithm [5] is designed to work with a discrete-continuous action space and is proved to outperform other existing methods, such as P-DQN [6], on the Half Field Offense (HFO) domain [4]. As the HFO domain is similar to the AI world cup domain, it can be useful to use in this research. In 2019, the two teams in the final both used different approaches. One team used a machine learning technique based on multi-agent deep deterministic policy gradient (MADDPG) algorithm [7], whereas the other team used a pure rule-based approach [8]. Because this research is aimed at training a single agent, a deep deterministic policy gradient algorithm for single agents [9] is more fitting than the multi-agent counterpart. As for the rule-based approach, [10] describes a man-marking strategy to organize the defense, but apart from that there are no other rule-based approaches researched before.

1.4 Contribution

Since this paper focuses on a single defender, a single-agent deep deterministic policy gradient [9] is implemented and compared with a self-made rule-based strategy described in section 2.2.2 for covering the goal. Furthermore, the rule-based algorithm has been extended to include teamwork between the goalkeeper and the defender. In addition, a rule-based method for covering opponent players is presented.

2 Methodology

To answer each of the sub-questions, several experiments have been created to measure the performance of the solutions. The webots [11] framework in which the AI world cup is simulated will be used. In this framework, the behavior of both teams can be controlled, which is ideal to conduct experiments with. Every frame, all players can perform one of 4 actions:

Go to, kick, jump and turn. These actions form the basis of the algorithms presented in this section.

2.1 Marking strategies

For the question "On which factors should we decide which player(s) should oppose the ball carrier?", two strategies have been considered: Man marking and zonal marking. These strategies are the two marking strategies used in normal soccer. In man-marking, the defender that is between the defender's goal and the opponent with the ball and closest to the ball is the defender that is assigned to the attacker. By doing this, it is made sure that the attacker that has possession of the ball can not dribble directly to our goal, as the defender can oppose the attacker and cover either shot access to the goal or any passing possibilities to other attackers. In zonal marking, each defender covers an opponent when the opponent is in the zone where the defender is responsible for. As evaluating marking strategies requires control over all the players of the defending team and this research focuses on one defender opposing the ball carrier, the marking strategies described in this section are difficult to evaluate. Therefore, the evaluation of the marking strategies is left for future research.

2.2 Covering the goal

To answer the question "How could we decide where to position a defender to hinder shot access to the uncovered regions of the goal?" two algorithms are implemented and compared. The first algorithm is a rule-based method where the defender defends the side of the goal that is uncovered. The experiment is then run 100 times to see how this rule-based approach performs in terms of goals conceded. Afterward, the DDPG Reinforcement Learning algorithm is implemented, the experiment is run again and the performance is compared to the rule-based approach.

2.2.1 Experiment setup

To measure the performance of the two algorithms, a 1v2 experiment is created. The goal of this experiment is to measure the impact of the position of the defender in the rule-based and Reinforcement Learning approach. The experiment is set up as follows.

The defender starts at a random position between the penalty area and 1 meter before the halfway line, and at least 1 meter from the sidelines, see figure 1. This is done to replicate a real match situation, as the defender will usually not be at the side of the field or in the penalty area. The attacker starts at a random position on the halfway line. Once both players are at their random position, the attacker dribbles straight towards the middle goal. When the defender is within 1 meter of the attacker, the attacker shoots at the goal. This is done to prevent that the defender dispossesses the attacker, as that is already researched [12]. The attacker is only shooting low shots. The reason for this is that the attacker can shoot instantly when shooting a low shot, but it needs time to set up a high shot. In the time the attacker needs to set up the high shot, the defender is already so close to the attacker that it is impossible to shoot past the attacker.

The goalkeeper used in this experiment is a goalkeeper that moves over the goal line [13]. The goalkeeper moves to the side of the goal where the ball is, so if the ball is on the right side of the field the goalkeeper slightly to the right side of the goal and vice versa. To measure the performance of the defender and goalkeeper, it is recorded if the defending team conceded a goal or not.

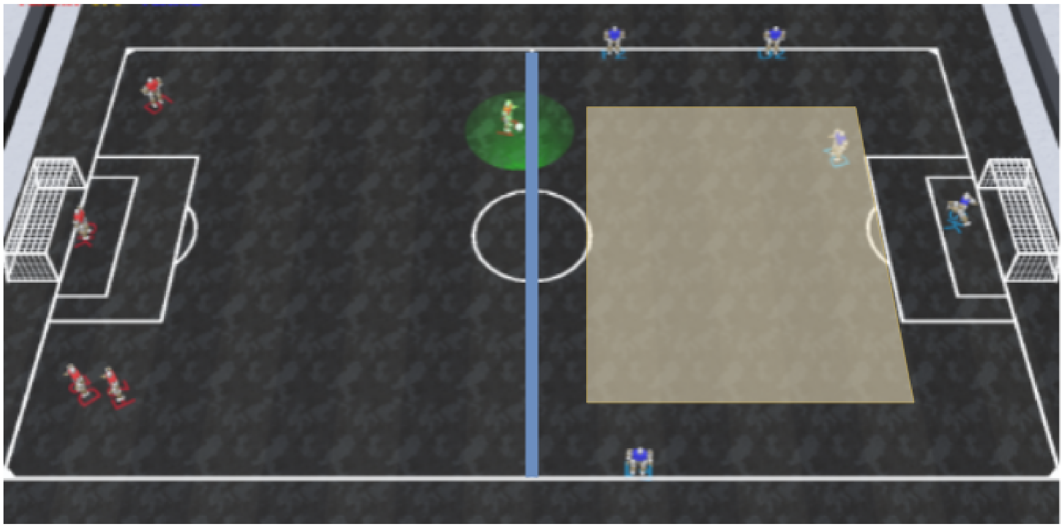

Figure 1: Experiment setup. The attacker starts with the ball on the blue line and the defender starts at a random position in the yellow area

2.2.2 Rule-based algorithm

To know which part of the goal is uncovered, the position of the goalkeeper has to be taken into account. Since the goalkeeper that is the most effective moves to the side where the ball is [13], the side of the goal that the goalkeeper is not covering is naturally the side that has to be covered. To cover this side of the goal, the defender covers the area next to the goalpost. This is done in the following way. First, the line from the ball to the goalpost is calculated. Then the defender calculates the closest point on this line to its own position. If the distance to this point is bigger than the distance from this point to the attacker, this point is taken as the target point. If this is not the case, the defender calculates a new point on this line closer to the goalkeeper, where the defender can reach this point faster than the attacker. This is done to prevent the defender from ending up behind the attacker and essentially giving the attacker a shooting opportunity without any cover of the defender. Once the defender is on at the target line, it moves a bit to the side of the line, such that the side of the robot touches the line, as depicted in figure 2. In this way, the defender is covering the largest possible area next to the goalpost. When the defender is at the green point displayed in figure 2, it turns toward the ball and moves to the ball. The closer the defender is to the ball, the bigger the area of the goal is that the defender is covering. Every frame the desired position on the line is recalculated to quickly react to any movements of the opponent, such as when the attacker is not going directly towards the goal.

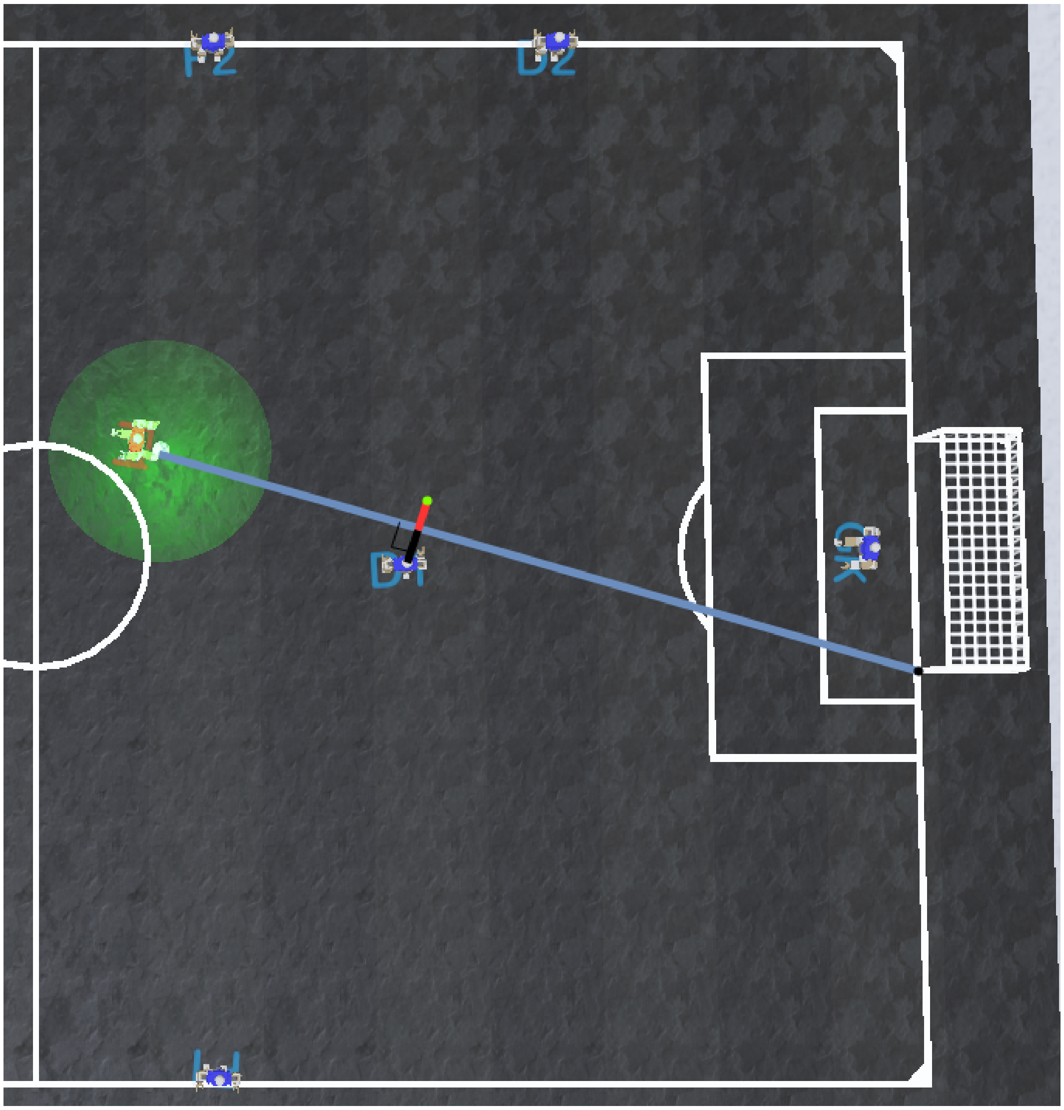

Figure 2: Rule based positioning of the defender. The green point is the target position of the defender. This position is half the width of a robot from the line that has to be defended. In this way, the ball can not be shot past the defender in the direction of the goalpost it is covering.

2.2.3 DDPG

The DDPG Reinforcement learning algorithm is implemented according to previous research [9]. Since the defending of the goal is solely based on positioning the defender, the wheel speeds of the defender are used as parameters to train. In this way, the results of training the algorithm are not influenced by the implementation of a GoTo(x,y) action.

The main reward used in training the algorithm comes from not conceding a goal. However, such a reward is too sparse for the agent to learn a good strategy. To overcome this problem, a reward is added for moving closer to the ball. This component provides a scalar reward corresponding to the change in distance from the agent to the ball d(a,b). This makes the total reward function as follows:

$$r_t = d(a,b)_{t-1} - d(a,b)_t + R_t^{goal-prevented}$$

One episode is defined as one run of the experiment described in section 2.2.1. The episode starts when the ball crosses the halfway line and ends when the ball is scored, out of play or the goalkeeper or defender blocked the ball.

2.2.4 Rule-based teamwork between the goalkeeper and defender

When implementing the rule-based algorithm, high shots were not taken into account. Since high shots aimed at the top corners of the goal are a weakness of the line goalkeeper[13], this is a part of covering the goal where the defender and goalkeeper can work together to improve the defensive performance.

For the purpose of testing the teamwork, the experiment described in section 2.2.1 is slightly adjusted. Instead of the attacker walking straight to the goal, it moves to a random position on the half of the defender. The defender starts at the (-2, 0) coordinate and starts covering the goal when the attacker crosses the halfway line.

In order to defend the top corners, the algorithm demonstrated in section 2.2.2 is extended. Since teamwork needs coordination, the defender takes the initiative. When executing the teamwork rule-based algorithm, the defender takes the initiative by moving to the target position, as depicted in figure 2. When the defender is within 1 meter of the line and within 3 meters of the ball, the defender communicates to the goalkeeper that it is able to intercept a shot that is aimed at the top corner that the defender is covering. When the goalkeeper receives this signal and it moves to the side of the goal that the defender is not covering, as is displayed in figure 3. For the keeper to be able to save a shot in the top corner, it moves to the y coordinate -0.55 or 0.55, depending on the side which the goalkeeper is already standing. The x-coordinate of the goalkeeper stays the same, to cover the corner as quickly as possible.

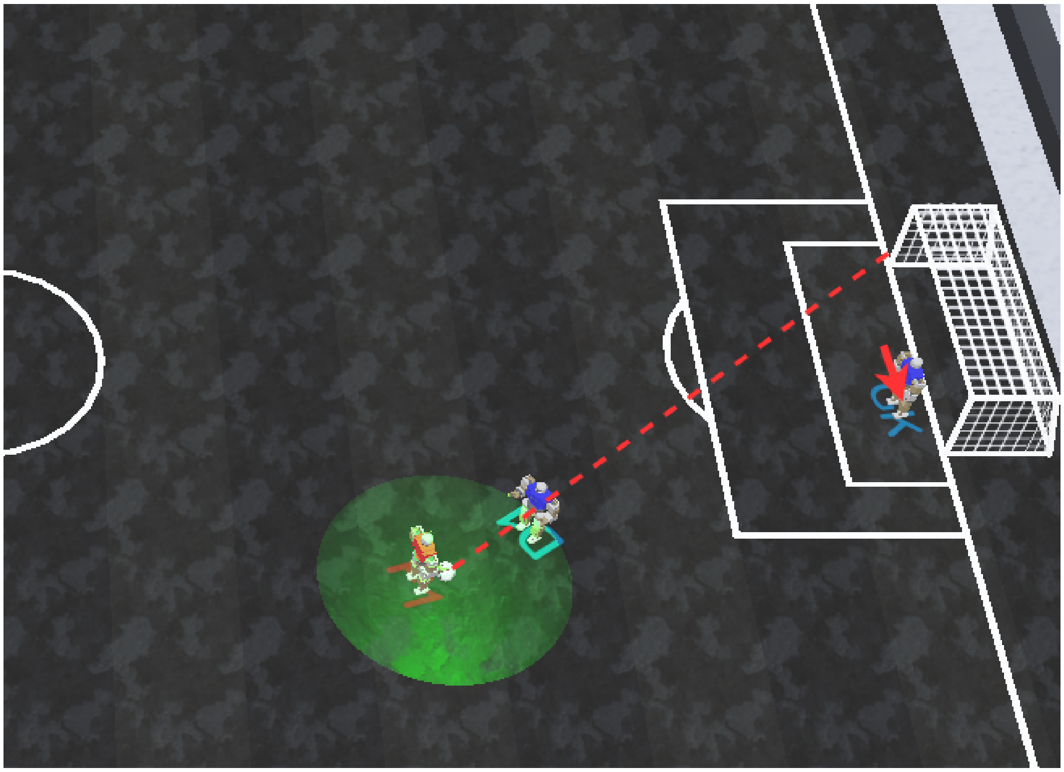

Figure 3: Rule based teamwork between the defender and the goalkeeper. When the defender is covering the far corner, the keeper moves over to the side that is not covered by the defender, so together they cover the whole goal.

To test the performance of the implemented teamwork algorithm, the adjusted experiment will be run 100 times without moving the keeper as a baseline and 100 times with moving the keeper after the defender is in position.

2.3 Covering other players

The last sub-question to answer is "How could we decide where to position a defender to hinder passing possibilities to the opponent's team members?". In order to answer this question, a rule-based algorithm is proposed, based on results of related work [14]. To hinder a passing possibility from the attacker carrying the ball to any of the other attackers, the objective of the defender is to be positioned on the line between the attackers. The coordinate on this line closest to the defender is calculated and subsequently, the defender moves to this coordinate.

To anticipate any actions of the player the defender is covering, the defender will mimic the movements of the attacker. For example, if the attacker turns away from the ball to prepare a movement, the defender will mimic this turn. In this way, the defender is always ready to act on any movements of the attacker that is supposed to receive the ball.

3 Experimental results

For training the DDPG network, the following parameters have been used: a gamma of 0.9, batch size of 64, buffer size of 100000, an actor learning rate of 10^{-5} and a critic learning rate of 10^{-4}. After training the DDPG algorithm on more than 4000 episodes, there is a slight uptrend visible in the average reward per frame at the end of the training procedure, see figure 4.

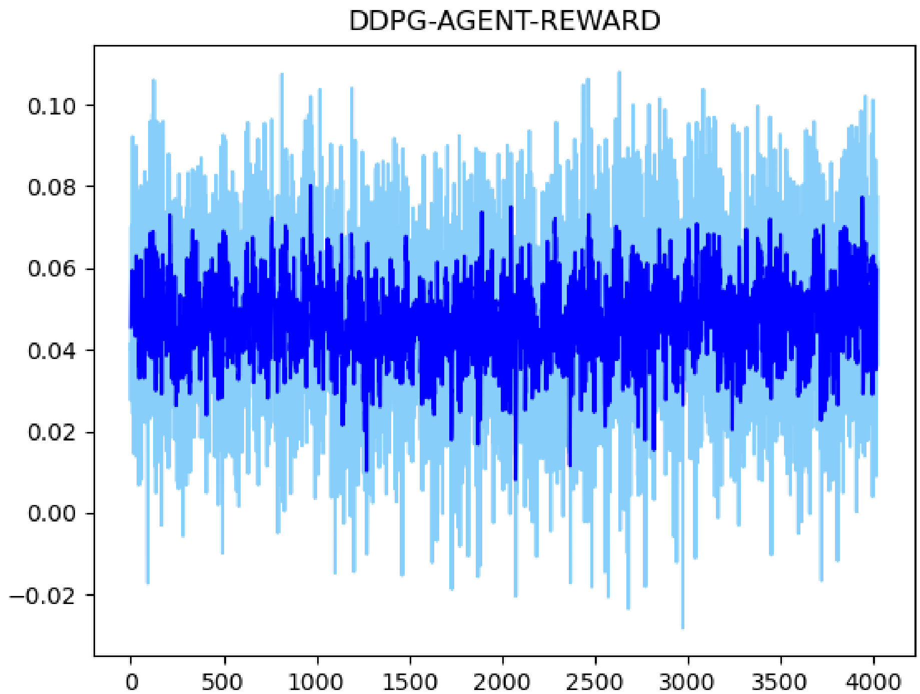

Figure 4: Dark blue: Average reward per frame over all episodes. Light blue: Average reward per frame of the episode.

The performance of the rule-based and DDPG algorithms has been evaluated by running the experiment described in section 2.2.1 100 times for each algorithm. This is 1 run. For both algorithms, 3 runs have been performed in order to minimize the impact of the randomness involved in the experiment. The results of these 3 runs are presented in table 1. It is clear that the rule-based algorithm outperforms the DDPG algorithm.

	Run 1	Run 2	Run 3	Mean
Rule-based	7	5	7	6.3
DDPG	11	12	14	12.3

Table 1: Amount of goals scored out of 100 shots each run for the basic rule-based algorithm and the DDPG algorithm.

The performance of the rule-based teamwork algorithm has been evaluated by running the experiment described in section 2.2.4 100 times for each algorithm. This is 1 run. For both algorithms, 3 runs have been performed in order to minimize the impact of the randomness involved in the experiment. The results of these 3 runs are presented in table 2.

	Goals conceded
Without teamwork	43
With teamwork	28

Table 2: Amount of goals conceded out of 100 shots aimed at one of the top corners.

4 Discussion and responsible research

Because the results of the basic rule-based algorithm and the DDPG are close to each other, the run of 100 shots is repeated two more times to validate the results of the experiment. For the teamwork performance, this is not done because the difference between the teamwork and no teamwork results is quite big and there are fewer random actions involved. As this research is conducted in a limited amount of time, there was little time to let the DDPG algorithm train for a longer period. This can slightly impact the performance of the DDPG algorithm in the results.

All the results of this research can be reproduced by following all the steps in the research. All steps of the algorithms and experiments are described in detail. Because all steps are described in detail, the code will not be made publicly available, as the code will be used in the future participation of the TU Delft team in the AI World cup.

5 Conclusions and future work

This paper is a first try at developing and comparing algorithms for facing the attacker carrying the ball. This problem is split up into three smaller subproblems: Which player is facing the ball carrier, how can the defender facing the ball carrier cover uncovered regions of the goal and how can the defender facing the ball carrier cover the other opponents. For deciding which player should face the ball carrier, the position of the defender is the most important factor to decide which player should face the ball carrier. When covering the goal with the defender, the DDPG algorithm failed to learn a better method of covering the goal than the rule-based solution proposed in this research. An extension for the rule-based covering of the goal was developed to minimize the number of goals scored when aiming at the top corners of the goal. The effectiveness of this extension was shown by comparing it against the non-extended version and it proved to be a significant increase in performance. Finally, a rule-based algorithm for covering the opponent's players has been presented. The results of covering other players are promising.

In the future, the algorithms described in this research can be integrated into a complete defensive tactic, which features a line goalkeeper [13], abilities to catch the opponent by surprise and dispossess the opponent [12], and uses either zonal marking or man-marking.

When the full defense is implemented, the performance of the man-to-man marking and zonal marking can be evaluated against different attacking strategies.

Next to integrating the algorithms presented in this research into a complete defensive tactic, the single-agent DDPG algorithm used in this research can be extended to the multi-agent DDPG [7]. In this way, all the players can learn how to position themselves when defending and the goalkeeper can learn where to position itself taking into account the positions of the defenders.

References

[1] M. Mozgovoy, M. Preuss, and R. Bidarra, "Team sports for game ai benchmarking revisited," *International Journal of Computer Games Technology*, vol. 2021, 2021.

[2] K. Prakash, "Ai world cup framework." unpublished, N.D.

[3] P. Peng, Y. Wen, Y. Yang, Q. Yuan, Z. Tang, H. Long, and J. Wang, "Multiagent bidirectionally-coordinated nets: Emergence of human-level coordination in learning to play starcraft combat games," *arXiv preprint arXiv:1703.10069*, 2017.

[4] M. Hausknecht, P. Mupparaju, S. Subramanian, S. Kalyanakrishnan, and P. Stone, "Half field offense: An environment for multiagent learning and ad hoc teamwork," in *AAMAS Adaptive Learning Agents (ALA) Workshop*, sn, 2016.

[5] C. J. Bester, S. D. James, and G. D. Konidaris, "Multi-pass q-networks for deep reinforcement learning with parameterised action spaces," *arXiv preprint arXiv:1905.04388*, 2019.

[6] J. Xiong, Q. Wang, Z. Yang, P. Sun, L. Han, Y. Zheng, H. Fu, T. Zhang, J. Liu, and H. Liu, "Parametrized deep q-networks learning: Reinforcement learning with discrete-continuous hybrid action space," *arXiv preprint arXiv:1810.06394*, 2018.

[7] R. Lowe, Y. Wu, A. Tamar, J. Harb, P. Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *arXiv preprint arXiv:1706.02275*, 2017.

[8] C. Hong, I. Jeong, L. F. Vecchietti, D. Har, and J.-H. Kim, "Ai world cup: Robot soccer-based competitions," *IEEE Transactions on Games*, 2021.

[9] M. Hausknecht and P. Stone, "Deep reinforcement learning in parameterized action space," *arXiv preprint arXiv:1511.04143*, 2015.

[10] A. Kutsenok *et al.*, *Swarm AI: A solution to soccer.* PhD thesis, Citeseer, 2004.

[11] Webots, "Webots: Robot simulator," 2020. (Version R2020a) [Computer Software].

[12] S. Heslenfeld, "Ai soccer: the most effective methods to dispossess the opponent player," Tech. Rep. CSE3000-G40-5, Delft University of Technology, Fac. EEMCS, 2021.

[13] R. Oude Elferink, "Defensive positioning techniques for a goalkeeper in the ai world cup," Tech. Rep. CSE3000-G40-2, Delft University of Technology, Fac. EEMCS, 2021.

[14] T. van den Hurk, "Ai soccer: covering an opponent not in possession of the ball," Tech. Rep. CSE3000-G40-2, Delft University of Technology, Fac. EEMCS, 2021.