P-HRL: An Adaptive and Flexible Prediction-based Hierarchical Reinforcement Learning for Robot Soccer

Zongyuan Zhang, Tianyang Duan, Zekai Sun, Xiuxian Guan, Shengliang Deng, Yong Cui, Hongbin Liang, Heming Cui

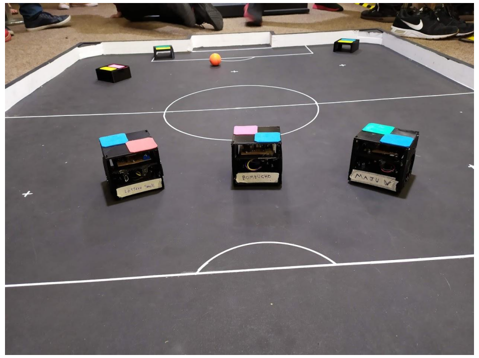
The University of Hong Kong

Tsinghua University

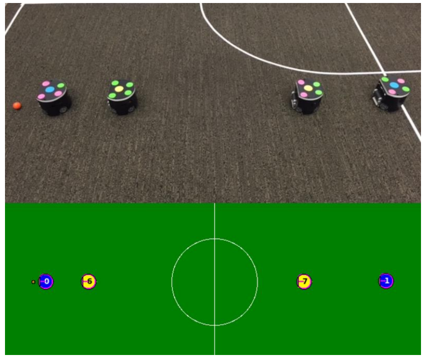
Southwest Jiaotong University

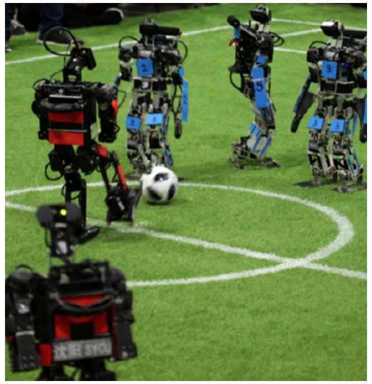
Motivation

- **Robot soccer** is a robot sports game in which both parties control a team of robots and cannot be manually controlled during the game.
    - A robot cooperation scenario: need effective team strategies and joint decision-making processes
    - Several leagues have been successfully organized in recent years.
        - e.g., RoboCup, IEEE Very Small Size Soccer (IEEE VSSS).



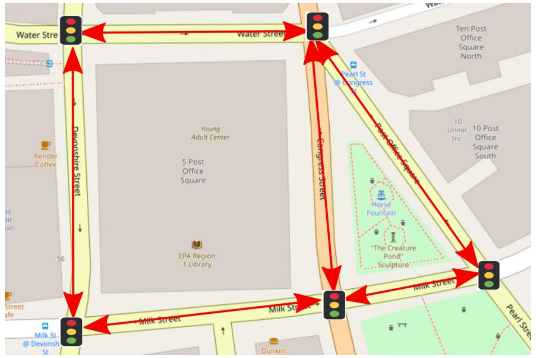
- Promoted research in academic scenarios
    - e.g., path planning, humanoid robot research





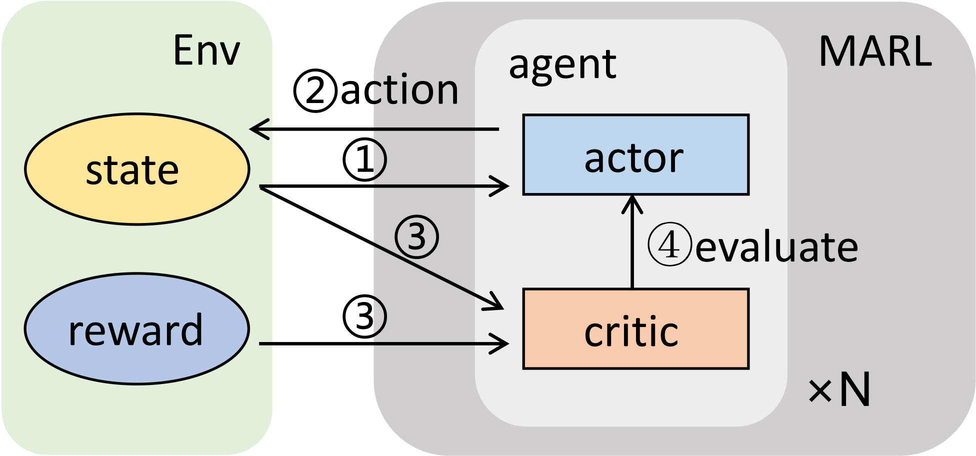
Existing MARL Method

- **Multi-agent reinforcement learning (MARL)** has achieved outstanding success on cooperative scenarios.
  - e.g., video games(Google research football), cooperative traffic light control.



- **Thus, the MARL approach has great potential in robot soccer.**

- MARL controls the actions of multiple agents **based on rewards** to maximize the expected rewards by continuously interacting with the environment.
  - Actor-critic structure
    - e.g., MADDPG, MATD3.
  - A "high-efficient" multi-agent reinforcement learning environment：diverse reward feedback and simple state-action space

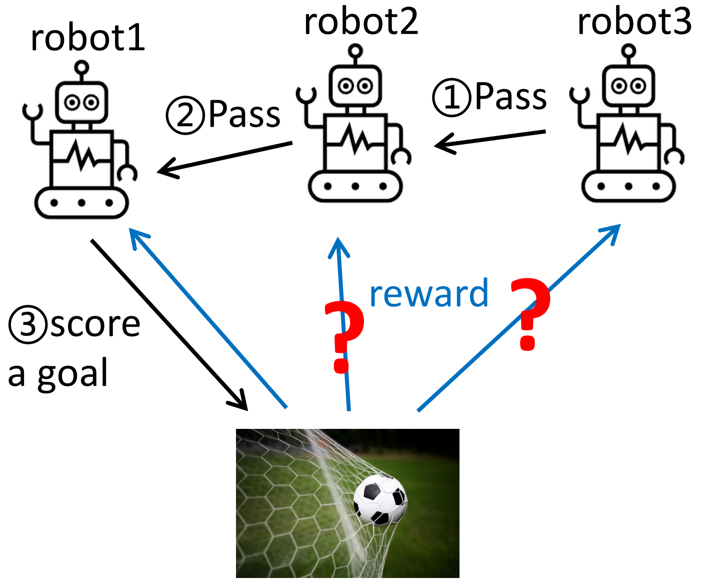


Problems of Existing MARL in Robot Soccer

- A "high-efficient" multi-agent reinforcement learning environment：
  - ① Continuous and timely reward feedback
  - ② Different reward feedback for each agent
  - ③ Low-dimensional and limited state-action space

- However, in robot soccer：
  - ① **Sparse reward**
    - Long-term process; occurs infrequently
  - ② **Global reward**
    - Cannot assign to each robot
  - ③ **High-dimensional and continuous state-action spaces**
    - Difficult to fully explore

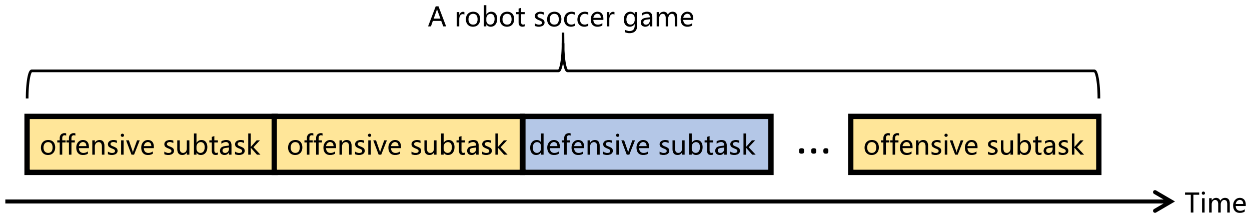
- It difficult for MARL to learn to collaborate in robot soccer.



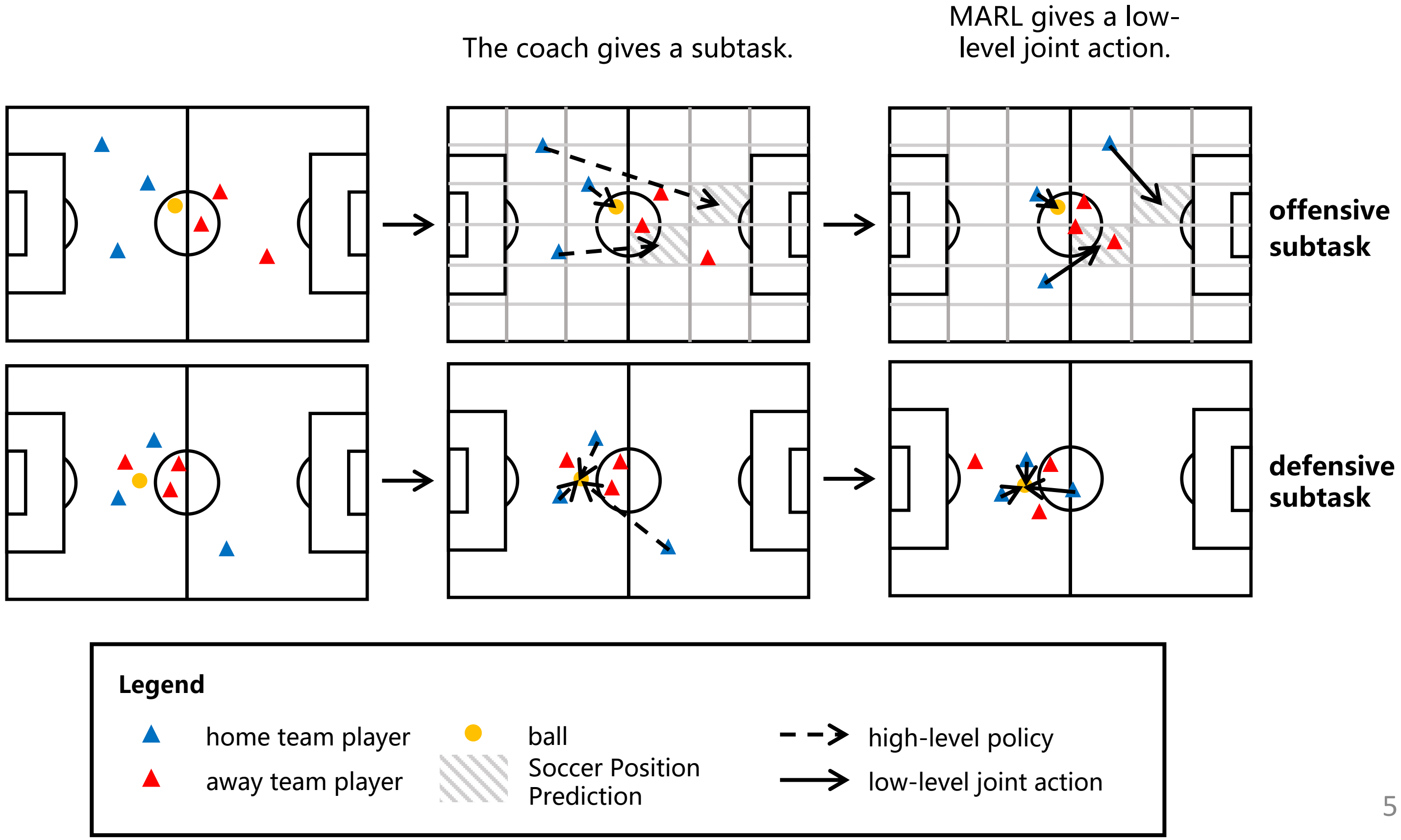
	Video game (Google research football)	Robot soccer (IEEE VSSS)
action space	**Only 20 actions** (e.g., left, short Pass.)	Left wheel speed: **range [-1,1]** Right wheel speed: **range [-1,1]**

Key Idea: Subtask Decomposition by Coach

- The **coach** assigns different subtasks to each robot.
  - Give each robot different and continuous reward feedback
- Two types of subtasks: **offensive subtask** and **defensive subtask**
  - **Guarantee possession of the ball**
  - Select based on ball possession

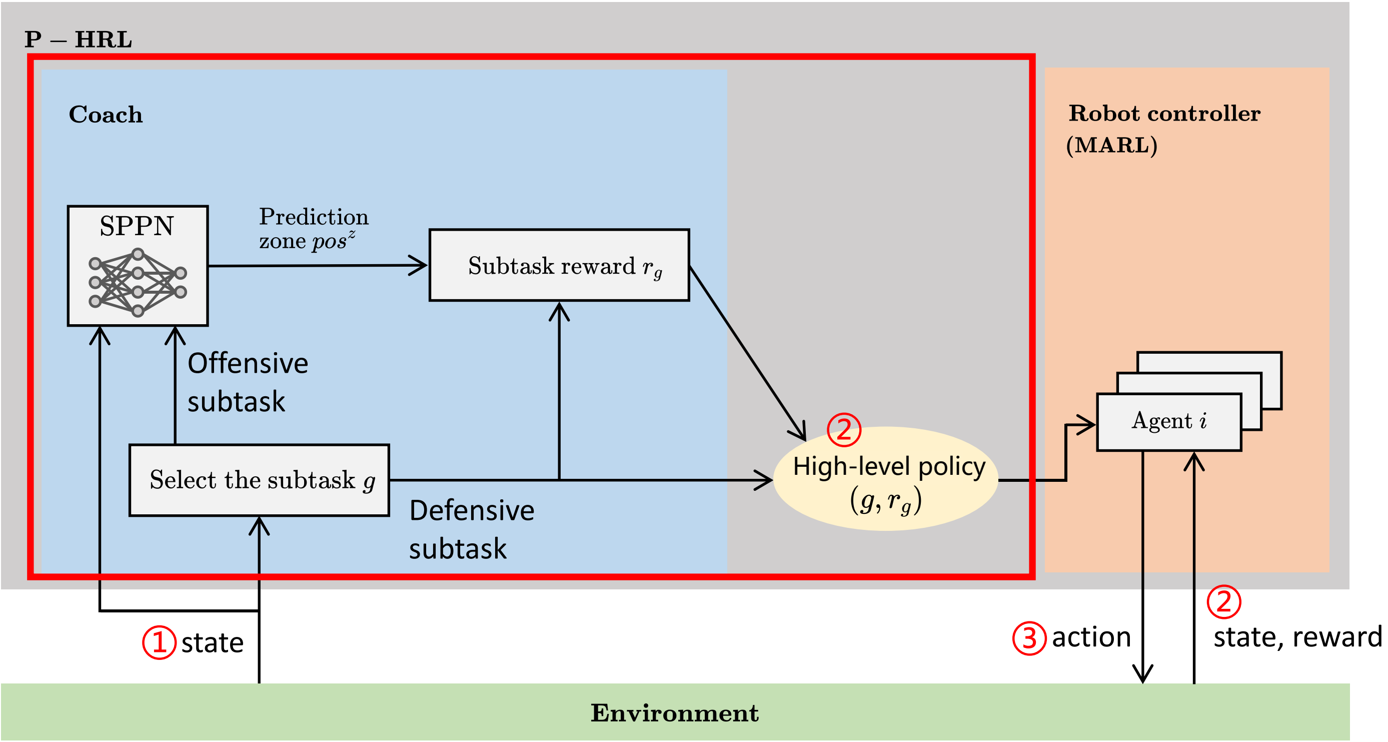


	Offensive subtask g^A	Defensive subtask g^D
Initiation condition I_g	Our robots are in possession.	Opposing robots are in possession.
Termination condition β_g	(1) Opposing robots are in possession. (2) The ball reaches the predicted zone (3) After T_p time steps.	Our robots are in possession.



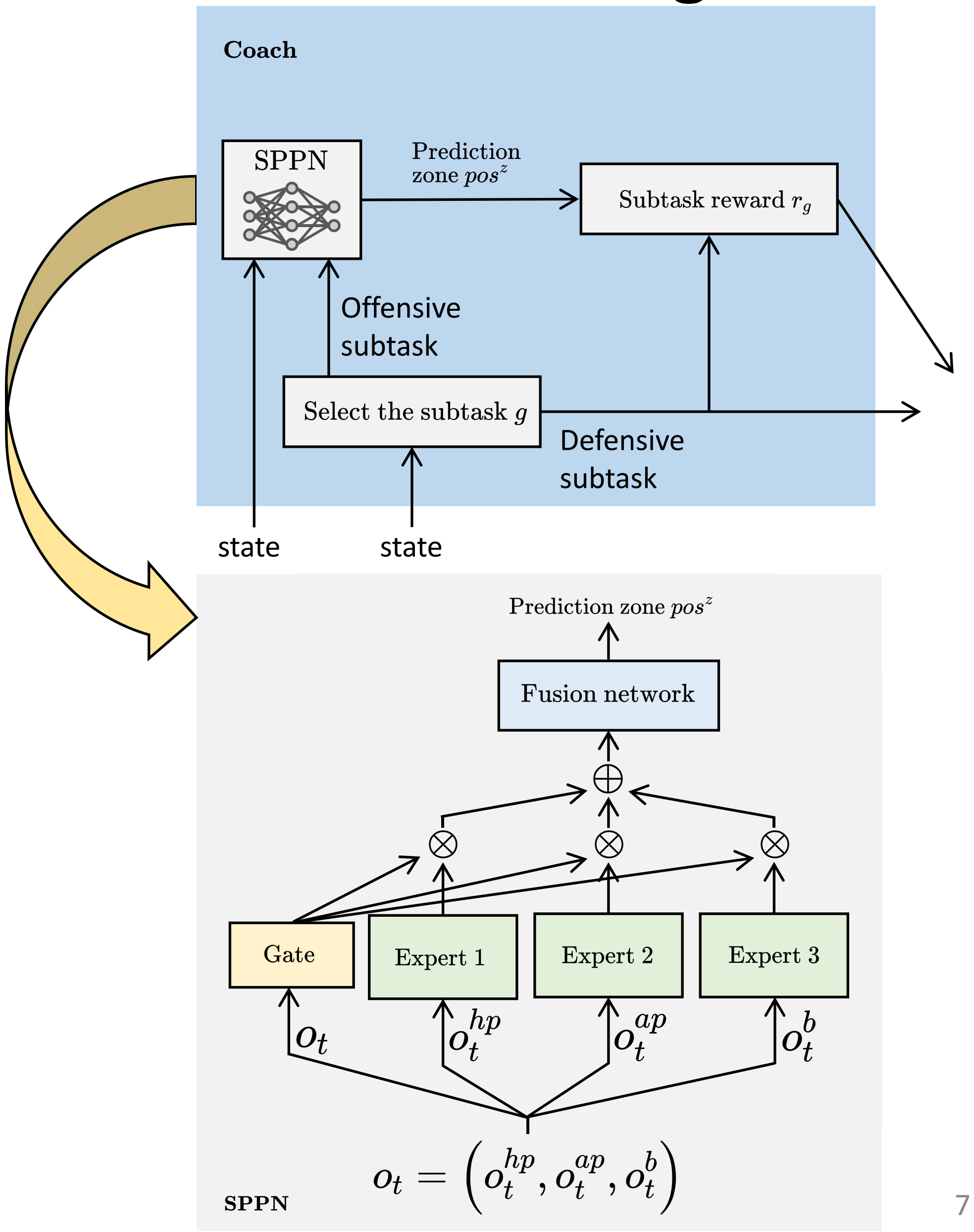
P-HRL: Prediction-based Hierarchical Reinforcement Learning

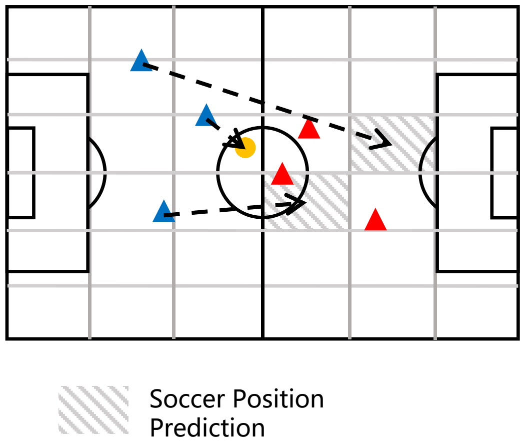
- **Hierarchical structure**
  - The coach provides a subtask and subtask reward to the robot controller(MARL).
  - Coach: adjusting soccer tactics
  - Robot controller: learning robot control.
- The robot controller uses a MARL method(MADDPG).
  - Added subtasks to actors and critics

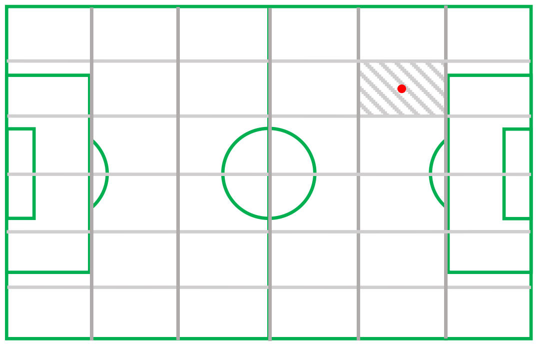


P-HRL: Prediction-based Hierarchical Reinforcement Learning

- In the offensive subtask, the robot needs to make judgments about the trajectory of the ball.
- The coach helps robot to judge the trajectory of the ball by predicting its position.
- **Soccer Position Prediction Network(SPPN)**
  - Mixture of experts(MoE) network
    - Independence of data in state
    - Easy to convergence
  - Zoning prediction



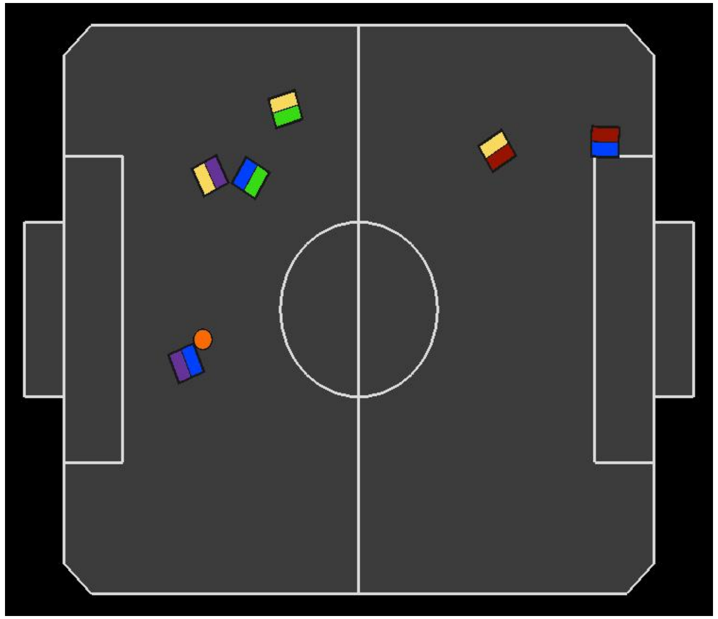




An example of zoning prediction

Implementation and Evaluation

- **Evaluation environment:**
  - rSoccer - IEEE VSSS multi-agent environment to simulate a robot soccer scenario.
- **Baseline opponent:**
  - MATD3 - a state-of-the-art MARL method that has been used in many multi-agent cooperative tasks similar to robot soccer.
- **Evaluation settings:**
  - 3 vs 3 robot soccer match
  - No hardware on robots for dribbling or kicking the ball
  - Follow the rewards shaping in rSoccer
  - Each match lasting 2000 time steps

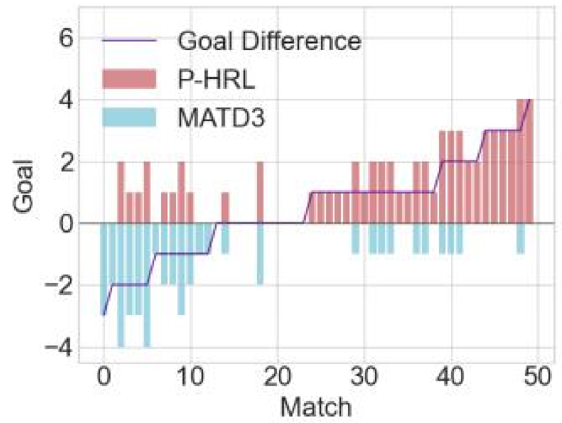


Evaluation Questions

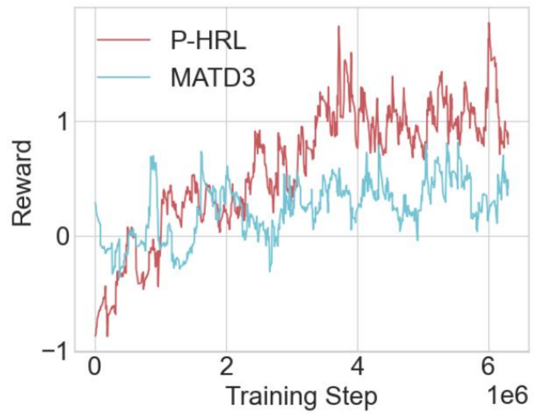
- How is P-HRL compared to baseline in terms of end-to-end performance?
- How does the coach contribute to the overall system?
- How well does the P-HRL adapt to the new opponent compared to the baseline?

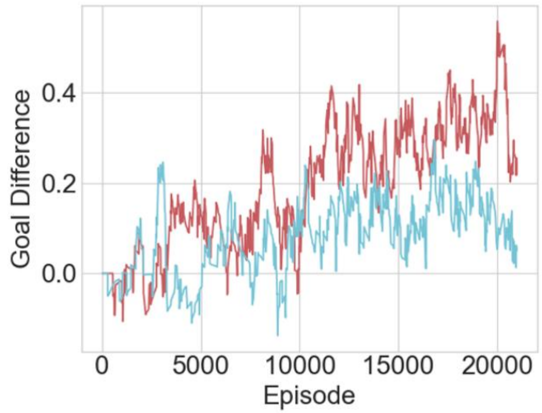
End-to-end Performance

- **P-HRL has 52% win rate**, 22% draw rate and 26% loss rate.
    - 50 matches: **P-HRL won 26 matches and tied 11 matches.** MATD3 won 13 matches.
- **Training method**：
    - Stage 1: Train P-HRL and MATD3 respectively, using rSoccer built-in agent as the opponent.
    - Stage 2: Train P-HRL and MATD3 in a mutual confrontation.
    - Each stage lasts until the average episode reward does not rise with the episode (about 1×10^6 steps).



- **Goal difference**：subtracting the number of goals scored by a team from the number of goals conceded in a match.

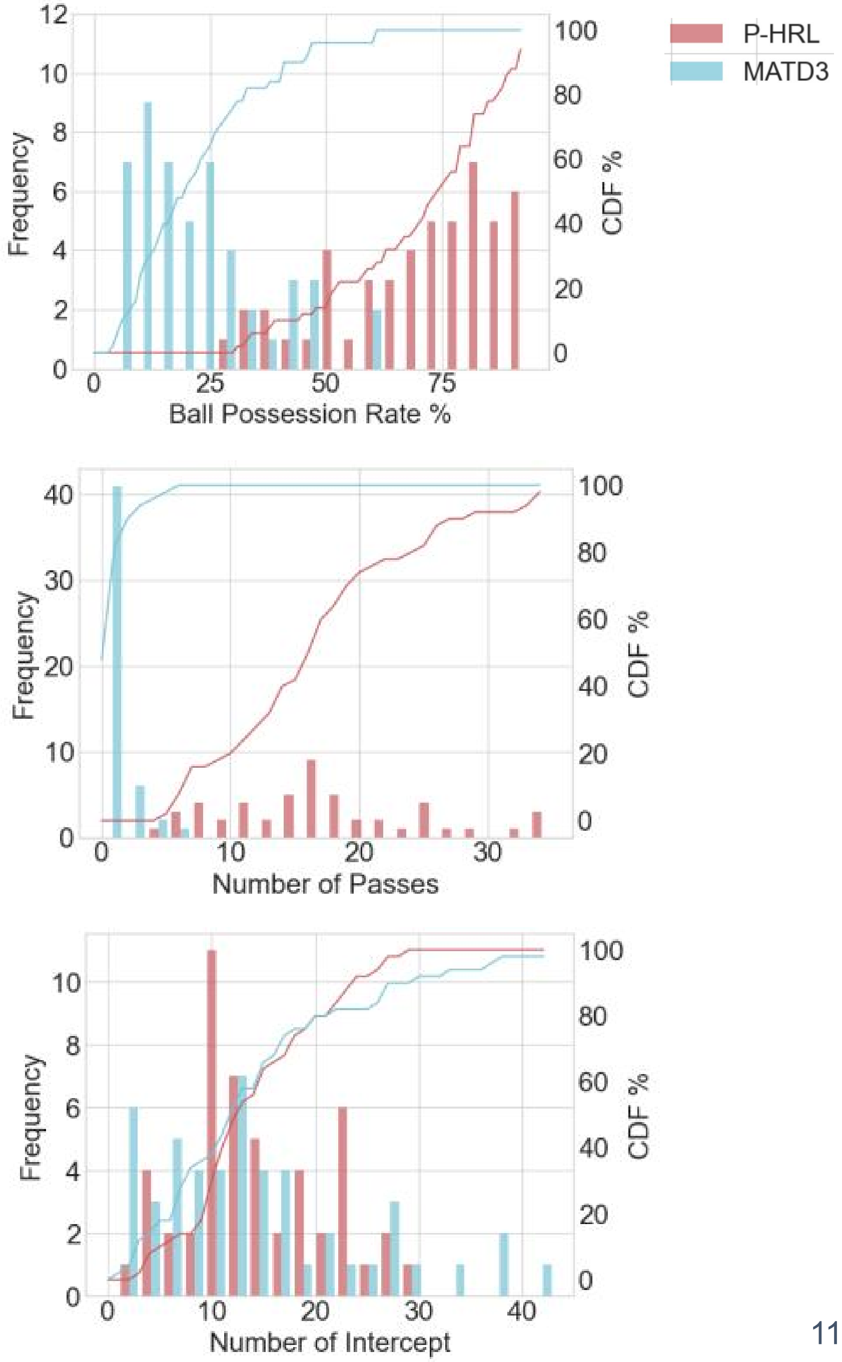




End-to-end Performance

- **Ball Possession Rate:** team possession of the ball as a percentage of total time
- **Number of Passes:** total number of passes on the team
- **Number of Interception:** total number of interceptions on the team
- **P-HRL outperforms MATD3 in ball possession rate and the number of passes.**

	Ball possession rate (per match)	Number of passes (per match)	Number of interception (per match)
P-HRL	**70.25%**	**14.32**	14.40
MATD3	17.14%	1.92	**14.44**

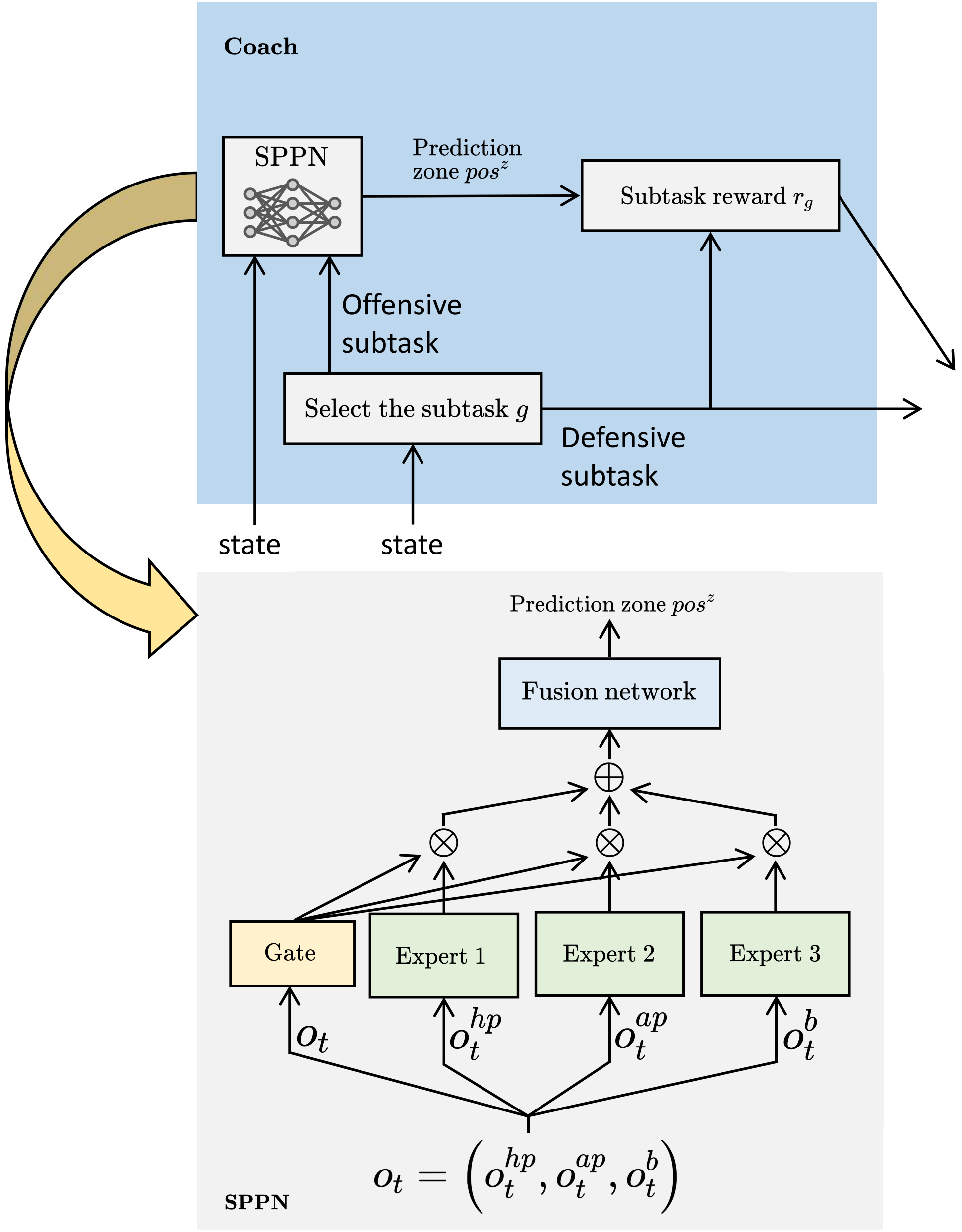


The Effect of the Coach to P-HRL

- **The coach brings better performance for P-HRL than baseline.**
  - NN Coach: Using a neural network with the same number of parameters as SPPN in the coach (but without the structure of MoE).
  - Random Nearest Coach: Using a method that selects two random adjacent zones of the current zone as the prediction results in the coach.

Coach	*Team Score : Opponent Score*	*Acc Top 1*	*Acc Top 2*
SPPN Coach	**2.40±1.78 : 1.48 ± 1.05**	**0.70±0.06**	**0.85±0.04**
NN Coach	2.15±1.22 : 1.55 ± 1.31	0.59±0.06	0.80±0.05
Random Nearest Coach	1.85±1.35 : 2.15 ± 1.39	0.06±0.11	0.12±0.06

Match results for different types of coach (vs. MATD3 in 50 matches)



Conclusion

- In this talk, we presented P-HRL, a prediction-based hierarchical reinforcement learning.
  - P-HRL consists of a coach for soccer tactics and a robot controller for robot motion control.
  - In matches against the state-of-the-art baseline MATD3, P-HRL has 52% win rate, 22% draw rate and 26% loss rate.
- We designed several key performance indicators (KPIs) for robotic soccer (e.g., ball possession) to more fully evaluate the performance of the P-HRL.
  - P-HRL has better cooperation between robots, with 70.25% possession rate compared to 17.14% for baseline.

Future work

- P-HRL has been submitted to the International Conference on Automated Planning and Scheduling (ICAPS2023).
- Short-term work
  - Add additional baselines, e.g., MAAC.
- Long-term work
  - Deploying P-HRL to real robots for soccer matches instead of evaluating it in the simulation environment.
    - Quickly correct deviations between the simulation environment and the real environment.
  - Using large-scale training models in robot soccer.
    - Optimize distributed training process.