Information Sharing for Cooperative Robots via Multi-Agent Reinforcement Learning

Ayesha Siddiqua
Department of Computer Science
Missouri State University
Springfield, MO, USA
as995s@missouristate.edu

Siming Liu
Department of Computer Science
Missouri State University
Springfield, MO, USA
simingliu@missouristate.edu

Razib Iqbal
Department of Computer Science
Missouri State University
Springfield, MO, USA
riqbal@missouristate.edu

Logan Ross
Joplin High School
Joplin, MO, USA
loganross@joplinschools.org

Brian Zweerink
Marshfield High School
Marshfield, MO, USA
brian.zweerink@mjays.us

Ryan Eskridge
Missouri State University
Springfield, MO, USA
rya132@live.missouristate.edu

***Abstract*—Facilitating collaboration within a team of robots poses a challenging question for the field of multi-agent reinforcement learning (MARL) in smart environments. Many existing cooperative MARL methods utilize centralized or decentralized frameworks leveraging global or local information for decision-making without sufficiently considering information exchange among agents. This research presents an innovative information-sharing approach for MARL, aiming to enhance collaboration among robots and improve overall team performance in multi-agent systems. In particular, our approach introduces an Information Sharing Matrix (ISM) that combines scenario-independent spatial and environmental information with each robot's local observations, thereby enhancing the performance of individual robots and improving their global awareness during the MARL learning process. To assess the efficacy of our approach, we conducted experiments on three cooperative multi-agent scenarios with varying difficulty levels implemented in *Unity ML-Agents Toolkit*. The experimental results indicate that robots employing our approach have effectively learned collaborative abilities, enabling them to maximize space coverage while avoiding conflicts among themselves. The robots utilizing our ISM-Shared variation outperformed those using decentralized MARL. They achieved performance comparable to robots employing centralized MARL, where complete global information is used for decision-making during the execution. Additionally, our ISM-MARL is adaptable across team sizes and consistently maintains high performance when transferring knowledge to teams of varying sizes, without being explicitly learned during the training phase. This suggests a resilient MARL learning technique that can adapt to changing environments.**

***Index Terms*—Deep reinforcement learning, multi-agent system, information-sharing, cooperative robots, Unity ML-Agent Toolkit**

I. Introduction

Modern smart environments have transformed into information-intensive cyber-physical multi-agent systems (MAS), propelled by the integration of technologies such as advanced robotics, the Internet of Things (IoT), and artificial intelligence (AI). In today's era of the ongoing industrial revolution, integrating machine learning (ML) and reinforcement learning (RL) with autonomous robots can greatly enhance industrial processes in smart environments. It is possible to devise a multi-agent system framework to further this integration by modeling an industrial production setting alongside autonomous robots. Although most RL research focuses on single-agent automation, there is an opportunity for enhancement within the realm of multi-agent smart environments. MARL extends single-agent RL and provides learning techniques to a group of evolving agents in cooperative and competitive tasks by maximizing rewards through agents' interaction with the environment and among themselves [1]. In this paper, we propose a MARL framework with a novel information-sharing mechanism for real-time task assignments, navigation control, and collaboration of autonomous robots to improve joint team performance in dynamic smart environments.

Various challenges emerge when defining problems in the domain of MARL, and one of these involves effectively facilitating internal communication among robots. Training MARL robots solely based on their limited local observations can easily lead to getting stuck in local optima and may struggle to learn cooperation in complex smart environments. Furthermore, within cooperative MASs, robots must act as a cohesive, coordinated entity to maximize the shared group reward. A comprehensive grasp of the team's dynamics and the surrounding environment around the robots becomes imperative to achieve team objectives. Numerous research initiatives, including those discussed in [2], [3], have addressed this issue by utilizing extensive information to eliminate the need for modeling communication among agents. Providing perfect global information during the RL training phase is usually called centralized learning. Although centralized learning can improve team performance, it often requires substantial computational expenses and may be unavailable in real-world scenarios. Numerous researchers have focused on modeling information exchange and communication among agents within smart environments to bolster team performance in collaborative tasks [4], [5].

In order to lessen reliance on complete information and promote team-wide information sharing, we propose the introduction of a novel multi-layer Information Sharing Matrix (ISM). This matrix serves as a shared knowledge repository aimed at augmenting the decision-making abilities of robots engaged in collaborative tasks within MAS. Furthermore, we present a variation of ISM called ISM-Shared that combines the local observations of individual robots in a team to emulate a global state, particularly in real-world scenarios where complete information may not be readily accessible. To assess the efficacy of ISM, we built multiple scenes to simulate warehouse and office building scenarios where a team of autonomous robots work together to collect packages while avoiding undesired objects like trash. To enable our robots to work efficiently and collaboratively across diverse situations, we employed our algorithms and trained robots with a carefully designed reward system as detailed in Section III-A. The robots controlled by ISM-MARL are trained to take optimal actions, which include selecting tasks, navigating, and cleaning, in three predefined scenarios. The results indicate that ISM improves the performance of MARL quantitatively with feasible computational expenditure. The robots employing ISM effectively learned collaborative skills and in Section IV, we show that robots utilizing our ISM-Shared approach outperformed those using decentralized MARL and achieved comparable performance to robots employing centralized MARL, utilizing complete global information for decision-making during the execution.

II. Related Work

Several research endeavors are dedicated to incorporating advanced AI techniques in MAS to achieve autonomous smart environments as autonomous systems and AI progress. Deep RL techniques readily apply to modeling challenges encountered in cooperative or competitive MAS. Li et al. discussed real-world applications of RL in various fields, including robotics and transportation [6]. Kober et al. showed RL application in robotics reduces the need for specific system engineering [7]. This study focuses on developing distributed autonomous robots to optimize cooperative tasks using MARL methods within the context of smart environments, where multiple trainable agents are considered in interactions. Extensive research has been done on achieving team goals in a cooperative MAS using RL algorithms. Early works on RL methods centered around single-agent domains. Watkins and Dayan proposed Q-Learning for agents to act optimally in single-agent Markovian problems [8]. Konda et al. introduced the Actor-Critic (AC) RL algorithm, which combines the advantages of both Q-Learning and policy gradient to enhance the RL learning performance [9]. Schulman et al. proposed the Proximal Policy Optimization (PPO) algorithm, which outperforms other policy gradient approaches in several applications [10]. Yu et al. extended the PPO into Multi-Agent PPO (MAPPO), specializing in multi-agent systems [11]. This work illustrates the efficacy of a MARL framework on multiple autonomous robots performing cooperative tasks. Panait et al. offered a comprehensive overview of cooperative multi-agent learning. They discussed the challenges of tackling joint team tasks, highlighting the significance of team learning and effective communication among learning agents [2]. Li et al. utilized agents' communication in cooperative MARL. The results show that effective inter-agent communication improved the team learning outcome [3].

In our novel information-sharing approach, we focus on addressing challenges regarding team learning within the realm of MARL in smart environments. In our prior experiments, we enabled inter agent communication through Team Information Matrix that showcased performance enhancement within smart environment [12]. In this research, we present ISM, which employs an encoding technique to significantly reduce the state space's dimension and enhance the MARL learning performance. Li et al. studied cooperative MARL in partially observable settings [13] and introduced a hierarchical relation graph to enhance cooperation among agents. However, the associated computational cost of generating a hierarchical relation graph far exceeds that of our ISM calculations. Moreover, ISM's abstraction of global information, independent of environmental parameters, leads to extensive adaptability across diverse scenarios and varying agent counts. Anthony et al. have also investigated abstracted global information as a form of multi-agent influence map [14]–[16] in semi-centralized MARL settings. Their approach revolves around agents' global influence, while our approach centers on combining agents' local observation in a hierarchical ISM. The convergence of MARL and intelligent manufacturing was investigated by Agrawal et al. [17]. They proposed an RL architecture for autonomous robots where each robot communicates with a central server for both learning and execution purposes. Fragapane et al. have shown the effectiveness of smart autonomous mobile robots in the production system of the process industry [18]. Compared to the above-mentioned works, this study focuses on applying different information-sharing mechanisms to enhance the efficiency and collaboration of agents operating across diverse scenarios in multi-agent environments.

III. Methodology

In this section, we present our experimental setup constructed in *Unity* to illustrate our multi-agent scenarios and then provide details of our proposed ISM-MARL approach.

A. Experimental Setup

We formulate the experimental scenarios using Markov games, an extension of Markov Decision Processes (MDP) designed for multiple agents, as exemplified in [19], [20]. A Markov game contains a set of states of the environmental status and agents' observations, as well as a set of actions that the agents could take with immediate rewards from the environments. In all of our experiments, we employ the notation of $A_1, A_2, ..., A_N$ to represent the robots' actions, while $O_1, O_2, ..., O_N$ signifies the robots' observations. Here, N denotes the number of learning robots in a scenario. Each robot takes an action following a policy π at each environmental step

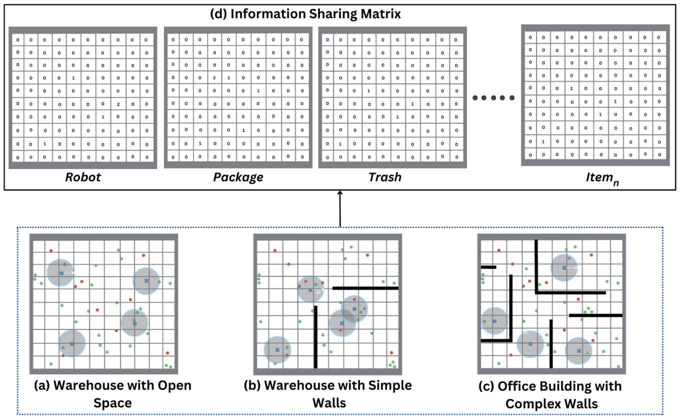


Fig. 1. ISM Generated from Multi-Agent Scenarios

and earns a reward r, where r represents the state transition from S to S' ($S \times \{A_1, ..., A_N\} \rightarrow S'$).

To evaluate the effectiveness of our ISM-MARL approach, we designed three multi-agent scenarios with various complexity levels. We opted for the *Unity ML-Agents Toolkit* as our research platform. *Unity ML-Agents Toolkit* is an open-source framework designed for creating and interacting with multi-agent simulations [21]. In these scenarios, we introduced four autonomous robots, depicted as blue cubes, to gather packages and red-colored trash items. In each scene, 25 packages and 10 trashes reappear randomly in the open space at a low speed after being collected, creating a simulation of a continuous production environment. This setup facilitates an extended training period for the multi-agent intelligent environment. The random reappearance of packages and trashes at different positions mirrors a dynamic production environment, maintaining a favorable product-to-trash ratio. The first scenario simulates an open production warehouse, as illustrated in Fig. 1a. To investigate the robustness of ISM-MARL, we added two more variations in the open production area by incorporating differing quantities of impenetrable walls. The second scenario, shown in Fig. 1b, features a warehouse scenario with open space partitioned by two walls. The third scenario, shown in Fig. 1c, comprises a set of complex walls and narrow corridors, representing layouts commonly found in office buildings.

The learning robots, represented by a cube in Fig. 1, can select a package or trash and navigate autonomously in a distributed manner with the shared objective of collecting as many packages and cleaning as much trash as possible in a given amount of time. Each robot has a sensing area of 20×20 and observes packages, trashes, robots, and walls within their sensing limit. In the simulation, the robot needs to collide with package to collect it. To simulate the cleaning task effectively and to distinguish it from the package collection task, we equipped each robot with a laser, which is capable of targeting and "cleaning" red balls and simulates a cleaning action. We implemented a reward system defined in (1):

$$R(s) = \begin{cases} +2 & \text{if collect a package} \\ +0.2 & \text{if clean a trash} \\ -2 & \text{if shoot a robot or collide a trash} \end{cases} \tag{1}$$

The reward values are selected based on our experimental results to effectively train our robots. We trained our robots using PPO and comprehensively compared our ISM-MARL approaches utilizing global and local observations.

B. *ISM-MARL Architecture*

The information-sharing MARL architecture is shown in Fig. 1. Here, we use a three-layer ISM as an input for the MARL model in our ISM-MARL architecture. The first layer represents the number of robot information, the second layer carries the information of packages, and the third layer contains the number of trashes. Although we have used only three types of information in our experiments, the ISM can be easily extended to represent more information to suit various smart environments. Based on global information availability during execution, we formulated two variations: ISM-Global and ISM-Shared. In ISM-Global, we operate under the assumption that the system has access to comprehensive environmental data, including precise object locations obtained through cameras situated in the production area. In ISM-Shared, however, the system does not need access to complete environmental information. ISM-Global and ISM-Shared are discussed in detail in Section III-B1 and Section III-B2.

Our experiments place the objects in a 100×100 unit square floor. However, incorporating all the data from a 100×100 region as input for a MARL model poses a substantial computational burden for collaborative learning. Furthermore, in practical smart environments, the operational area for robots can surpass the dimensions of our simulation, intensifying the difficulty of inputting data into the MARL model even further. Therefore, we abstracted the data to a 10×10 matrix using (2) to decrease the computational complexity.

$$ISM\_Pos = \lfloor \frac{op + \frac{1}{2} \times l}{d} \rfloor \tag{2}$$

In this equation, op represents the original coordinate of an object in a given scenario, l denotes the original scene length, given that the floor has a square shape, and d represents the new target dimension of each matrix layer of ISM. We then calculate the new position in ISM for each object and update their count in their respective two-dimensional layer. As depicted in Fig. 2, in ISM-MARL, we utilize multiple robots operating with a single shared neural network within a multi-task learning framework. This setup results in computational savings during both training and inference, as only one network needs to be evaluated. Each robot's local observation, combined with ISM, forms the input space for individual robots, guiding their collaborative task selection and navigation throughout the MARL learning process.

1) ISM-Global: In situations where complete information is accessible, we calculate ISM using the entirety of the global state, named as ISM-Global. This involves the continuous observation of the production area through a central camera unit, capturing the position and velocity of each robot, package, and trash in every frame, constituting what we term "complete global information." Here, an intermediary controller hub

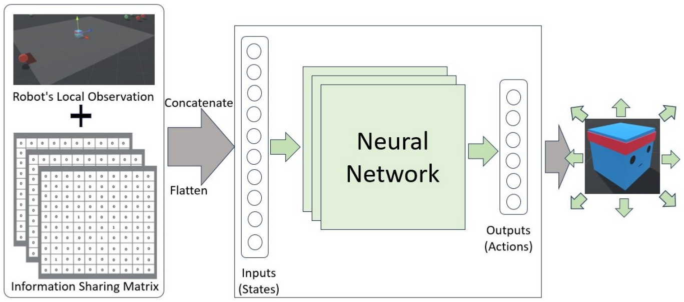


Fig. 2. ISM-MARL Robot Control Architecture

receives the real-time location data of each object and regularly updates the ISM. Each robot fetches and utilizes ISM-Global with its local information via ISM-MARL architecture to make optimal actions. ISM-Global abstracts comprehensive environmental information, facilitating the robots' cooperative learning in MAS. This abstraction notably diminishes the complexity of the extensive environmental state space, significantly reducing the processing requirements of multi-agent reinforcement learning.

2) ISM-Shared: Given the potential limitations in obtaining complete global information during deployments, particularly in settings like office buildings where implementing a centralized camera system is challenging, we developed the ISM-Shared model. This model aims to mimic complete information by establishing a shared knowledge repository for the team. In this variation, robots share and upload their individual observations to a central hub controller, where the ISM-Shared data is continuously maintained and distributed to all robots in real-time. This shared information assists them in their assigned tasks and navigation activities. To summarize, the difference between ISM-Global and ISM-shared is that, in ISM-Global, the hub controller receives all the environmental information, but in ISM-Shared, each robot communicates with the hub to share its knowledge to calculate ISM.

Algorithm 1 ISM-Shared Generation from Individual Robots

1: $ISM[][][] \leftarrow$ 3-layer ISM Initialize to 0
2: $robots[] \leftarrow$ Collect each robot's observation data
3: **for** $robot$ in $robots[]$ **do**
4: $\quad RP \leftarrow$ ISM position from $robot$.position
5: $\quad ISM[0][RP.x][RP.y]++$
6: $\quad itemList[] \leftarrow$ list of item type to track
7: $\quad$ **for** $item$ in $itemList[]$ **do**
8: $\qquad Obj[] \leftarrow$ untracked $item$ in robot's viewpoint
9: $\qquad matLayer \leftarrow$ designated matrix Layer for $item$
10: $\qquad$ **for** $object$ in $Obj[]$ **do**
11: $\qquad\quad P \leftarrow$ ISM position from $object$.position
12: $\qquad\quad ISM[matLayer][P.x][P.y]++$

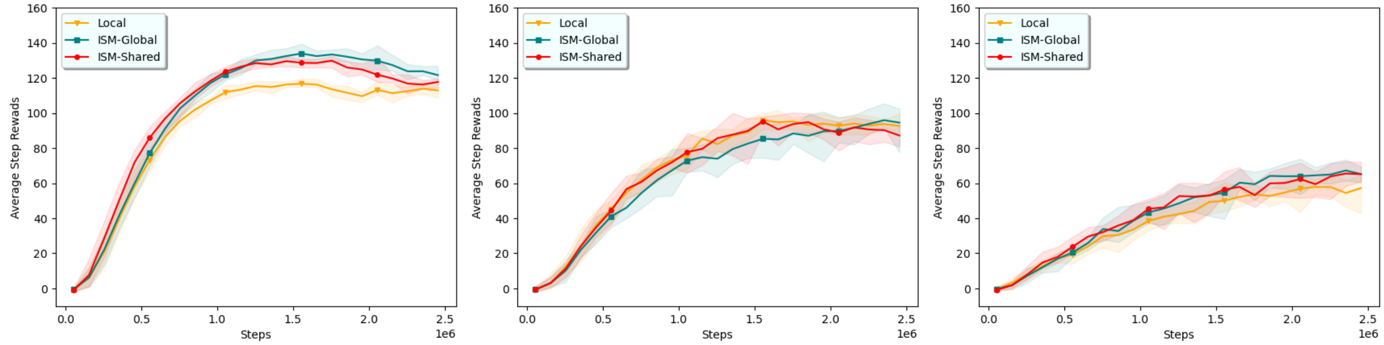
To further illustrate, in Fig. 1, the circular area surrounding each robot represents a robot's observation range. Algorithm 1 outlines the calculation process for constructing ISM-Shared. After initializing ISM for the shared knowledge among robots, we iterate over each robot to update the information in ISM. For each robot, we calculate the position of the ISM using Equation 2, as stated in line 4 of Algorithm 1. We increment the value in the position of ISM by 1. Then, in line 6, we take the type of items to track and their designated matrix layer information. In our experiments, we tracked only two types of objects, packages and trashes, and designated the second and third layers of ISM to represent them. Next, we iterate over each object from the robot's observation range and continue updating the ISM. Subsequently, if a package or trash is already being tracked and updated in the ISM by another robot, we remove the object from the robot's observation list. Next, we iterate through each package and trash, calculate their new position in the ISM using Equation 2, and increment the respective position values by 1. Hence, we can see that the ISM-Shared implementation computes the ISM exclusively from locally uploaded observations of individual robots and is more widely applicable in practice due to its independence from global information.

IV. Results and Discussion

We assessed ISM-enabled MARL performance across the scenarios representing warehouses and office buildings with varying numbers of walls described in Section III-A. Additionally, we evaluated the efficacy of our trained models across these scenarios, considering different quantities of robots. Our evaluation employed various performance metrics, including average step reward and average score. In the following subsections, we present the ISM performance for specific scenarios.

A. Performance of ISM-MARL on Open Warehouse Scenario

Initially, we examined the teamwork effectiveness of four robots using Local-MARL, ISM-Global MARL, and ISM-Shared MARL algorithms, respectively. This evaluation focused on team-based assignments, specifically tasks related to collecting packages and cleaning trash within a warehouse setting characterized by open space, as depicted in Fig. 1a. Here, Local-MARL denotes a conventional MARL approach where robots are trained solely through their local observations, without information sharing among their peers'. We trained our learning robots up to 2.5 million steps on three MARL approaches in all three predefined scenarios. As illustrated in Fig. 3a, robots operating under Local-MARL demonstrated ongoing learning and steady improvement in team performance over 2.5 million training steps. Notably, the peak average reward attained by Local-MARL reached 120 across 11 distinct runs with varied random seeds. Note that packages and trashes consistently respawn immediately after being consumed in simulated scenarios for an extended duration of MARL training and execution. In this context, the results indicate that the robots can achieve local optimal with high performance without information sharing. Next, we enabled information-sharing mechanisms among robots to further enhance the overall performance of a team. We integrated both ISM-Global with abstracted global information and local observations into the robots' decision-making process to help promote the discovery of team objectives and



(a) Avg. Reward on Warehouse Scenario with Open Space (b) Avg. Reward on Warehouse Scenario with Simple Walls (c) Avg. Reward on Office Building Scenario with Walls

Fig. 3. Results of MARL Learning across Scenarios during Training

TABLE I
MAXIMUM AVERAGE REWARD DURING TRAINING

Method	Open Space	Simple Walls	Complex Walls
Local-MARL	120	101	86
ISM-Global MARL	**139**	**105**	95
ISM-Shared MARL	132	101	**98**

global awareness. We abstracted the perfect global information in a three-dimensional matrix as described in Section III and trained our neural networks using the combination of ISM-Global and local observation in the warehouse scenario. As shown in Fig. 3a, the robots utilizing ISM-Global outperformed the Local-MARL significantly. The maximum average reward reached 139, which is 16% higher than the Local-MARL performance. This indicates that the ISM derived from global information is certainly enhancing the robots' ability to work collaboratively in decision-making. We further evaluated the performance of robots using ISM-Shared MARL, considering the fact that complete information might not always be available in real-world applications. ISM-Shared is integrated from all the robots' local observations and used to imitate the global information for MARL learning. Fig. 3a shows our ISM-Shared robots outperformed the robots using only local observations by almost 10% and achieved a maximum average reward of approximately 132.

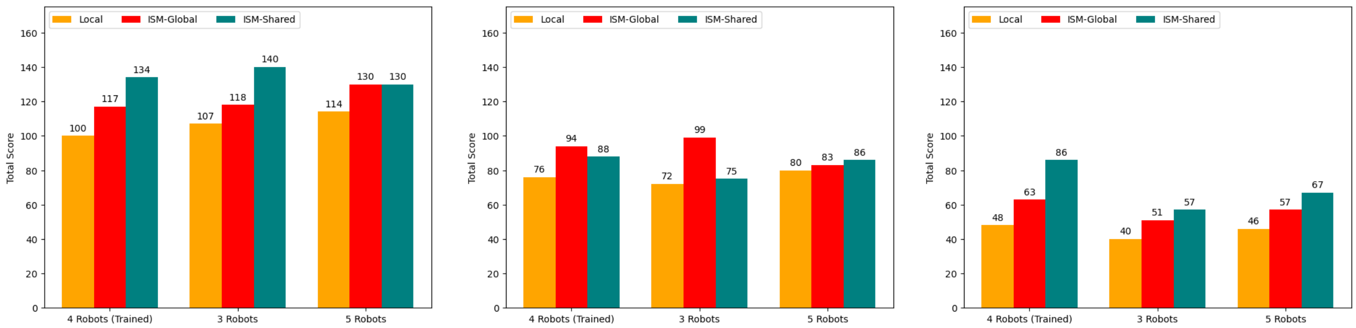
B. Performance of MARL on Scenarios with Walls

Fig. 3b shows that Local-MARL robots successfully learned to work collaboratively with each other and achieved the maximum average reward of 101 in Fig. 1b scenario featuring warehouses with open space. In contrast to the peak score of 120 achieved in the open warehouse scenario, the introduction of walls and obstacles resulted in a diminished performance. Despite this, the robots continued to acquire collaborative task-solving skills, albeit with reduced efficiency. We further evaluated the Local-MARL robots in the most complex office building scenario where the area is partitioned into various rooms and corridors by 7 distinct walls, as illustrated in Fig. 1c. As depicted in Fig. 3c, the robots exhibited progressive performance enhancement over the training period and achieved a peak average reward of 86, nearly 72% of the reward obtained in the warehouse scenario with open space.

TABLE II
GENERALIZABILITY ON VARYING NUMBER OF ROBOTS (50,000 STEPS)

Team	Scenario	Method	Package	Trash	inc (%)
3 Robots (Unseen)	Open Warehouse	Local	164.60	**3.80**	-
		ISM-Global	181.20	0.00	9.91
		ISM-Shared	**213.00**	3.60	**30.34**
	Simple Walls	Local	110.30	1.40	-
		ISM-Global	**151.40**	0.20	**37.96**
		ISM-Shared	113.40	**2.40**	4.60
	Office Buildings	Local	60.60	**2.40**	-
		ISM-Global	80.00	0.20	29.16
		ISM-Shared	**85.00**	5.40	**44.16**
5 Robots (Unseen)	Open Warehouse	Local	290.40	**3.40**	-
		ISM-Global	**335.40**	0.00	**14.56**
		ISM-Shared	232.71	1.86	14.04
	Simple Walls	Local	206.20	0.80	-
		ISM-Global	209.80	0.40	2.98
		ISM-Shared	**220.00**	**5.20**	**7.44**
	Office Buildings	Local	116.60	4.40	-
		ISM-Global	146.80	1.20	22.61
		ISM-Shared	**165.00**	**13.6**	**46.96**

As Local-MARL performance deteriorated with the increasing complexity of the wall, we wanted to explore how ISM-Global and ISM-Shared perform in comparison. Here, the robots utilizing ISM-Global achieve a maximum average reward of 105 which is 4% higher than Local-MARL. In the office building scenario with complex walls, the maximum average reward is 78 which is 10% greater than the local observation. For the robots with ISM-Shared MARL, in the warehouse scenario with simple walls, there was minimal variation between Local-MARL and ISM-Global MARL. Nonetheless, the robots utilizing ISM-Shared MARL demonstrated notable enhancement in navigating the intricate walls of the office building scenario, surpassing the performance of Local-MARL and nearly matching the effectiveness of the ISM-Global approach. Our ISM-Shared approach outperformed local observation by 14% and reached the maximum average reward of 98. Table I elaborates the comparison of maximum average reward across all three scenarios during the training process and we can see that the ISM-Global approach is best performed across all three scenarios in training. Nonetheless, the ISM-Shared approach outperforms local-MARL greatly and shows a very close performance compared to ISM-Global.



(a) Avg. Score on Warehouse Scenario with Open Space (b) Avg. Score on Warehouse Scenario with Simple Walls (c) Avg. Score on Office Building Scenario with Walls

Fig. 4. Evaluation Results of Average Score attained per Robot utilizing MARL Methods across Scenarios with Varying Numbers of Co-operative Robots

C. Evaluation of the Best Performed Robots

After the training phase of the four-robot team across scenarios with three different difficulty levels, we chose the top-performing neural network models utilizing Local-MARL, ISM-Shared, and ISM-Global MARL in all three scenarios for assessment. Furthermore, to evaluate the generalizability of learned models, we examined robots trained with scenarios involving four robots and tested them with varying numbers of robots, gauging the adaptability and generalizability of a robot operating with an unseen team size. Table II detailed the results collected on each MARL method undergoing 50,000 steps. For each combination of scenario, method, and team size, we observed the average number of packages collected, trash cleaned, cumulative team score, and the percentage increase in score from Local-MARL to ISM-MARL approaches. In the open warehouse scenario as shown in Fig. 4a, each robot achieved an average score of 100 when collaborating with a team of four robots. However, robots utilizing ISM-Global and ISM-Shared achieved average scores of 117 and 134, respectively, representing a 25% improvement in performance on average with ISM-MARL approaches. Similarly, in the scenario with simple walls described in Fig. 4b, robots trained with Local-MARL in four robot teams attained 76 for each robot, whereas those utilizing ISM-Global and ISM-Shared achieved scores of 94 and 88, respectively, reflecting a 20% performance improvement. Furthermore, in the office building scenario as illustrated in Fig. 4c, Local-MARL four robot teams scored 48 points for each robot, while ISM-Global and ISM-Shared robots achieved 63 and 86 points, respectively. This represents a 48% improvement in performance with ISM-MARL approaches over Local-MARL. Overall, our ISM-MARL outperformed robots with only local observations by 26% on average. In addition, the ISM-Shared MARL model showcased superior performance over ISM-Global in most experimental setups and demonstrated comparable performance in other cases. Given that ISM-Shared operates independently of complete information, the evaluation results suggest its adaptability in diverse MASs. Fig. 4 also demonstrates that, despite being trained to collaborate within a team consisting of only 4 robots, our ISM-MARL methods consistently outperform when applied to smaller teams, such as 3 robots, and larger teams, such as those comprising 5 robots. In the open warehouse scenario with a team of three robots, the ISM-Global method surpassed the baseline local MARL by 9.91%, while the ISM-Shared method outperformed the baseline by 30.34%. Similarly, the robot team consistently outperformed the baseline in the other two scenarios by an average of 13% and 36%, respectively. Furthermore, our experimental results with a team of five robots also exhibit a similar trend. As shown in Table II, ISM-MARL methods outperform the baseline in the open warehouse, simple walls, and office buildings scenarios by an average of 14%, 5%, and 32%, respectively. The evaluation result suggests that our ISM representation is scenario-independent and capable of adapting to function effectively across various team sizes. This highlights the capability of ISM-MARL approaches to maintain high-performance levels despite fluctuations in the workforce. In real-world production scenarios, where robots may be added or removed for various reasons, a MARL approach necessitating additional training each time the environment setting slightly changes would incur production overhead. Thus our robust ISM-MARL approach can effectively improve the overall performance with fewer training initiatives.

V. Conclusion and Future Work

This paper presents a novel Information Sharing Matrix (ISM) designed to facilitate inter-agent communication. We integrate this matrix with local observations to efficiently guide robots in learning collaborative tasks within intricate multi-agent smart environments. We conducted experiments on diverse multi-agent scenarios of varying difficulty levels and experimental results demonstrated that our robots successfully collaborated in both simple and complex scenarios. Moreover, the ISM-enhanced MARL consistently exhibited performance improvements and outperformed local models across teams comprising different numbers of robots. In the future, we plan to include transfer learning to evaluate the performance of our trained models with new space information and investigate the efficacy of ISM in maximizing team objectives to promote the generalization and robustness of MARL. Furthermore, we intend to investigate the role of human intervention during the training of collaborative robots and its influence on performance outcomes. Additionally, we plan to investigate the outcome of our MARL model in dynamic training environments and assess the efficacy of the training robots

VI. Acknowledgement

This material is based upon work supported by the National Science Foundation under Award No. 2302060.

References

[1] L. Buşoniu, R. Babuška, and B. De Schutter, *Multi-agent Reinforcement Learning: An Overview*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 183–221.

[2] L. Panait and S. Luke, "Cooperative multi-agent learning: The state of the art," *Autonomous Agents and Multi-Agent Systems*, vol. 11, no. 3, pp. 387–434, Nov 2005.

[3] D. Li, Z. Xu, B. Zhang, G. Zhou, Z. Zhang, and G. Fan, "From explicit communication to tacit cooperation: A novel paradigm for cooperative marl," in *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, ser. AAMAS '24. Richland, SC: International Foundation for Autonomous Agents and Multiagent Systems, 2024, p. 2360–2362.

[4] S. Sukhbaatar, R. Fergus *et al.*, "Learning multiagent communication with backpropagation," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[5] C. Zhang and V. Lesser, "Coordinating multi-agent reinforcement learning with limited communication," in *Proceedings of the 2013 international conference on Autonomous agents and multi-agent systems*, 2013, pp. 1101–1108.

[6] Y. Li, "Reinforcement learning applications," *arXiv preprint arXiv:1908.06973*, 2019.

[7] J. Kober, J. A. Bagnell, and J. Peters, "Reinforcement learning in robotics: A survey," *The International Journal of Robotics Research*, vol. 32, no. 11, pp. 1238–1274, 2013.

[8] C. J. C. H. Watkins and P. Dayan, "Q-learning," *Machine Learning*, vol. 8, no. 3, pp. 279–292, 1992.

[9] V. Konda and J. Tsitsiklis, "Actor-critic algorithms," *Advances in Neural Information Processing Systems*, vol. 12, 1999.

[10] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv:1707.06347*, 2017.

[11] C. Yu *et al.*, "The surprising effectiveness of ppo in cooperative multi-agent games," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 611–24 624, 2022.

[12] A. Siddiqua, S. Liu, R. Iqbal, F. A. Irfan, L. Ross, and B. Zweerink, "Tim-marl: Information sharing for multi-agent reinforcement learning in smart environments," in *2024 IEEE 21st Consumer Communications & Networking Conference (CCNC)*. IEEE, 2024, pp. 1044–1045.

[13] Y. Li, X. Wang, J. Wang, W. Wang, X. Luo, and S. Xie, "Cooperative multi-agent reinforcement learning with hierarchical relation graph under partial observability," *International Conference on Tools with Artificial Intelligence*, pp. 1–8, 2020.

[14] A. Harris and S. Liu, "Maidrl: Semi-centralized multi-agent reinforcement learning using agent influence," *IEEE Conference on Games*, pp. 01–08, 2021.

[15] A. S. Nipu, S. Liu, and A. Harris, "Maidcrl: Semi-centralized multi-agent influence dense-cnn reinforcement learning," *IEEE Conference on Games*, pp. 512–515, 2022.

[16] P. Brackett, S. Liu, and Y. Liu, "Sc-mairl: Semi-centralized multi-agent imitation reinforcement learning," *IEEE Access*, vol. 11, pp. 57 965–57 976, 2023.

[17] A. Agrawal, S. J. Won, T. Sharma, M. Deshpande, and C. McComb, "A multi-agent reinforcement learning framework for intelligent manufacturing with autonomous mobile robots," *Proceedings of the Design Society*, vol. 1, p. 161–170, 2021.

[18] G. Fragapane, D. Ivanov, M. Peron, F. Sgarbossa, and J. O. Strandhagen, "Increasing flexibility and productivity in industry 4.0 production networks with autonomous mobile robots and smart intralogistics," *Annals of Operations Research*, vol. 308, no. 1, pp. 125–143, Jan 2022.

[19] R. Lowe, Y. I. Wu, A. Tamar, J. Harb, O. Pieter Abbeel, and I. Mordatch, "Multi-agent actor-critic for mixed cooperative-competitive environments," *Advances in neural information processing systems*, vol. 30, 2017.

[20] J. Foerster, G. Farquhar, T. Afouras, N. Nardelli, and S. Whiteson, "Counterfactual multi-agent policy gradients," *AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[21] A. Juliani *et al.*, "Unity: A general platform for intelligent agents," *arXiv preprint arXiv:1809.02627*, 2020. [Online]. Available: https://arxiv.org/pdf/1809.02627.pdf