AI

MDPI

Article

Development of an Attention Mechanism for Task-Adaptive Heterogeneous Robot Teaming

Yibei Guo, Chao Huang and Rui Liu *

Cognitive Robotics and AI Laboratory (CRAI), College of Aeronautics and Engineering, Kent State University, Kent, OH 44240, USA; yguo27@kent.edu (Y.G.); chuang26@kent.edu (C.H.)
* Correspondence: ruiliu.robotics@gmail.com

Abstract: The allure of team scale and functional diversity has led to the promising adoption of heterogeneous multi-robot systems (HMRS) in complex, large-scale operations such as disaster search and rescue, site surveillance, and social security. These systems, which coordinate multiple robots of varying functions and quantities, face the significant challenge of accurately assembling robot teams that meet the dynamic needs of tasks with respect to size and functionality, all while maintaining minimal resource expenditure. This paper introduces a pioneering adaptive cooperation method named inner attention (***innerATT***), crafted to dynamically configure teams of heterogeneous robots in response to evolving task types and environmental conditions. The ***innerATT*** method is articulated through the integration of an innovative attention mechanism within a multi-agent actor–critic reinforcement learning framework, enabling the strategic analysis of robot capabilities to efficiently form teams that fulfill specific task demands. To demonstrate the efficacy of ***innerATT*** in facilitating cooperation, experimental scenarios encompassing variations in task type ("Single Task", "Double Task", and "Mixed Task") and robot availability are constructed under the themes of "task variety" and "robot availability variety." The findings affirm that ***innerATT*** significantly enhances flexible cooperation, diminishes resource usage, and bolsters robustness in task fulfillment.

Keywords: inner attention; multi-agent reinforcement learning; adaptive cooperation; heterogeneous multi-robot team

Citation: Guo, Y.; Huang, C.; Liu, R. Development of an Attention Mechanism for Task-Adaptive Heterogeneous Robot Teaming. *AI* **2024**, *5*, 555–575. https://doi.org/10.3390/ai5020029

Academic Editor: Demos T. Tsahalis

Received: 13 February 2024
Revised: 15 March 2024
Accepted: 22 April 2024
Published: 23 April 2024

1. Introduction

A heterogeneous multi-robot system (HMRS) encompasses an assembly of robots differing in shape, size, and functionality, collaborating to achieve collective goals. Owing to its functional diversity, scalability in team size, and enhanced control resilience, HMRS finds extensive applications in executing tasks on a grand scale. Notably, in the context of search and rescue operations following natural disasters [1–3], HMRS's ability to conduct parallel operations allows for the expansive coverage of surveillance areas and the efficient rescue of numerous victims. Similarly, in the realm of traffic management, including the regulation of traffic flows and the scheduling of public transport systems [4–6], deploying a multitude of specialized robots to form a cohesive team proves to be more cost effective than relying on a single robot endowed with multiple capabilities for the entire operation. Additionally, when addressing tasks that require comprehensive area coverage and the navigation of complex search missions [7–9], HMRS demonstrates its capacity to mitigate task intricacies by distributing responsibilities among the team members.

However, the integration of heterogeneous robots into effective teams for real-world applications is significantly hampered by the diversity of tasks. Initially, the fluctuating demands of tasks across various environments complicate the determination of the necessary type and scale of assistance [10–13]. Furthermore, the challenge extends to precisely aligning the capabilities of a robot team with the diverse requirements of tasks [14]. Tasks distributed across different locations and times introduce variability in requirements,

with even identical work areas experiencing dynamic changes in task demands over time. For instance, traffic management scenarios demonstrate variability in traffic flow between urban and rural areas and fluctuations at the same intersection over time, necessitating teams with adaptable and complementary skills for effective cooperation. The dynamic and evolving nature of these requirements complicates the formation of teams that are adequately matched in both type and scale to the tasks at hand, thereby posing a challenge to flexible team composition [15].

In addition, the operational readiness of robots to contribute to tasks is affected by real-world issues such as motor wear, sensor malfunctions, and the overall working status of the robots [16–18]. Robots experiencing faults may disseminate inaccurate data within the team, compromising the team's ability to fulfill its tasks effectively. The unpredictability and detrimental effects of such faults constrain the selection of competent team members, thus hindering the formation of a capable and appropriately sized robot team to meet the tasks' demands.

Lastly, the task of aligning robot team capabilities with task requirements is further complicated by the dynamic nature of task needs, robot availability, and environmental constraints [14]. The variability of assistance needs due to task changes, the impact of obstacles and weather conditions on the timing and viability of robot contributions, and the varying proximities of available robots to the required locations underscore the complexities involved. The disparity between task varieties and the capabilities of robot teams undermines the effective deployment of HMRS in practical scenarios. Overlooking these diversities in real-life situations can detrimentally affect the performance of HMRS and its precise alignment with task necessities, thereby significantly restricting the practical application of HMRS.

Hence, there exists a critical necessity to flexibly assemble teams of heterogeneous robots that can adeptly meet task requirements and optimally leverage robot capabilities, addressing the aforementioned challenges.

This research addresses the highlighted challenges by proposing a novel method for flexible robot teaming termed inner attention (***inner-ATT***). This method is realized through the incorporation of an innovative attention mechanism within a multi-agent actor–critic reinforcement learning framework, as illustrated in Figure 1. The ***innerATT*** mechanism empowers a robot to focus on communications with its available teammates, thereby recognizing and integrating cooperative factors essential for team formation; this enables robots to selectively form teams that are adaptive to the dynamics of the environment. The attention mechanism central to ***innerATT*** is refined and perfected through deployment training. The contributions of this paper are threefold:

- A novel multi-robot teaming method, ***innerATT***, is developed to guide the flexible cooperation among heterogeneous robots as the task complexity varies in target number, target type, and robot work status.
- A theoretical analysis is conducted to validate the robustness of ***innerATT*** in guiding flexible cooperation, providing a theoretical foundation for implementing ***innerATT*** in general disturbance-involved multi-robot teaming in future similar research.
- A deep reinforcement learning-based simulation framework, which integrates the simulation platform of a multi-agent particle environment, the multi-agent deep reinforcement learning algorithms, and robot models, is developed to provide a standard pipeline for simulating flexible robot teaming.

This paper is organized into the following sections to explore the development and implications of the inner attention (innerATT) method. Following the introduction, Section 2 reviews existing methodologies and highlights the gap innerATT aims to fill. Section 3 details the theoretical foundation and implementation of innerATT and the experimental design. Section 4 meticulously presents and analyzes the performance of innerATT under various scenarios, including task variety and robot availability. Section 5 synthesizes the findings, discusses their practical implications, and suggests directions for future research.

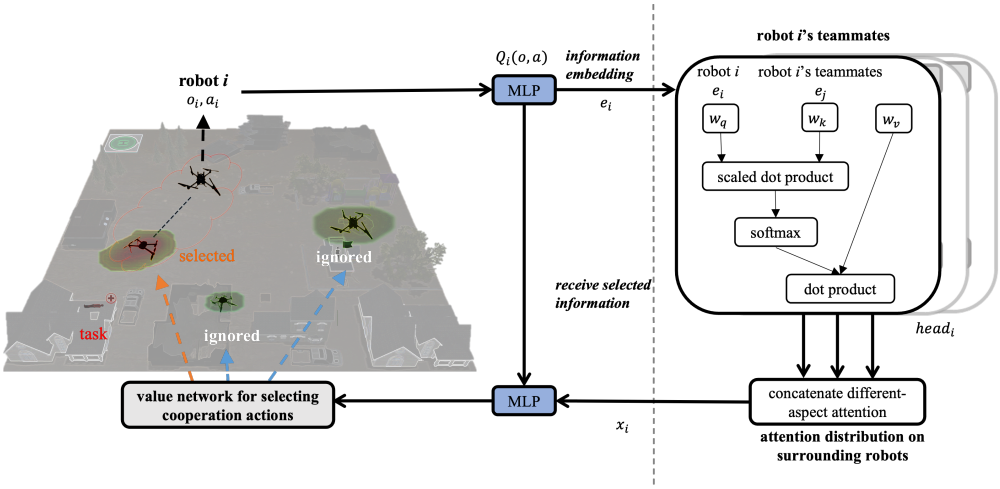

Figure 1. The architecture of ***innerATT***. The inner attention mechanism determines the attention weights between robots. As the left figure shows, the input is a robot's observations, status, actions, and cooperability-related information from the robot's teammates. The output is the Q-value network for cooperation strategy selection, in which a robot pays different-level attention to others to form a team for a given task. In the right figure, multiple attention heads are used to evaluate different aspects of the cooperability between a robot and its potential teammates.

2. Related Work

To flexibly compose heterogeneous robot teams, much research was conducted to optimize task allocation among robot members inside a team. From early work on centralized and homogeneous robot systems to more recent work on decentralized and heterogeneous robot systems, various kinds of algorithms have been proposed to increase overall task performance by translating the multi-robot task allocation problem into an optimization problem. In [19–23], human pre-defined robot utility functions, including flexible teaming, time efficiency, and resource consumption, and intelligent optimization algorithms, including mixed-integer optimization, genetic algorithm, ant colony algorithm, and particle swarm algorithm were used to enable robot teams to perform tasks that require cooperation. During optimization, the utility function was maximized, and robots were assigned different sub-tasks based on robot capability. However, the pre-optimized task allocation strategy can represent general tasks with low applicability, especially for applications with dynamic task requirements, as these approaches deal primarily with a task allocation problem based on fixed task requirements. When new tasks are presented or the task location changes, a revised optimization process is required to reach a flexible teaming strategy based on the refreshed task requirements. Such a revision is difficult to achieve in order to satisfy the needs of real-time task executions. To overcome the above-mentioned limits, a reinforcement learning algorithm has been used to enable robot performance in a dynamic environment derived from experience in maximizing or minimizing human-designed utility functions equivalently. In [24,25], a Q-learning algorithm was used to discern optimal policy from which robots select the optimal strategy based on environmental adaptation and team capability. Refs. [26–29] proved deep reinforcement learning to be effective in enabling sophisticated behaviors of individual robots in dynamic environments. Although reinforcement learning-based methods can perform real-time robot guidance based on current observations, including robot status, task type, and location, these approaches do not correct for sensor and robot failures. The ***innerATT***, multi-robot task allocation method based on a deep reinforcement learning algorithm, is resilient to these failures by selectively recruiting functional robot team members and isolating failed robots from the team. More importantly, the ***innerATT*** can also adapt to different task complexities and real-world disturbances without retraining, which is beneficial to real-world applications.

To reduce the influence of robot failures on HMRS performance, research in HMRS self-healing has been conducted. Refs. [30–32] investigated methods for mobile robot

networks to maintain the logical and physical topology of the network when robots fail and must be replaced within a formation. They further demonstrate the stability of motion synchronization under their topological repair mechanism. However, these research works mainly focused on replacing broken robots, which ignores the danger of partial failures likely to be encountered in real-world deployments. Recently, Refs. [33–37] limited the negative influence of partial robot failures on the HMRS team by protecting the swarm through resilience by restricting robot updates to values of neighbors near their own. Their results for swarms meeting connectivity requirements and based on communication of constant or time-varying values by faulty robots showed convergence of the swarm to correct headings. However, such passive strategies usually require high robot connectivity and specification of tolerable values that are difficult to qualify in advance. Inspired by [37], the negative influence was limited by decreasing the communication quality between the failed robots and other robots. The multi-robot teaming method increased HMRS team resilience based on inner attention mechanism, which can selectively attend to robot communication connectivity. In addition, the attention weights used in ***innerATT*** can be automatically obtained, releasing humans from the burden of monitoring robot behaviors and assigning corresponding weights to their communication connectivity.

The attention mechanism remains associated with the selection of stimulus or response to processes. In cognitive theory, the attention mechanism plays a critical role in the capacity to choose task-relevant versus task-irrelevant information [38]. The attention mechanism is used in both daily and industrial scenarios to increase robot execution efficiency and safety [39,40]. Social robots use human-like gestures in a conversation to increase human engagement by paying attention to the human gaze, facial expressions, and behaviors [41–43]. In situations (specific) with limited communication channels, an attention mechanism is used to precisely determine whether the communication is necessary or not by assigning different weights to various related factors [44–46]. Inspired by the benefits of the cognitive attention mechanism, in this work, a multi-head attention mechanism [47] is used to weigh information values differently and selectively discourage low-quality communications among robots, which is similar to [37]. Finally, robot behaviors with minimum requirements on human cognitive load are identified. This research develops a novel attention-based multi-robot teaming method, ***innerATT***, based on a multi-head attention mechanism with an actor–critic multi-agent deep reinforcement learning framework. With attention-based teaming capability, ***innerATT*** supports robot team adaptation to dynamic task requirements and variable teammate availability by selectively selecting compatible teammates.

3. Materials and Methods

3.1. Inner Attention-Supported Adaptive Cooperation

When task requirements and real-world situations change, robots are expected to flexibly select teammates to satisfy task requirements and effectively utilize robot capabilities. The ***innerATT*** helps a robot in a team selectively pay different attention to different robots by using the inner attention mechanism. As shown in Figure 1, given the inputs of all statuses and observations of the robot, the inner attention mechanism automatically determines the amount of attention to different robots.

3.1.1. Heterogeneous Teaming Supported by Multi-Agent Reinforcement Learning with Centralized Training and Decentralized Execution

The basic robot teaming framework is supported by a multi-agent actor–critic deep reinforcement learning (MAAC) algorithm [48], which has an advantage of modeling entangled decision-making of multiple members in a heterogeneous robot team. MAAC has been proven to be effective in guiding the dynamic cooperation of the multi-robot [49,50]. In this paper, deep reinforcement learning is defined by the number of robots, N; state space, S; a set of actions for all robots, $A = \{A_1, \ldots A_N\}$; transition probability function over the next possible states, T: $S \times A_1 \times \ldots \times A_N \rightarrow P(S)$; a set of observations for all

robots, $O = \{O_1, \ldots O_N\}$; and reward function for each robot R_i: $S \times A_1 \times \ldots \times A_N \rightarrow R$. The application scenario of multi-robot cooperation in this paper is designed as a fully observable environment in which each robot i receives an observation, O_i, which is a simplified communication method for robots to exchange location information, share and allocate task goals, and maintain connectivity. The discrete action space includes moving up, down, left, and right at each time step. The observation in this paper includes positions of obstacles, victims, and robots; the injury level of victims and capacities of robots; and the speed and acceleration of robots. By using reinforcement learning for guiding the cooperation, each robot learns an individual policy function, π_i: $O_i \rightarrow P(A_i)$, which is a probability distribution on potential cooperation actions. The goal of multi-agent reinforcement learning is to learn an optimal cooperation strategy for each robot which can maximize their expected discounted returns:

$$J_i(\pi_i) = E_{a_* \sim \pi_*; s \sim T}[\sum_{t=0}^{\infty} \gamma^t r_{it}(s_t, a_{1t}, \ldots, a_{Nt})] \quad (1)$$

where $J_i(\pi_i)$ represents the expected cumulative rewards for robot i following policy π_i, which maps observations to actions. Actions a^* are based on the combined policies π^* of all robots, with $s \sim T$ indicating state transitions as per the environment dynamics. The discount factor γ prioritizes immediate versus future rewards, guiding strategic balance. The reward function r_i assesses the immediate utility of actions by all robots in the current state s_t, pivotal for optimizing collaborative strategies in heterogeneous multi-robot systems. Here, * represents $\{1, \ldots N\}$; $\gamma \in [0, 1]$ is the discount factor that determines the degree to which the policy favors immediate reward over long-term gain.

The actor-critical policy gradient algorithm is a learning process to solve reinforcement learning problems, which targets modeling and optimizing the policy directly. To maximally improve team performance given the current status of all robots, a robot's policy is updated by encouraging updating along the gradient:

$$\bigtriangledown_\theta J(\pi_\theta) = \bigtriangledown_\theta log(\pi_\theta(a_t \mid s_t)) Q_\psi(s_t; a_t) \quad (2)$$

where θ denotes the parameters for the policy, $\log(\pi_\theta(a_t|s_t))$ emphasizes the likelihood of selecting action a_t under policy π_θ, reinforcing effective actions, and Q is an approximation function of the expected discounted returns, estimating the total expected rewards from taking action a_t in state s_t and following the policy thereafter:

$$Q_\psi(s_t; a_t) = E[\sum_{t'=t}^{\infty} \gamma^{t'-t} r_{t'}(s_{t'}, a_{t'})] \quad (3)$$

It can ameliorate policy gradient methods' high variance issue by replacing the original return term in the policy gradient estimator. For each cooperation step, the action value Q for the robot i needs to observe its neighbors' status o and actions a and learned by off-policy temporal difference learning by minimizing the regression loss:

$$L_Q(\psi) = E_{(s,a,r,s') \sim D}[(Q_\psi(s, a) - y)^2] \quad (4)$$

where $y = r(s, a) + \gamma E_{a' \sim \pi(s')}[Q_{\bar{\psi}}(s', a')]$, $Q_{\bar{\psi}}$ is the target Q-value function, which is simply an exponential moving average of the past Q-functions, D is the experience replay buffer, which stores the previous robot cooperation experience to further reduce the loss, and $\bar{\psi}$ and ψ denote the parameters for the target critics and critics, respectively.

In the scenario of multi-robot cooperation, each robot's environment is non-stationary, with the dynamics in teammates, task requests, and environmental conditions. This non-stationary environment challenges the performance of the actor–critic reinforcement learning algorithm in both learning stability and past experience exploitation. Therefore, in this paper, an extended actor–critic framework is used to train all the robots in a centralized

way and to support an individualized cooperation strategy in a distributed way. More concretely, in the gradient of the expected return for robot i, the $Q_{\pi_i}(x; a_1, \ldots, a_N)$ is calculated centrally with a global objective of improving the whole team's performance by taking the actions of all robots as input, in addition to the robots' statues x, and then outputted the Q-value for robot i. In the simplest case, x could consist of the observations of all robots, $x = (o_1, \ldots, o_N)$. However, additional state information could also be included if available. Given that each robot may have different cooperation requirements, different teammates available, and limited perceiving capability, each robot distributively implements a cooperation policy. This centrally learned and distributively used methodology support flexible teaming for heterogeneous robots, such that the cooperators' and cooperation actions are adjusted dynamically.

3.1.2. Robot Inner Attention for Team Adaptability Modeling

In the extended actor–critic framework consisting of centralized training with decentralized execution, to calculate the Q-value function $Q_i(o, a)$ for the robot i, the critic receives observations, $o = (o_1, \ldots, o_N)$, and actions, $a = (a_1, \ldots, a_N)$, for all robots which take redundant information into account. In addition, the action space also increases exponentially with the number of robots. Given that, each robot should pay more attention to task-relevant information based on task requirements and robot availability. For example, to rescue heavily injured victims, the medical assistant robot should pay more attention to the closest and available food delivery robots. Therefore, it is necessary to train the critic for each robot with the ability to filter task-relevant information. That is, each robot is aware of which robots it should pay attention to rather than simply considering all robots at every step of decision-making. To achieve that, the inner attention mechanism is used as a complementary part of the extended actor–critic framework. Intuitively, in the robot's decision-making process, the contributions of other robots' status can be evaluated by the ***innerATT*** generating different attention weights for different robots. The more important a teammate is, the higher attention weight it should have. With the ***innerATT***, the robots can selectively cooperate with proper team members to flexibly satisfy dynamic task needs with limited team sources.

To generate the attention weights, the embedding function g_i is a two-layer multiple-layer perception (MLP), which takes robots' observations and actions as input. The embedded information is fed into the ***innerATT*** to obtain the Q-value function $Q_i(o; a)$ for robot i, which is a function of robot i's embeddings as well as other robots' contributions:

$$Q_i(o; a) = w^{2T}\sigma(w^1, < e_i, x_i >) \tag{5}$$

where σ is rectified linear units (ReLU), and w^1 and w^2 are the parameters of critics. Similar to the query-key system, the inner attention mechanism also has shared query (w_q), key (w_k), and value (w_v) matrixes. $\langle e_i, x_i \rangle$ represents the interaction between the robot's embedding e_i and the context x_i, encapsulating the robot's perceived environment and its own state. Each agent's embedding e_i can be linearly transformed into q_i, k_i, and v_i separately. The contribution from other robots, x_i, is a weighted sum of other robots' values:

$$x_i = \sum_{j \neq i} \alpha_{ij} v_j \tag{6}$$

where v_j symbolizes the value vector of robot j, the attention weight α_{ij} compares the similarity between k_j and q_i, and the similarity value can be obtained from a softmax function:

$$\alpha_{ij} = \frac{exp(S_{ij})}{\sum_{k=1}^{N} exp(S_{ik})} = \frac{exp(e_j w_k^T w_q e_i)}{\sum_{k=1}^{N} exp(e_k w_k^T w_q e_i)} \tag{7}$$

where the similarity score S_{ij} compares robot i's query with robot j's key, influencing the amount of attention robot i pays to robot j. In the experiments, P set of parameters

$(w_q^p, w_k^p, w_v^p)_{p=1}^P$, which gives rise to an aggregated contribution from all other robots to the robot i, is used. Then, the contributions from all set parameters can be simply concatenated as a single vector. Note that the matrix for extracting queries, keys, and values is shared across all agents, which encourages a common embedding space. The sharing of critical parameters between robots is possible because multi-robot value-function approximation is, essentially, a multi-task regression problem.

As for the reward function that encourages the robots to cooperate in dynamic environments, in the learning process, the corresponding reward based on their behavior is given. At the time step t, the robot obtains its observation o_t and the contribution from other robots x_t. The robot is likely to execute the action with the highest reward. To describe the reward function accurately, the expectations for the robots in the cooperation tasks should be introduced first. Each robot is expected to avoid collisions with other robots and obstacles in the environment and cooperate with other robots to rescue victims based on the following rules: (1) One robot can only cooperate with another proper kind of robot; (2) One robot should rescue its closest victim only if it is not occupied by other tasks.

In other words, the tasks we encourage robots to perform are rewarded positively, while behavior we wish the robots to avoid is rewarded negatively. So at time step t, each robot seeks policy $\pi(a_t \mid o_t, x_t)$ that could reach the expected goals. Reward function R_t for each robot is as follows:

$$R_t = Rewards + Collisions + Steps \tag{8}$$

Here, R_t is the combination of three aspects: rewards from interacting with the environment, collision with other robots or walls, and step cost for rescuing per victim.

$$Rewards = -\min_{j \in C} Dist(robot_i, victim_j) \tag{9}$$

which represents the robot's expected action to rescue the closest victim, cooperate with proper candidates, and C is a set of victims and robots that need $robot_i$ to rescue and cooperate separately according to expected cooperation.

$$Collisions = \begin{cases} -5 \times \sum_{wall \in Walls} I(robot_i, wall) \\ -1 \times \sum_{j \neq i} I(robot_i, robot_j) \end{cases} \tag{10}$$

implies that the robot should avoid collision with obstacles, and $I(*)$ is the indication function indicating whether $robot_i$ collides with the wall or/and other robots or not. The average $Steps$ needed for rescuing one victim are used to take resource consumption into account.

3.1.3. Theoretical Analysis of ***innerATT***'s Robustness

To clarify the theoretical foundation underlying our claims about the ***innerATT*** method's robustness, this section provides a detailed mathematical framework illustrating its resilience to component failures and sensor inaccuracies within a multi-robot teaming context.

Robustness in the context of our hybrid multi-robot system (HMRS) refers to the system's capacity to minimize the impact of incorrect or uncertain information transmitted by malfunctioning robots on the collective's flexible teaming performance. Specifically, we demonstrate that the attention weights calculated by operational robots remain substantially unaffected by a malfunctioning robot. Similar to the critic neural network, a two-layer ReLU neural network is considered to analyze the robustness of ***innerATT***. The weights of the first layer can be denoted as w^1, the weights of the second layer as w^2, and the ReLU function is represented by $\sigma(*)$. Then, the output of the two-layer neural network, when the input is x, can be written as:

$$f(x) = w^{2T}\sigma(w^1, x), x =< e_i, x_i > \tag{11}$$

When employing the inner attention mechanism, the robots' robustness to failures or sensor malfunctions are significantly enhanced [51].

We consider that a small perturbation is added to a particular robot $\bar{j}$'s embedding, such that $e_{\bar{j}}$ is changed to $e_{\bar{j}} + \triangle e$ while all the other robots' embeddings remain unchanged. How much will this perturbation affect attention weights α_{ij}? For a particular $i (i \neq j)$, the

$$S_{ij} = e_j w_k^T w_q e_i \tag{12}$$

is only changed by one term since:

$$S'_{ij} = \begin{cases} S_{ij} + \triangle e w_k^T w_q e_i, & if (j = \bar{j}). \\ S_{ij}, & otherwise. \end{cases} \tag{13}$$

where S'_{ij} denotes the value after the perturbation. Therefore, with the perturbed input, each set of $\{S_{ij}\}_{j=1}^{N}$ only has one term being changed for the perturbation part. Obviously, if there are two or more broken robots, the number of terms changed in each set of $\{S_{ij}\}_{j=1}^{N}$ is the same as that of broken robots in HMRS. We assume $\| \triangle e \| \leq \delta_1$, $\|e_i\| \leq \delta_2$, and $\{e_i\}_{i=1}^{N}$ are d-dimensional vectors uniformly distributed on a sphere. The value $E[S'_{i\bar{j}} - S_{i\bar{j}}] = E[Me_i]$ where $M = \triangle e w_k^T w_q$ is a fixed vector, and it is easy to derive that $\|M\| \leq \|w_q\| \|w_k\| \delta_1$. Due to rotation invariance and $\|e_i\| \leq \delta_2$, $E[e_i^T [1, 0, \ldots, 0]] \leq \frac{\delta_2}{\sqrt{d}}$. If we build the orthogonal coordinate system based on the direction of M, the expected value becomes

$$E[S'_{ij} - S_{ij}] \leq \frac{\|w_q\| \|w_k\| \delta_1 \delta_2}{\sqrt{d}} \tag{14}$$

Therefore, as the norm of w_q, w_k values are not too large (usually regularized by L_2 during training) and dimension d is large enough, there is a significant amount of i such that S'_{ij} is perturbed negligibly. This theoretical exploration, rooted in the mathematical properties of high-dimensional spaces and the controlled norm sizes of the network's weights, decisively supports the claim of ***innerATT***'s robustness. That means the flexible teaming performance of the robots in good condition is not affected dramatically by the wrong information delivered by the broken robot.

3.2. Experiment Settings

To validate ***innerATT***'s effectiveness in improving HMRS adaptability, a cooperative environment with two typical scenarios, "task variety" and "robot availability variety", and three different situations with different task complexities are designed. These two scenarios and three situations frequently occur and generally represent the deployment dynamics of HMRS in the real world. Therefore, by validating ***innerATT*** effectiveness in these two scenarios, we hope to obtain a general conclusion on the efficiency of ***innerATT*** in improving HMRS adaptability. Figure 2 illustrates the robot task scenario.

The experimental setting is depicted in Figure 3. In flood disasters, there are trapped victims with different injury levels. The victims with high injury levels (Task 1) need rescuing robots providing them with food, water, and emergency medical treatment, while the victims with low injury levels (Task 2) need other kinds of rescuing robots providing food, water, and useful information to guide them to safer places. The main robots team is expected to split into different sub-teams that can rescue these victims effectively. The environment leverages the multi-agent particle environment (MPE) framework, an open-source platform [48]. MPE is characterized by a straightforward multi-agent domain where agents navigate a continuous observation space with discrete actions, supplemented by basic simulated physics principles. This framework is advantageous for constructing experimental scenarios that involve multiple robots, intricate situations, and varied interactions among robots while simplifying aspects related to control and perception. In the custom-designed

environment, the implementation of discrete action spaces and a rudimentary physics engine facilitates the simulation of robots' movements, incorporating robot momentum to mirror real-world dynamics closely. The dimensions of the synthetic environment are established at 2×2, which suffices for the number of robots necessary to evaluate ***innerATT*** without leading to insufficient exploration issues. Robot parameters, as delineated in Table 1, are calibrated to mirror the specifications of actual robots. The environment is configured as a continuous space, allowing for robots to traverse any location on the map based on their velocity and acceleration attributes. The overarching objective of the system is to maximize the number of victims rescued within a specified timeframe.

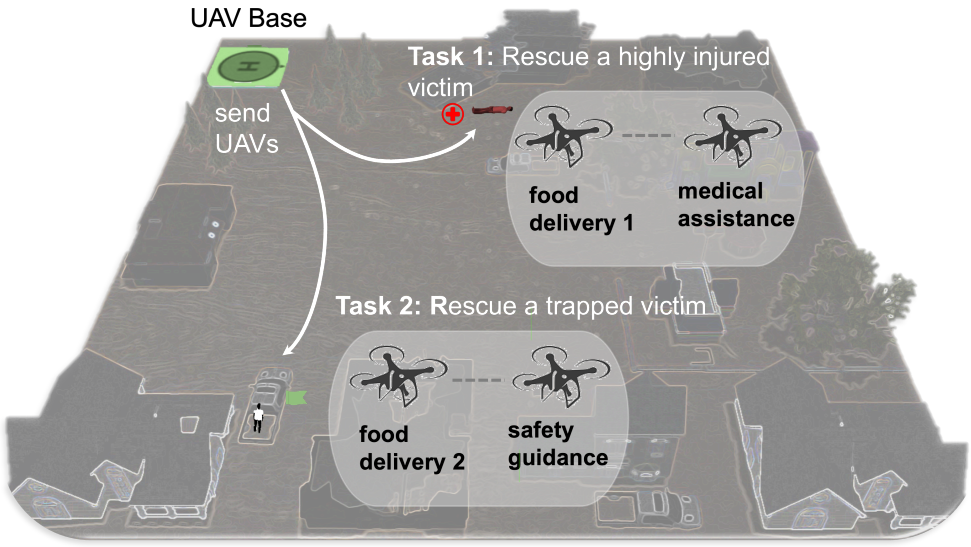

Figure 2. An illustration of the robot task scenario used by the designed ***innerATT*** model. In this scenario, the food delivery robots can flexibly participate in Task 1, "medical assistance for heavily injured victims", or Task 2, "navigation assistance for victims in good conditions".

Table 1. The configurations of robots.

Type	Speed	Mass	Ability
Food Delivery	1.0 m/s	1.0 kg	Food
Safety Guidance	1.5 m/s	0.5 kg	Information
Medical Assistance	1.5 m/s	0.5 kg	Medicine

In this environment, there are two victims with different injury levels and four rescuing robots with different capabilities. To fully utilize the robot team's functionality, each robot needs to cooperate to rescue victims with different injury levels. Of the rescuing robots, two are food delivery robots providing living supplies such as food and water, and one is a safety guidance robot providing victims with useful information about the location of safer places. The remaining robots are medical assistance robots, which are mainly used to provide medical treatments to heavily injured victims. As for the victims, one of them is heavily injured, requiring both food and medical assistance for survival, defined as "Task 1", while another victim who is trapped but in good health needs food delivery as well as safety guidance for moving to a safer place, defined as "Task 2". As for the typical "task variety" scenario, food delivery robots are needed for both kinds of tasks. Therefore, the food delivery robots should flexibly adapt to different tasks and satisfy different task requirements. As for the "robot availability variety" scenario, the medical assistance robot or safety guidance robot's motors could be broken due to mechanical failures, which has negative impacts on food delivery robots' cooperator availability. Therefore, this scenario can be used to evaluate food delivery robots' robustness to real-world disturbances. The simulation settings can also be readily extended to demonstrate the feasibility of the

proposed method in scenarios involving the cooperation of different types of HMRS. In the simulations, ground robots and UAVs may be assigned specific roles based on their unique capabilities. For example, UAVs could perform aerial surveillance to identify points of interest, while ground robots might execute tasks requiring physical interaction, such as collecting samples or clearing debris. The simulation environment can be configured to test various scenarios wherein the cooperation between ground robots and UAVs proves critical, including areas challenging for ground robots to access or tasks necessitating rapid response times, where UAVs can swiftly scout the area. To facilitate adaptive cooperation, the ***innerATT*** is employed to dynamically configure teams of heterogeneous robots, thereby enabling them to efficiently form teams that meet specific task demands.

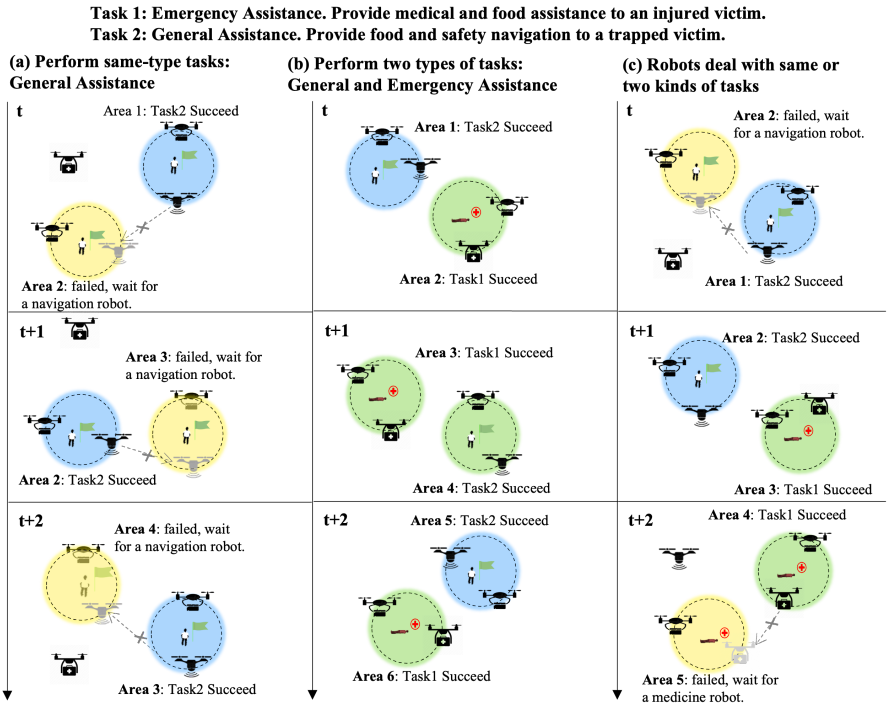

Figure 3. Simulated environment illustration. (**a**) "(S_1) Single Task" in which only one kind of task pops up. (**b**) "(S_2) Double Task" in which two kinds of tasks always pop up. The robot team always deals with Task 1 and Task 2. (**c**) "(S_3) Mixed Task," which is the combination of situations one and two.

To better analyze the effectiveness of the ***innerATT*** method, we design three situations—from simple to complex—for each scenario. In the first situation, "(S_1) Single Task", only one task is presented at random locations: "Task 1" or "Task 2" randomly appear in each episode. In the second situation, "(S_2) Double Task", "Task 1" and "Task 2" always present together in each episode with random locations. The third situation, "(S_3) Mixed Task", is a combination of S_1 and S_2. In addition, two deep reinforcement learning algorithms based on temporal difference (TD) and proximal policy optimization (PPO) are also designed. In the method without inner attention, attention weights α are simply fixed to $\frac{1}{(N-1)}$. Given that only the values of attention weights are changed to a fixed value, both ***innerATT*** and methods without inner attention are implemented with an approximately equal number of parameters. To better analyze the effectiveness of the ***innerATT*** method in improving HMRS adaptability to different task complexities and real-world disturbances, we mainly train our model in S_3 and then analyze its performance in different situations (S_1, S_2, and S_3) and different scenarios ("task variety" and "robot availability variety") without retraining the model.

In the training regimen, the system employs an off-policy variant of the actor–critic methodology, augmented for maximum entropy reinforcement learning, across a span of

25,000 episodes. The computational architecture facilitates parallel processing via 12 threads for handling training data, alongside a replay buffer designed to archive experience tuples of (o_t, a_t, r_t, o_{t+1}) at each timestep. Episodes are set to a duration of 100 steps, with the environment undergoing a reset after each episode, which includes random repositioning of victims and robots. Within these episodes, subsequent to the completion of a rescue operation, only the task parameters are reset, specifically the locations of the victims. Following the conclusion of each episode, both the policy network and the attention critic network undergo quadruple updates. This process involves the selection of 1024 tuples from the replay buffer, followed by the refinement of the Q-function parameters and the policy objectives via policy gradients. The Adam optimization algorithm is utilized for this purpose, with an initial learning rate established at 0.001 and a discount factor γ set to 0.99. For encoding embedded information, a hidden dimension of 128 is selected, and the inner attention mechanism is equipped with four attention heads, optimizing the system's focus and response within the training environment.

4. Results and Discussion

Performance was assessed by the total number of rescued victims and the total length of trajectory per episode. As shown in Figure 4, the methods with ***innerATT*** are competitive when compared to the methods without attention model, which means the methods with ***innerATT*** take fewer steps when rescuing the same number of victims compared with the baseline method.

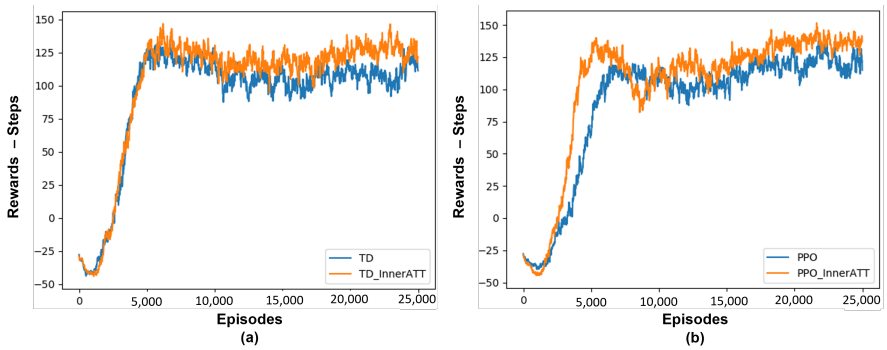

Figure 4. The efficiency comparison of ***innerATT*** and the baseline method when rescuing the same number of victims. (**a**) Training rewards of TD-based methods. (**b**) Training rewards of PPO-based methods.

In the following subsections, ***innerATT*** is mainly analyzed in two directions with different task complexities: In Section 4.1, the ability to adapt to task varieties is discussed. The relationship between robot behavior and their inner attention weights is also analyzed to prove that the inner attention mechanism is beneficial to robots' flexible teaming to different tasks. In Section 4.2, the robustness of HMRS is presented, especially when there are malfunctioning robots in the team. ***innerATT***'s efficiency in resource consumption (moving steps needed) to rescue one victim was also analyzed. In the real-world HMRS application, it is important to pay attention not only to the cooperation strategy that can accomplish complex tasks efficiently, but also to other real-world factors such as resource consumption, distance cost, etc. Since robots usually carry limited energy, if the energy resources of some robots are consumed too fast, there will not be enough energy left for future tasks.

4.1. Adapting to Task Varieties

To analyze ***innerATT***'s ability to adapt to task varieties, the simulated environment "task variety" includes two different kinds of tasks: in Task 1, victims are heavily injured and need medical treatments, while in Task 2 the victims who are in good health need useful information guiding them to a safer place. This means food delivery robots should

learn to dynamically cooperate with proper robots and participate in different tasks based on dynamically changing situations. For example, to rescue victims who are heavily injured, the robot providing medical treatment should cooperate with the closest and available food delivery robot rather than the food delivery robot far away from it or occupied by other rescuing tasks.

To quantitatively measure robots' flexibility, the cooperation rate between food delivery robots and other rescuing robots in a period of time (80 episodes) is calculated using the following formulation:

$$rate_{ij} = \frac{Num_{ij}}{\sum_{k=1}^{N} Num_{ik}} \tag{15}$$

where $\sum_{k=1}^{N} Num_{ik}$ is the total number of victims rescued by robot i and Num_{ij} is the total number of victims rescued by the cooperation of robot i and robot j. The results are shown in Table 2, in the three situations. The average cooperation rates of food delivery robots trained by ***TD-innerATT*** and ***PPO-innerATT***are 0.47/0.53 and 0.48/0.52, respectively, in Task 1, which is similar to uniform distribution with 95% confidence. The cooperation rates of food delivery robots trained by ***TD*** and ***PPO*** methods are 0.82/0.18 and 0.32/0.68, respectively, which is not enough evidence to prove that it is similar to the uniform distribution. Similar results were shown for Task 2: the robots trained by ***TD-innerATT*** or ***PPO-innerATT*** are more flexible than those trained by the methods without the attention model. As suspected, the baseline model's critics use all information non-selectively, while ***innerATT*** can learn which robots to pay more attention to through the inner attention mechanism. Thus, the method with ***innerATT*** is more flexible and sensitive to dynamically changing tasks.

Table 2. UAV participate rate comparison. Numbers in column "Food delivery 1" and "Food delivery 2" are the corresponding cooperation rates.

		Food Delivery 1	**Food Delivery 2**	χ_1^2 **(a = 0.05)**
Task 1	TD-innerATT	0.47	0.53	$0.36 < 3.84$
	TD	0.82	0.18	$81.9 > 3.84$
Task 2	TD-innerATT	0.56	0.44	$1.44 < 3.84$
	TD	0.18	0.82	$81.9 > 3.84$
Task 1	PPO-innerATT	0.48	0.52	$0.16 < 3.84$
	PPO	0.32	0.68	$25.9 > 3.84$
Task 2	PPO-innerATT	0.45	0.55	$1.00 < 3.84$
	PPO	0.73	0.27	$42.3 > 3.84$

To further prove that the inner attention-supported multi-robot teaming method is beneficial to the robot's flexible adaptation to different tasks, Figure 5 demonstrates the effect of the attention head on the robot during the training process by showing the entropy of the attention weights for each robot. A decrease in entropy to approximately 1.02 indicates that the ***innerATT*** mechanism effectively trains robots to focus selectively on certain team members. This decrease signifies not just a concentration of attention but an evolved capability of the robots to prioritize task-relevant interactions, enhancing their collaborative efficiency in dynamic environments. Importantly, the reduction in entropy reflects an improvement in selective attention among robots, crucial for dynamic adaptation to task requirements and robot availability. The metric of entropy is used as a relative measure of the system's ability to filter and focus on relevant information, crucial for effective teamwork. As such, a ''good" level of entropy is contextually defined by the system's improved task performance and adaptability, validating the inner attention mechanism's role in fostering efficient and focused collaborative behavior.

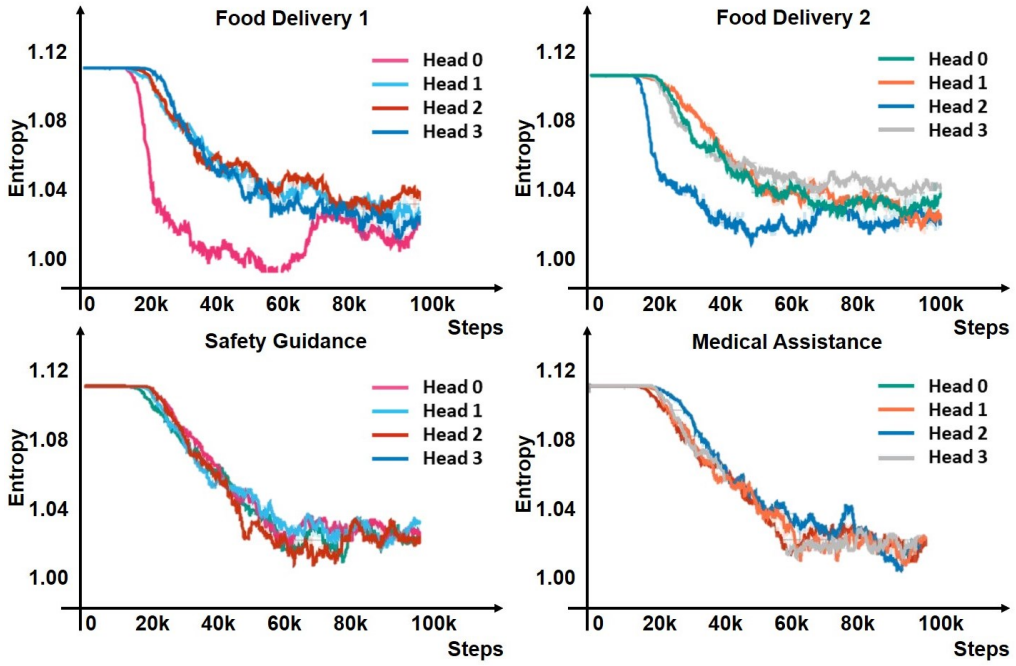

Figure 5. The attention entropy for each attention head throughout the training period within the context of multi-robot cooperation is analyzed. A diminished entropy value signifies that the robots have progressively honed their ability to selectively concentrate their attention on a particular teammate.

Besides that, the relationship between robot behavior and inner attention weights of the robots was analyzed to illustrate attention support in adjusting robot behaviors for flexible teaming. Figure 6A is an illustration of a specific scenario occurring during the experiment. In the pre-stage, Food delivery 1 robot first cooperates with the medical assistance robot to rescue the heavily injured victim (Task 1). At this moment, Food delivery 1 robot needs to pay more attention to the medical assistance robot. After finishing Task 1, in the middle stage and the post-stage, it changes to cooperate with a safety guidance robot to rescue the trapped victim in good health (Task 2). At this time, Food delivery 1 robot needs to pay more attention to the safety guidance robot. Figure 6B shows the curves of Food delivery 1 robot's total attention weights over the other three robots. In the pre-stage, the curve of total attention weights paid to the medical assistance robot has the highest values, which supports Food delivery 1 robot to selectively cooperate with the medical assistance robot. In the middle stage and in the post-stage, the curves of total attention weights paid to the medical assistance robot and safety guidance robot are decreasing and increasing separately, which supports Food delivery 1 robot to transfer its attention from the medical assistance robot to the safety guidance robot. Therefore, the ***innerATT*** can support robot flexible teaming behaviors to different tasks. Figure 6C shows the curves of Food delivery 1 robot's attention weights, generated by each attention head, over other rescuing robots.

Furthermore, to evaluate whether or not the robot cooperation is reasonable, "awkward cooperation" and "expected adaptive cooperation" were introduced based on the above-mentioned cooperation rules. In Figure 7, these typical cases are common in all situations. The typical cases were divided into three categories according to the distances between robots and task: both food delivery robots are closer to the victim that is close to the cooperating robot, both food delivery robots are closer to the victim that is far away from cooperating robot, and one of food delivery robot is closer to one victim, the other one is closer to another victim. For all the cases, the awkward cooperation (victims are not rescued by the closest food delivery robot) is illustrated by orange arrow lines, while the expected adaptive cooperation (victims are rescued by the closest food delivery robot) is

presented by green arrow lines. In awkward cooperation cases, longer time is required for the robots to rescue victims, which is unacceptable, especially in disaster search and rescue. Therefore, even though awkward cooperation can rescue victims, it has a lower cooperation quality than expected adaptive cooperation.

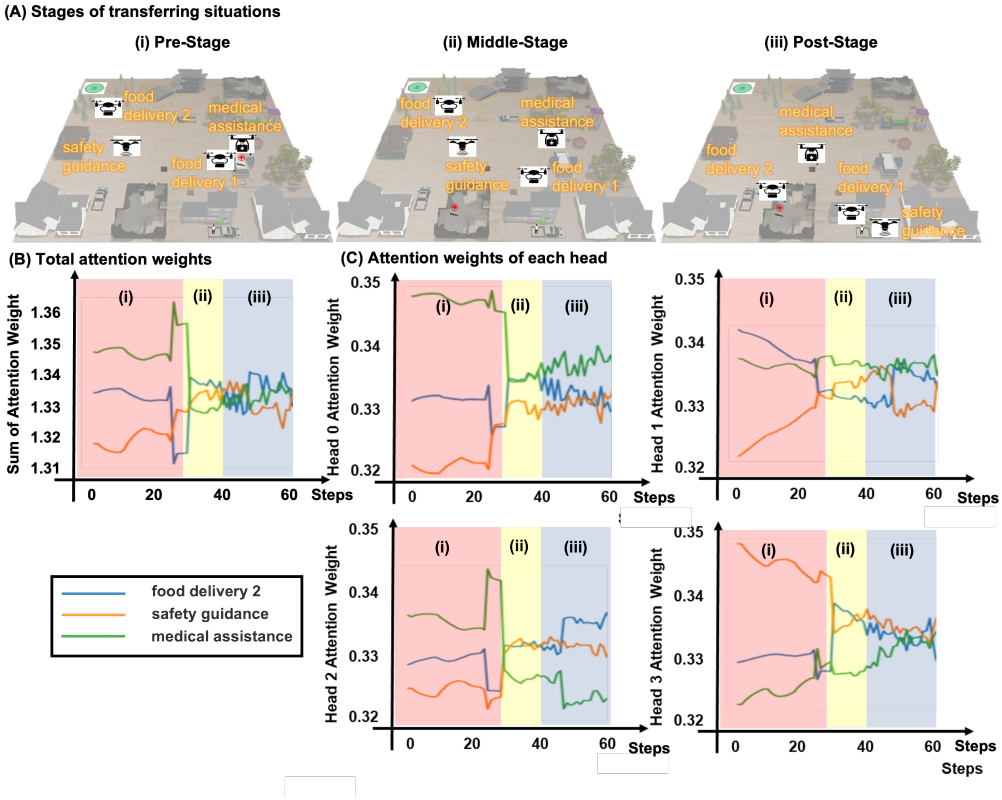

Figure 6. Relationships between Food delivery 1 robot's conduct and its internal attention weights within adaptive collaboration settings in a multi-robot cooperation context under scenario two (S_2) are explored. (**A**) delineates three phases of Food delivery 1 robot's dynamic teaming approach. Initially, in the pre-stage (i), collaboration is established between Food delivery 1 robot and the medical assistance robot. During the intermediate stage (ii), Food delivery 1 robot modifies its behavior by leveraging an intrinsic attention mechanism. In the concluding post-stage (iii), Food delivery 1 robot engages in cooperation with the safety guidance robot. (**B**) encapsulates the aggregate attention weight that Food delivery 1 robot allocates to its robotic peers. (**C**) details the individual attention weights generated by each attention head within Food delivery 1 robot.

To quantitatively measure robot cooperation quality, the rates of awkward cooperation and expected adaptive cooperation were calculated. As shown in Table 3, for the method ***TD-innerATT*** and ***PPO-innerATT***, after 20 episodes, the average rates of Food delivery 1 robot's awkward cooperation in Task 1 and Task 2 are 0.17, 0.22 and 0.26, 0.30, respectively. Food delivery 2 robot's average awkward cooperation rates in Task 1 and Task 2 are 0.17, 0.16, and 0.30, 0.18, respectively. As for the results for methods ***TD*** and ***PPO***, the average rates of Food delivery 1 robot's awkward cooperation in Task 1 and Task 2 are 0.43, 0.45, and 0.44, 0.47. Food delivery 2 robot's average awkward cooperation rates in Task 1 and Task 2 are 0.40, 0.49, and 0.46, 0.44, respectively. The results show that methods with ***innerATT*** have fewer awkward cooperative actions in different tasks than the baseline method because the robots with inner attention mechanism can flexibly cooperate with different robots based on the current situations instead of always cooperating with the same robot. Therefore, methods with ***innerATT*** can have more meaningful cooperative actions based on the dynamically changing environments compared with the baseline method.

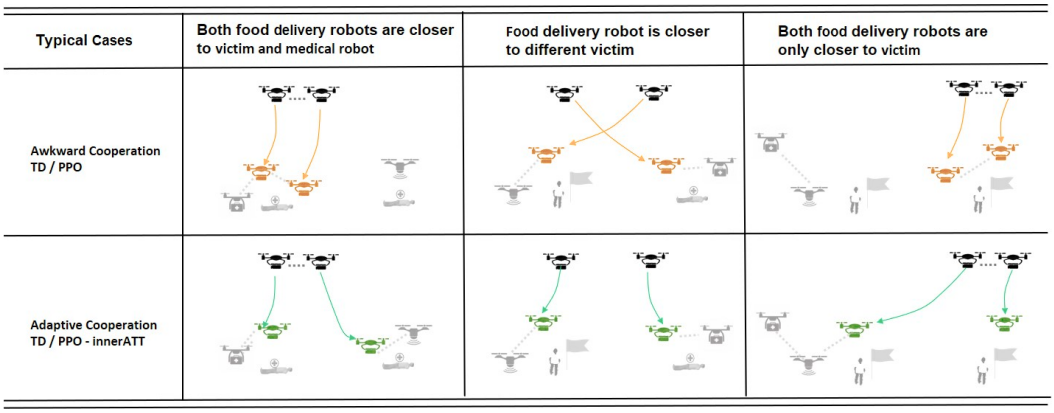

Figure 7. Typical awkward (orange color) and adaptive cooperation (green color) in the three task situations "S_1 Single Task, S_2 Double Tasks, s_3 Mixed Tasks".

Table 3. Average awkward cooperation rate.

		Food Delivery 1	Food Delivery 2
Task 1	TD-innerATT	0.17	0.17
	TD	0.43	0.40
Task 2	TD-innerATT	0.22	0.16
	TD	0.45	0.49
Task 1	PPO-innerATT	0.26	0.30
	PPO	0.44	0.46
Task 2	PPO-innerATT	0.30	0.18
	PPO	0.47	0.44

To quantitatively prove that ***innerATT*** is much more efficient in energy consumption than the baseline method, the average trajectory distance needed to rescue one victim was calculated by the following formulation:

$$\overline{Distance_T} = \frac{Distance^T_{Total}}{Victims^T_{Total}} \tag{16}$$

where $\overline{Distance_T}$ is the average distance cost to rescue one victim during time T, $Distance^T_{Total}$ is the total distance calculated by summing all robots trajectory length in a period of time T, and $Victims^T_{Total}$ is the total number of rescued victims during time T. As Figure 8 shows, after 10,000 training episodes, the average trajectory distance cost of the model trained by ***TD-innerATT*** is 0.14 greater than that of the model trained by the baseline method. The model trained by ***TD*** is more efficient because the inner attention mechanism increases the complexity of the Deep Neural Network framework. So, the baseline method can learn faster than the method with ***innerATT*** and can rescue more victims at the beginning of the training phase. After 25,000 episodes of training, when ***innerATT*** is sufficiently trained, the average trajectory distance cost of the model trained by ***TD-innerATT*** is 0.10 less than that of the model trained by the baseline method. Similar results were also obtained from the experiment in which the models were trained based on the PPO method. After 25,000 episodes of training, the average trajectory distance cost of the model trained by ***PPO-innerATT*** was 0.19 lesser than that of the model trained by the baseline method.Therefore, the robots trained by the methods with ***innerATT*** are more efficient than the robots trained by the baseline method.

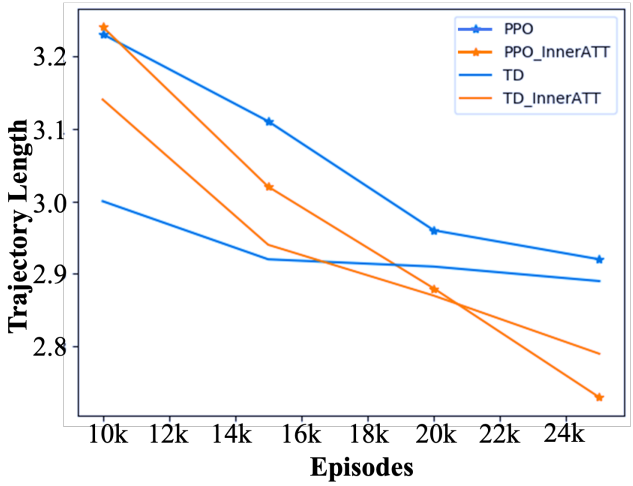

Figure 8. Average trajectory length for rescuing one victim in (S_2).

4.2. Adapting to Robot Availability

In addition to robot flexible teaming, robustness to real-world disturbances is important in HMRS. If the robots cannot flexibly adapt to real-world disturbances, for example, if some robots are broken in the robot team or if the faults are caused by sensor failures, then there may be undesirable and uncontrollable effects on other teammates. What is more, broken robots may share incorrect information with other members of the team, leading to incorrect cooperation behaviors.

With the ***innerATT***, the HMRS team is more robust to sensor failure or broken units, which has been theoretically proved in the Methods section. To practically measure the robustness of HMRS, the typical robot failure issue "motor broke" is simulated. Then, the food delivery robots' cooperation rates are calculated to estimate their robustness to the "motor broken" disturbance. In the ideal cases, if the food delivery robots are robust enough, they have an equal chance to participate in reusing tasks. That means food delivery robots are not influenced by faulty robots. As Table 4 shows, considering Task 1 when the safety guidance robot is broken, the average cooperation rates of food delivery robots trained by ***TD-innerATT*** and ***PPO-innerATT*** are 0.51/0.49 and 0.51/0.49, respectively, which is similar to uniform distribution with 95% confidence; the cooperation rates of food delivery robots trained by ***TD*** and ***PPO*** methods are 0.82/0.18 and 0.34/0.66, which means the food delivery robots are significantly influenced by the broken robot. As for Task 2 when the medical assistance robot is broken, similar results are observed. Therefore, the robots trained by the methods with ***innerATT*** are more robust to robot failure than those trained by the baseline method.

To further prove that the ***innerATT*** is beneficial to robot robustness to real-world factors, the relationship between robot behavior and their inner attention weights was analyzed to illustrate attention supports in adjusting robot behaviors for increasing robot resilience. Figure 9A is an illustration of a specific scenario occurring during experiments. In the pre-stage, Food delivery 1 robot is initially cooperating with the medical assistance robot to rescue the heavily injured victim (Task 1). At this moment, the medical assistance robot needs to pay more attention to Food delivery 1 robot. After finishing Task 1, in the middle-stage and the post-stage, DFod delivery 2 robot cooperates with the medical assistance robot to rescue another heavily injured victim (Task 1). At this time, the medical assistance robot needs to pay more attention to Food delivery 2 robot. Figure 9B shows the curves of the medical assistance robot's total attention weights over the other three robots. In the pre-stage, the curve of total attention weights paid to Food delivery 1 robot has the highest values, which supports the medical assistance robot to selectively cooperate with Food delivery 1 robot. In the middle stage and in the post-stage, the curves of total attention weights paid to Food delivery 1 robot and Food delivery 2 robot are decreasing and increasing separately, which supports the medical assistance robot to transfer its attention from

Food delivery 1 robot to Food delivery 2 robot. Therefore, the inner attention mechanism can increase robot robustness to real-world robot failures by adjusting robot behaviors for increasing robot resilience. Figure 9C shows the curves of the medical assistance robot's attention weights, generated by each attention head, over other rescuing robots.

Table 4. UAV participate rate when one robot is broken. Numbers in column "Food delivery 1" and "Food delivery 2" are the corresponding cooperation rates.

		Food Delivery 1	Food Delivery 2	χ_1^2 (a = 0.05)
Task 1	TD-innerATT	0.51	0.49	$0.04 < 3.84$
	TD	0.82	0.18	$81.9 > 3.84$
Task 2	TD-innerATT	0.57	0.43	$0.36 < 3.84$
	TD	0.17	0.83	$87.1 > 3.84$
Task 1	PPO-innerATT	0.51	0.49	$0.04 < 3.84$
	PPO	0.34	0.66	$20.4 > 3.84$
Task 2	PPO-innerATT	0.47	0.53	$0.36 < 3.84$
	PPO	0.72	0.28	$38.7 > 3.84$

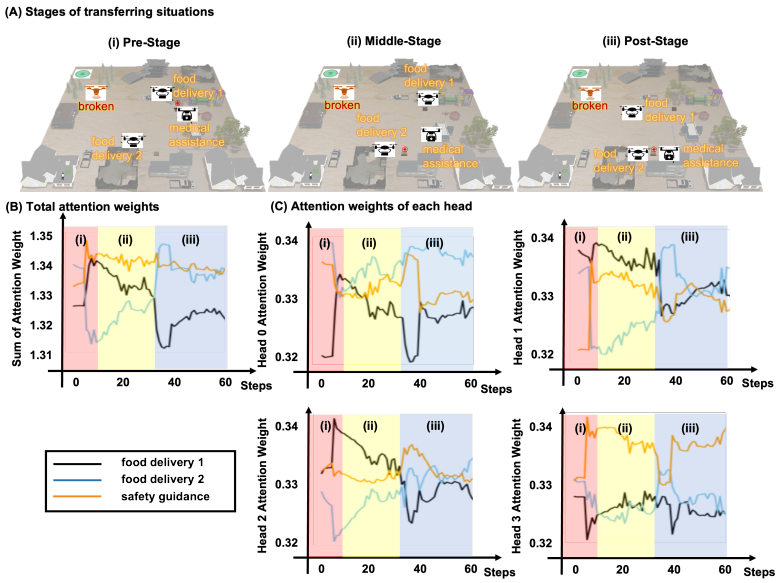

Figure 9. Relationships between medical assistance robot's behavior and its inner attention weights in adaptive teaming in the multi-robot cooperation environment with situation two (S_2). (**A**) Three stages of medical assistance robot's flexible teaming. In the pre-stage (i), the medical assistance robot cooperates with Food delivery 1 robot. In the middle stage (ii), the medical assistance robots change their behavior based on the inner attention mechanisms. In the post-stage (iii), the medical assistance robot cooperates with Food delivery 2 robot. (**B**) Medical assistance robot's total attention weight paid to other robots. (**C**) The medical assistance robot's attention weights are obtained from each attention head.

5. Conclusions

- **Summary.** This study introduces a novel inner attention mechanism, ***innerATT***, designed to facilitate adaptive cooperation among multi-heterogeneous robots in response to varying task requirements. Through the deployment of scenarios with diverse task configurations, such as a single task, a double task, and dynamically mixed tasks, the efficacy of the ***innerATT*** model in promoting flexible team formation

is empirically confirmed. The model adeptly navigates the challenge of distributing limited robot resources across fluctuating task demands.

- **Potential Application.** Additionally, the theoretical framework of this model offers potential for broad application, including the coordination of ground and aerial vehicles, as well as integration between vehicular units and human operatives. Consequently, the attention-driven flexible teaming model unveiled in this research holds substantial promise for practical implementation across a spectrum of multi-robot applications, ranging from disaster response to wildlife conservation and the management of airport traffic flows.
- **Novelty.** While building upon foundational research in multi-agent reinforcement learning [52,53], our work introduces a distinctive approach by focusing on the strategic formation and adaptability of robot teams to task demands and environmental changes, utilizing innerATT for adaptive cooperation. This unique contribution addresses unexplored challenges in the field, extending beyond the scope of prior studies, and highlights the innovative potential of attention mechanisms in enhancing HMRS operations.
- **Practical Challenges.** The primary goal of this research is to assess the feasibility of using attention mechanisms to flexibly assemble heterogeneous robot teams. Given the differences between simulated environments and their real-world counterparts, the model developed in this study might not perform identically in practical settings. Nonetheless, the model acts as an essential initial step for further development and adjustment to actual conditions. It is vital to acknowledge the difficulties in applying simulation-based methods in real environments. The unpredictability and sensor inaccuracies inherent in real-world scenarios necessitate thorough validation and improvement of any theoretical model. Consequently, our future endeavors aim to narrow the gap between simulation outcomes and real-world implementation. This will include a comprehensive analysis and the integration of specific sensors (e.g., LIDARs, cameras, GPS) for various applications, alongside the examination of robust communication protocols to facilitate efficient team coordination amidst the challenges of bandwidth and latency in operational environments. Additionally, future research will explore robot behavior analysis and the creation of models to measure and improve human trust in heterogeneous multi-robot systems with the goal of enhancing their real-world efficacy.

Author Contributions: Conceptualization, Y.G., C.H. and R.L.; methodology, Y.G., C.H. and R.L.; software, Y.G., C.H. and R.L.; validation, Y.G., C.H. and R.L.; formal analysis, Y.G., C.H. and R.L.; investigation, Y.G., C.H. and R.L.; resources, Y.G., C.H. and R.L.; data curation, Y.G., C.H. and R.L.; writing—original draft preparation, Y.G., C.H. and R.L.; writing—review and editing, Y.G., C.H. and R.L.; visualization, Y.G., C.H. and R.L.; supervision, R.L.; project administration, R.L. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: The study was conducted in accordance with the Declaration of Helsinki and approved by the Kent State Institutional Review Board (IRB) (FWA 00001853, Expires 2 September 2026). This article does not contain any studies with animals performed by any of the authors.

Informed Consent Statement: Not applicable.

Data Availability Statement: The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

HMRS	Heterogeneous Multi-Robot System
inner-ATT	Inner Attention
MAAC	Multi-Agent Actor–Critic
UAV	Unmanned Aerial Vehicle
TD	Temporal Difference
PPO	Proximal Policy Optimization

References

1. Matos, A.; Martins, A.; Dias, A.; Ferreira, B.; Almeida, J.M.; Ferreira, H.; Amaral, G.; Figueiredo, A.; Almeida, R.; Silva, F. Multiple robot operations for maritime search and rescue in euRathlon 2015 competition. In Proceedings of the OCEANS 2016-Shanghai, Shanghai, China, 10–13 April 2016; pp. 1–7.
2. Mouradian, C.; Yangui, S.; Glitho, R.H. Robots as-a-service in cloud computing: Search and rescue in large-scale disasters case study. In Proceedings of the 2018 15th IEEE Annual Consumer Communications and Networking Conference (CCNC), Las Vegas, NV, USA, 12–15 January 2018; pp. 1–7.
3. Beck, Z.; Teacy, N.R.; Rogers, A.C. Online planning for collaborative search and rescue by heterogeneous robot teams. Association of Computing Machinery. In Proceedings of the AAMAS'16: International Conference on Agents and Multiagent Systems, Singapore, 9–13 May 2016.
4. Alotaibi, E.T.S.; Al-Rawi, H. Multi-robot path-planning problem for a heavy traffic control application: A survey. *Int. J. Adv. Comput. Sci. Appl.* **2016**, *7*, 10.
5. Digani, V.; Sabattini, L.; Secchi, C.; Fantuzzi, C. Towards decentralized coordination of multi robot systems in industrial environments: A hierarchical traffic control strategy. In Proceedings of the 2013 IEEE 9th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 5–7 September 2013; pp. 209–215.
6. Digani, V.; Sabattini, L.; Secchi, C.; Fantuzzi, C. Hierarchical traffic control for partially decentralized coordination of multi agv systems in industrial environments. In Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA), Hong Kong, China, 31 May–7 June 2014; pp. 6144–6149.
7. Broecker, B.; Caliskanelli, I.; Tuyls, K.; Sklar, E.I.; Hennes, D. Hybrid insect-inspired multi-robot coverage in complex environments. In Proceedings of the Conference Towards Autonomous Robotic Systems, London, UK, 3–5 July 2019; pp. 56–68.
8. Kolling, A.; Carpin, S. Multi-robot surveillance: An improved algorithm for the graph-clear problem. In Proceedings of the 2008 IEEE International Conference on Robotics and Automation, Pasaena, CA, USA, 19–23 May 2008; pp. 2360–2365.
9. Easton, K.; Burdick, J. A coverage algorithm for multi-robot boundary inspection. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 727–734.
10. Zhu, A.M.; Yang, S.X. An improved SOM-based approach to dynamic task assignment of multi-robot. In Proceedings of the World Congress on Intelligent Control and Automation, Jinan, China, 7–9 July 2010; pp. 2168–2173.
11. Fazli, P.; Davoodi, A.; Mackworth, A.K. Multi-robot repeated area coverage. *Auton. Robot.* **2013**, *34*, 251–276. [CrossRef]
12. Boardman, M.; Edmonds, J.; Francis, K.; Clark, C.M. Multi-robot boundary tracking with phase and workload balancing. In Proceedings of the International Conference on Intelligent Robots and Systems, Taipei, Taiwan, 18–22 October 2010; pp. 3321–3326.
13. Zhu, D.Q. ; Huang, H.; Yang, S.X. Dynamic task assignment and path planning of multi-AUV system based on an improved self-organizing map and velocity synthesis method in three-dimensional underwater workspace. *IEEE Trans. Cybern.* **2013**, *43*, 504–514.
14. Vergnano, A.; Thorstensson, C.; Lennartson, B.; Falkman, P.; Pellicciari, M.; Leali, F.; Biller, S. Modeling and optimization of energy consumption in cooperative multi-robot systems. *IEEE Trans. Autom. Sci. Eng.* **2012**, *9*, 423–428. [CrossRef]
15. Parker, L.E. Adaptive heterogeneous multi-robot teams. *Neurocomputing* **1999**, *28*, 75–92. [CrossRef]
16. Kim, J.; Cauli, N.; Vicente, P.; Damas, B.; Bernardino, A.; Santos-Victor, J.; Cavallo, F. Cleaning tasks knowledge transfer between heterogeneous robots: A deep learning approach. *J. Intell. Robot. Syst.* **2020**, *98*, 191–205. [CrossRef]
17. Prorok, A.; Hsieh, M.A.; Kumar, V. Fast redistribution of a swarm of heterogeneous robots. In Proceedings of the 9th EAI International Conference on Bio-inspired Information and Communications Technologies (formerly BIONETICS), New York City, NY, USA, 3–5 December 2015; pp. 249–255.
18. Saribatur, Z.G.; Patoglu, V.; Erdem, E. Finding optimal feasible global plans for multiple teams of heterogeneous robots using hybrid reasoning: An application to cognitive factories. *Auton. Robot.* **2019**, *43*, 213–238. [CrossRef]
19. Atay, N.; Bayazit, B. *Mixed-Integer Linear Programming Solution to Multi-Robot Task Allocation Problem*; Washington University: St. Louis, MO, USA, 2006.
20. Darrah, M.; Nil, W.; Stolarik, B. *Multiple UAV Dynamic Task Allocation Using Mixed Integer Linear Programming in a SEAD Mission*; Infotech at Aerospace: Arlington, VA, USA, 2005; p. 7165.
21. Mosteo, A.R.; Montano, L. Simulated annealing for multi-robot hierarchical task allocation with flexible constraints and objective functions. In Proceedings of the Workshop on Network Robot Systems: Toward Intelligent Robotic Systems Integrated with Environments, Beijing, China, 10 October 2006.

22. Juedes, D.; Drews, F.; Welch, L.; Fleeman, D. Heuristic resource allocation algorithms for maximizing allowable workload in dynamic, distributed real-time systems. In Proceedings of the International Parallel and Distributed Processing Symposium, Santa Fe, NM, USA, 26–30 April 2004; p. 117.
23. Kmiecik, W.; Wojcikowski, M.; Koszalka, L.; Kasprzak, A. Task allocation in mesh connected processors with local search meta-heuristic algorithms. In Proceedings of the Asian Conference on Intelligent Information and Database Systems, Hue City, Vietnam, 24–26 March 2010; pp. 215–224.
24. Iijima, N.; Sugiyama, A.; Hayano, M.; Sugawara, T. Adaptive task allocation based on social utility and individual preference in distributed environments. *Procedia Comput. Sci.* **2017**, *112*, 91–98. [CrossRef]
25. Lope, D.; Javier, D.; Quiñonez, Y. Response threshold models and stochastic learning automata for self-coordination of heterogeneous multi-tasks distribution in multi-robot systems. *Robot. Auton. Syst.* **2012**, *61*, 714–720. [CrossRef]
26. Elfakharany, A.; Yusof, R.; Ismail, Z. Towards multi-robot Task Allocation and Navigation using Deep Reinforcement Learning. *J. Phys. Conf. Ser.* **2020**, *1447*, 012045. [CrossRef]
27. Fan, T.X.; Long, P.X.; Liu, W.X.; Pan, J. Fully distributed multi-robot collision avoidance via deep reinforcement learning for safe and efficient navigation in complex scenarios. *arXiv* **2018**, arXiv:1808.03841.
28. Noureddine, D.B.; Gharbi, A.; Ahmed, S.B. Multi-agent Deep Reinforcement Learning for Task Allocation in Dynamic Environment. In Proceedings of the 12th International Conference on Software Technologies, ICSOFT 2017, Madrid, Spain, 24–26 July 2017; pp. 17–26. [CrossRef]
29. Luo, T.Z.; Subagdja, B.; Wang, D.; Tan, A. Multi-Agent Collaborative Exploration through Graph-based Deep Reinforcement Learning. In Proceedings of the 2019 IEEE International Conference on Agents (ICA), Jinan, China, 18–21 October 2019; pp. 2–7.
30. Harnett, B.M.; Doarn, C.R.; Rosen, J.; Hannaford, B.; Broderick, T.J. Evaluation of unmanned airborne vehicles and mobile robotic telesurgery in an extreme environment. *Telemed.-Health* **2008**, *14*, 539–544. [CrossRef] [PubMed]
31. Zhang, F.; Chen, W. Self-healing for mobile robot networks with motion synchronization. In Proceedings of the 2007 IEEE/RSJ International Conference on Intelligent Robots and Systems, San Diego, CA, USA, 29 October–2 November 2007; pp. 3107–3112.
32. Liu, Z.; Ju, J.; Chen, W.; Fu, X.Y.; Wang, H. A gradient-based self-healing algorithm for mobile robot formation. In Proceedings of the 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, 28 September–2 October 2015; pp. 3395–3400.
33. Mathews, N.; Christensen, A.L.; O'Grady, R.; Mondada, F.; Dorigo, M. Mergeable nervous systems for robots. *Nat. Commun.* **2017**, *8*, 439. [CrossRef] [PubMed]
34. Mathews, N.; Christensen, A.L.; Stranieri, A.; Scheidler, A.; Dorigo, M. Supervised morphogenesis: Exploiting morphological flexibility of self-assembling multirobot systems through cooperation with aerial robots. *Robot. Auton. Syst.* **2019**, *112*, 154–167. [CrossRef]
35. Pelc, A.; Peleg, D. Broadcasting with locally bounded byzantine faults. *Inf. Process. Lett.* **2005**, *93*, 109–115. [CrossRef]
36. Saulnier, K.; Saldana, D.; Prorok, A.; Pappas, G.J.; Kumar, V. Resilient flocking for mobile robot teams. *IEEE Robot. Autom. Lett.* **2017**, *2*, 1039–1046. [CrossRef]
37. Liu, R.; Jia, F.; Luo, W.; Chandarana, M.; Nam, C.; Lewis, M.; Sycara, K.P. Trust-Aware Behavior Reflection for Robot Swarm Self-Healing. In Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems, Montreal, QC, Canada, 13–17 May 2019; pp. 122–130.
38. Foerster, R.M.; Schneider, W.X. Task-Irrelevant Features in Visual Working Memory Influence Covert Attention: Evidence from a Partial Report Task. *Vision* **2019**, *3*, 42. [CrossRef] [PubMed]
39. Aly, A.; Griffiths, S.; Stramandinoli, F. Metrics and benchmarks in human-robot interaction: Recent advances in cognitive robotics. *Cogn. Syst. Res.* **2017**, *43*, 313–323. [CrossRef]
40. Huber, A.; Weiss, A. Developing human-robot interaction for an industry 4.0 robot: How industry workers helped to improve remote-HRI to physical-HRI. In Proceedings of the Companion of the 2017 ACM/IEEE International Conference on Human-Robot Interaction, Vienna, Austria, 6–9 March 2017; pp. 137–138.
41. Anzalone, S.M.; Boucenna, S.; Ivaldi, S.; Chetouani, M. Evaluating the engagement with social robots. *Int. J. Soc. Robot.* **2015**, *7*, 465–478. [CrossRef]
42. Daglarli, E.; Daglarli, S.F.; Gunel, G.O.; Kose, H. Improving human-robot interaction based on joint attention. *Appl. Intell.* **2017**, *47*, 62–82. [CrossRef]
43. So, W.C.; Wong, M.K.; Lam, W.Y.; Cheng, C.H.; Yang, J.H.; Huang, Y.; Ng, P.; Wong, W.L.; Ho, C.L.; Yeung, K.L. Robot-based intervention may reduce delay in the production of intransitive gestures in Chinese-speaking preschoolers with autism spectrum disorder. *Mol. Autism.* **2018**, *9*, 34. [CrossRef] [PubMed]
44. Jiang, J.C.; Lu, Z.Q. Learning attentional communication for multi-agent cooperation. In Proceedings of the Neural Information Processing Systems, Montreal, QC, Canada, 3–8 December 2018; pp. 7254–7265.
45. Geng, M.Y.; Xu, K.L.; Zhou, X.; Ding, B.; Wang, H.M.; Zhang, L. Learning to cooperate via an attention-based communication neural network in decentralized multi-robot exploration. *Entropy* **2019**, *21*, 294. [CrossRef] [PubMed]
46. Capitan, J.; Spaan, M.T.; Merino, L.; Ollero, A. Decentralized multi-robot cooperation with auctioned POMDPs. *Int. J. Robot. Res.* **2018**, *32*, 650–671. [CrossRef]
47. Iqbal, S.; Sha, F. Actor-attention-critic for multi-agent reinforcement learning. *Int. Conf. Mach. Learn.* **2019**, *97*, 2961–2970.

48. Lowe, R.; Wu, Y.I.; Tamar, A.; Harb, J.; Abbeel, O.P.; Mordatch, I. Multi-agent actor-critic for mixed cooperative-competitive environments. In Proceedings of the Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 6379–6390.
49. Patkin, M.L.; Rogachev, G.N. Construction of multi-agent mobile robots control system in the problem of persecution with using a modified reinforcement learning method based on neural networks. *IOP Conf. Ser. Mater. Sci. Eng.* **2018**, *32*, 012018. [CrossRef]
50. Gupta, J.K.; Egorov, M.; Kochenderfer, M. Cooperative multi-agent control using deep reinforcement learning. In *International Conference on Autonomous Agents and Multiagent Systems*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 66–83.
51. Hsieh, Y.L.; Cheng, M.H.; Juan, D.C.; Wei, W.; Hsu, W.L.; Hsieh, C.J. On the robustness of self-attentive models. In Proceedings of the Association for Computational Linguistics, Florence, Italy, 28 July–2 August 2019; pp. 1520–1529.
52. Hu, S.; Shen, L.; Zhang, Y.; Tao, D. Learning Multi-Agent Communication from Graph Modeling Perspective. In *The Twelfth International Conference on Learning Representations*; ICLR: Appleton, WI, USA, 2023.
53. Seraj, E.; Wang, Z.; Paleja, R.; Martin, D.; Sklar, M.; Patel, A.; Gombolay, M. Learning efficient diverse communication for cooperative heterogeneous teaming. In Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems, Auckland, New Zealand, 9–13 May 2022; pp. 1173–1182.