

Article

# Online Unmanned Ground Vehicle Path Planning Based on Multi-Attribute Intelligent Reinforcement Learning for Mine Search and Rescue

Shanfan Zhang <sup>†</sup> and Qingshuang Zeng <sup>\*,†</sup>

Space Control and Inertial Technology Research Center, Harbin Institute of Technology, Harbin 150001, China; zsanfan@outlook.com

\* Correspondence: zqshuang@hit.edu.cn

† These authors contributed equally to this work.

**Abstract:** Aiming to improve the efficiency of the online process in path planning, a novel searching method is proposed based on environmental information analysis. Firstly, a search and rescue (SAR) environmental model and an unmanned ground vehicle (UGV) motion model are established according to the characteristics of a mining environment. Secondly, an online search area path-planning method is proposed based on the gray system theory and the reinforcement learning theory to handle multiple constraints. By adopting the multi-attribute intelligent (MAI) gray decision process, the action selection decision can be dynamically adjusted based on the current environment, ensuring the stable convergence of the model. Finally, experimental verification is conducted in different small-scale mine SAR simulation scenarios. The experimental results show that the proposed search planning method can capture the target in the search area with a smoother convergence effect and a shorter path length than other path-planning algorithms.

**Keywords:** search and rescue (SAR); unmanned system; path planning; partially observable Markov decision process (POMDP); gray system



**Citation:** Zhang, S.; Zeng, Q. Online Unmanned Ground Vehicle Path Planning Based on Multi-Attribute Intelligent Reinforcement Learning for Mine Search and Rescue. *Appl. Sci.* **2024**, *14*, 9127. <https://doi.org/10.3390/app14199127>

Academic Editor: Yutaka Ishibashi

Received: 9 September 2024

Revised: 29 September 2024

Accepted: 7 October 2024

Published: 9 October 2024



**Copyright:** © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

UGVs can replace humans in performing hazardous tasks with their superior speed, mobility, and atmospheric independence [1]. In recent years, research into UGV control has concentrated on navigation, SAR, area coverage, and military tasks. As an important function of UGV, SAR has received significant attention in academia, with the key focus areas including search area prediction [2,3], SAR force configuration [4], and rescue path planning [5–7]. The complex structure of mining caves lacks environmental information, and there is a risk of secondary explosions and collapses during the rescue process. To address the above dilemma, UGVs (see more detailed information about mining UGVs in [8]) have been considered to replace humans in dangerous SAR missions due to their advantages of low cost and high efficiency. Additionally, their path planning is a prerequisite for subsequent rescue missions. Therefore, the path-planning problem derived from this has always been an important topic in the field of UGV applications [9–11].

Previously, the research focus of SAR was on point-to-point path planning. According to the difference in target information, it can be categorized as known and unknown target path planning [12]. When in different application backgrounds, it can be classified as indoor planning, outdoor planning, or maritime planning [13]. In addition, according to the types of application platforms, robots can be classified as isomorphic path planners or heterogeneous path planners [14–16]. What is noteworthy is that most of the above methods plan paths based on precise or fuzzy target location information, which has high dependence on the in-variance of the demand. However, path planning without the targets' location information is more in line with the general scenario and more challenging.

To address the dimension explosion caused by large amounts of data, most online POMDP methods are based on forward-looking searches, and these technologies reduce the complexity from different angles [17,18]. Among them, the most classic is PRIMAL [19]. These methods can usually be classified into three categories: Branch-and-Bound Pruning [20,21], Monte Carlo Sampling [22–26], and Heuristic Search [27–30]. However, most research using these methods has the following main shortcomings: (1) agents require much training before completing path-planning tasks, namely offline learning, and spend many resources on that; (2) it is also difficult for agents to respond when task maps differ significantly from training maps.

When the environmental structure is unfixed, observations are discrete through the sensor system's data acquisition process. In order to study path-planning methods, modeling the environmental information of mining caves is necessary. In addition, obviously, a path-planning problem with real-time reactions to environmental uncertainty is NP-hard, which has been proven by Ryan A. MacDonald and Stephen L. Smith using informative path planning [31]. And the partial observable Markov decision process (POMDP) search balances short-term tracking performance and the long-term final cost [32]. The environmental information objectively exists in a stable state of change, in which the subject generally lacks awareness. The external world is a generally steady resource, and stable information is considered to not require storage [33]. Targeting the challenges above, this paper proposes a gray Q-Learning (GQL) search area path-planning method using UGV as the agent. This method combines the even gray model (EGM) prediction model, the MAI gray decision process, and the Q-Learning (QL) action planning model. The main contributions of this paper are as follows:

- A method for constructing an environmental model and an information feature mining method based on EGM are proposed to address the lack of environmental information in mining SAR.
- An agent-centered path-planning model based on the RL theory is proposed for the online path-planning problem. The optimization reward function designed for multiple scenarios effectively solves the conflict problem between paths, obstacles, and traps.
- A heuristic decision-making strategy based on the gray system theory is proposed for our SAR problem, which helps the model accelerate convergence towards the target and improve the robustness of the intelligent agent decision-making process.

The remaining sections of this paper are organized as follows: Section 2 introduces the constraints in the search area path-planning process used in Section 3. Section 4 provides a detailed description of the algorithm, a GQL search area path-planning method for the mine SAR problem. The proposed model is compared with the A\* search algorithm [34], rapidly exploring the random tree (RRT) algorithm [35] and QL algorithm [36]. The simulation results and their discussion follow in Section 5, and the paper is concluded with a summary and an outlook for future work in Section 6.

## 2. Constraints in Search Area Path Planning

Although the natural environment undergoes constant changes, subtle changes occur in adjacent environments. Therefore, ground information is usually a stable resource that is always continuous when there are no obstacles [33]. Environmental information objectively exists in a stable state of change, but the subject generally lacks awareness of this fact. In new environments, it is impossible to obtain environmental data, but the observations with the scope measured by the sensor system describe the UGV's understanding of the environment. The assumptions are as follows:

**Assumption 1.** *The SAR environmental information is described as the measured value of the sensor system.*

**Assumption 2.** *The SAR environmental state always continues slightly.*

There may be multiple coupled constraints for point-to-point path planning with an unknown target position, and the constraints that this paper focuses on are the resource and time expenditure of the search target. Based on the above analysis, it is reasonable to evaluate the efficiency of search area path-planning methods from the following aspects.

### 2.1. Probability of Detection (POD)

The *POD* calculates the similarity between the characteristic values of the current position and the target position, as shown in (1), and describes the probability of a successful search at the current position of the target.

$$POD = \frac{1}{4} \frac{(1 + e^{0.1(c_T - c_a)})^2}{e^{0.1(c_T - c_a)}} \quad (1)$$

### 2.2. Relative Distance (RD)

The *RD* is defined as follows, and refers to the Euclidean distance between the agent's and target's current positions in real time.

$$RD = |p_t - p_a| \quad (2)$$

$p_t$  is the grid unit of the target, and  $p_a$  is the grid unit of the agent.

### 2.3. Characteristic Distance (CD)

The *CD* is defined in (3) and it describes the ability of agents to approach targets autonomously. The faster the characteristic distance converges, the higher the search efficiency.

$$CD = POD \times RD \quad (3)$$

## 3. SAR Environment Modeling

### 3.1. Environment Model

The environment is formulated as 2D grid worlds of size  $m \times m \in N^2$  with a cell size of  $c$  and a set of all possible positions  $M$ . The buildings and walls that the UGV cannot occupy are given by the set

$$W = \left\{ \left[ x_i^W, y_j^W \right]^T : \left[ x_i^W, y_j^W \right]^T \in M \right\} \quad (4)$$

The traps where the UGV will be punished are given by the set

$$B = \left\{ \left[ x_i^B, y_j^B \right]^T : \left[ x_i^B, y_j^B \right]^T \in M \right\} \quad (5)$$

The environment can be described by functions  $f_1, f_2, \dots, f_e$ , and the eigenvector at position  $p \in M$  is

$$F_p = (f_1(p), f_2(p), \dots, f_e(p)) \quad (6)$$

The state of the target is described through its position:

$$p_{\text{target}} = [x_{\text{target}}, y_{\text{target}}]^T \in M - W \quad (7)$$

The target tends to appear in positions within the feature interval, such as

$$F_{\text{low}} \leq F_{p_{\text{target}}} \leq F_{\text{up}} \quad (8)$$

The UGV moves within the limits of the grid world, and the state of the UGV is described through the following:

- Its position;
- Its operational status, either inactive or active;

- Its field of view (FOV), the grid world size with the cell size, and the set of all possible positions.

The distance that the UGV walks in a mission’s online time slot  $\delta_t$  is equivalent to the cell size  $c$ . The position of the UGV evolves according to the motion model given by

$$p(t+1) = \begin{cases} p(t) + a(t), & \varphi(t) = 1 \\ p(t), & \text{otherwise} \end{cases} \quad (9)$$

The evolution of the operational status  $\varphi(t)$  of the UGV is given by

$$\varphi(t+1) = \begin{cases} 0, & \varphi(t) = 0 \vee p(t) = p_{\text{target}} \\ 1, & \text{otherwise} \end{cases} \quad (10)$$

The end time point  $T$  is defined as the time slot when the UGV reaches its terminal state without actively operating. The following constraints restrict the UGV mobility model:

$$\begin{cases} p(t) \notin W, & (a) \\ t \leq T, & (b) \\ \varphi(0) = 1, & (c) \end{cases} \quad (11)$$

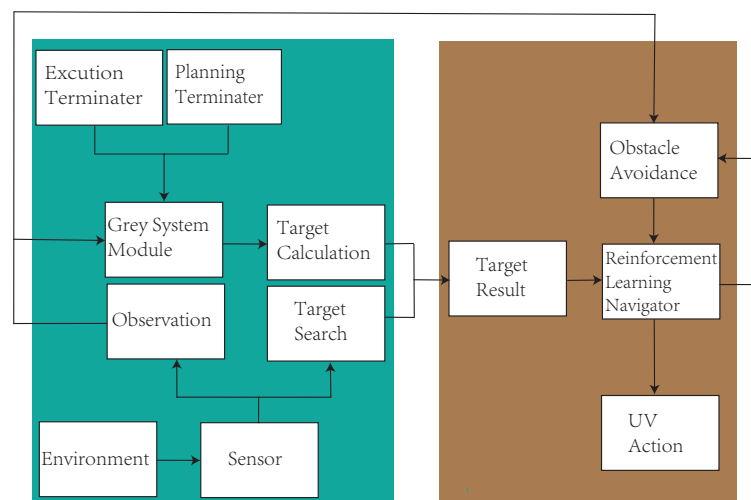
### 3.2. Optimization Problem

Using the environment model, the central aim of the UGV path-planning problem is to search the position of the target, given as  $H$ , while adhering to mobility constraints (8) and (11). The optimal problem is given over joint actions  $\times a(t)$  by

$$\begin{aligned} & \max_{\times a(t)} H \\ & \text{s.t. (8)(11)} \end{aligned} \quad (12)$$

## 4. Search Area Path-Planning Model for Mine SAR

Considering practical applications, the model work is performed as a two-step mapping process based on the visual and position sensors. Due to the state transition of the agent with the Markov property, the POMDP can provide a general framework [18]. Based on the above processes, the workflow of the gray online method is shown in Figure 1.



**Figure 1.** The workflow of the gray online method.

### 4.1. Local Path-Planning Reward Function

The state space  $s(t) \in S$  of search area path-planning problem consists of the environment and the agent. Meanwhile, the states in the observation space  $\Omega$  are given as

$f_{\text{map}} = f_{\text{information}} \times f_{\text{position}} : \text{Environment} \rightarrow \Omega$  and  $f_{\text{position}} \in N^2$  and  $f_{\text{information}} \in R$ . Observations  $o(t)$  ( $5 \times 5$  FOV in mission) are defined through the tuple

$$o(t) = (M(t), V(t), W(t), B(t), p_{\text{target}}(t), \{\varphi(t)\}) \tag{13}$$

A security controller is introduced into the system to implement an obstacle avoidance constraint (a). The security controller evaluates the point  $V_{i,j}(t)$  in the map and determines whether it can be reached. If not, it treats this point as  $\emptyset$ .

$$V_{ij}(t) = \begin{cases} \emptyset, & (x_i, y_j) \in W \\ V_{ij}(t), & \text{otherwise} \end{cases} \tag{14}$$

In addition, a safety controller evaluates the agent's action  $a(t)$  and determines whether the action should be punished. Based on the conclusion, the corresponding safety penalty value  $\beta$  is offered:

$$\beta(t) = \begin{cases} -1, & p(t) + a(t) \in B \\ 0, & \text{otherwise} \end{cases} \tag{15}$$

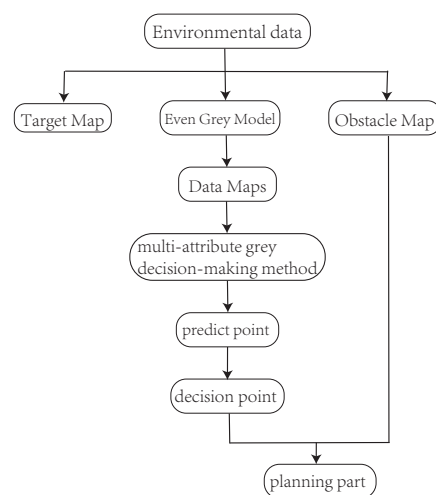
The reward function  $R : S \times A \times S \rightarrow R$  of the POMDP is defined as

$$r(t) = \tau \times \Phi(t) + \beta(t) + \varepsilon \tag{16}$$

The agent reaches a new state  $s(t + 1)$  which may depend on the state  $s(t)$  and action  $a(t)$ , and updates value Q.

#### 4.2. Environmental Data Prediction Process

In order to highlight the attribute of the uncertain information, the environment data are treated as gray numbers [37], that is, more attention is paid to the nature of the data than the values. Therefore, it is necessary to preprocess the data. However, the prediction system with limited information may be affected by interference. Firstly, for the information system with the gray attribute, the gray number and buffer operator [38] are combined to describe the data of the information system. Secondly, the data analyzed by the gray algorithm are preprocessed by the even gray model (as shown in Figure 2).



**Figure 2.** Environmental data prediction processing.

The FOV data sequence  $d^{(0)}(k)$  is described by generating the function  $G(o(k))$ , and its one time accumulating generation operator(1-AGO) sequence  $D^{(1)}$  can be calculated as

$$D^{(1)} = (d^{(1)}(1), d^{(1)}(2), \dots, d^{(1)}(n)) \tag{17}$$

Thus, the sequence  $Z^{(1)}$

$$Z^{(1)} = (z^{(1)}(1), z^{(1)}(2), \dots, z^{(1)}(n)) \tag{18}$$

can be obtained from

$$z^{(1)}(k) = \frac{1}{2}(d^{(1)}(k) + d^{(1)}(k - 1)) \tag{19}$$

So, the even gray model of the environmental data prediction process is

$$d^{(0)}(k) + az^{(1)}(k) = b \tag{20}$$

in which  $-a$  is the development coefficient and  $b$  is the gray actuating quantity. The prediction sequence is the basis for the subsequent algorithm to determine the optimal decision.

### 4.3. The Multi-Attribute Gray Decision Process

This paper presents a multi-attribute gray decision-making method to find solutions in exploration and development to enable UGVs to independently complete search tasks. In the new model, the following constraints of the observation space are regarded: feature constraints, update constraints, exclusion constraints, and attraction constraints. Four data maps are formed from the constraints, and the target map and the obstacle map constitute a complete observation space.

The feature constraint  $U^{(1)} = (u_{ij}^{(1)})$  is regarded as an interval number, and is a moderate-type objective with the upper limit  $G_{\max}$  and lower limit  $G_{\min}$  assumed.

When  $u_{ij}^{(1)} \in [G_{\min}, \frac{1}{2}(G_{\max} + G_{\min})]$ , the lower effect measurement function for a moderate objective is

$$r_{ij}^{(1)} = \frac{2(u_{ij}^{(1)} + G_{\min})}{G_{\max} - G_{\min}} \tag{21}$$

When  $u_{ij}^{(1)} \in [\frac{1}{2}(G_{\max} + G_{\min}), G_{\max}]$ , the upper effect measurement function for a moderate objective is

$$r_{ij}^{(1)} = \frac{2(G_{\max} - u_{ij}^{(1)})}{G_{\max} - G_{\min}} \tag{22}$$

The update constraint  $U^{(2)} = (u_{ij}^{(2)})$  is a benefit-type objective; the more significant the effect sample value, the better. The effect measurement function for the benefit-type objective is

$$r_{ij}^{(2)} = \frac{u_{ij}^{(2)} - \frac{5}{2}}{\max_i \max_j \{u_{ij}^{(2)}\} - \frac{5}{2}} \tag{23}$$

The exclusion constraint  $U^{(3)} = (u_{ij}^{(3)})$  is a benefit-type objective. The effect measurement function for this objective is

$$r_{ij}^{(3)} = \frac{u_{ij}^{(3)}}{\max_i \max_j \{u_{ij}^{(3)}\}} \tag{24}$$

Attraction constraint reflects the agent's preference for the map after starting. If the agent has no preference, the map should be searched evenly. The attraction constraint  $U^{(4)} = (u_{ij}^{(4)})$  is a cost-type objective. The effect measurement function for this objective is

$$r_{ij}^{(4)} = \frac{u_{ij}^{(4)}}{\min_i \min_j \{u_{ij}^{(4)}\}} \quad (25)$$

Furthermore, the decision weight of the  $k$ -th objective is  $\eta_k (k = 1, 2, 3, 4)$ . The uniform effect measurement matrix of the decision strategy  $s_{ij} \in S$  under the  $k$ -th objective is  $R^{(k)} = (r_{ij}^{(k)})$ . The synthetic effect measurement of decision strategy  $s_{ij}$  is

$$r_{ij} = \sum_{k=1}^s \eta_k r_{ij}^{(k)} \quad (26)$$

The synthetic effect measurement matrix is  $R = (r_{ij})$ . Using the synthetic effect measure matrix, effect values of different significance, dimensions, and characters can be comprehensively considered to determine the optimal decision strategy.

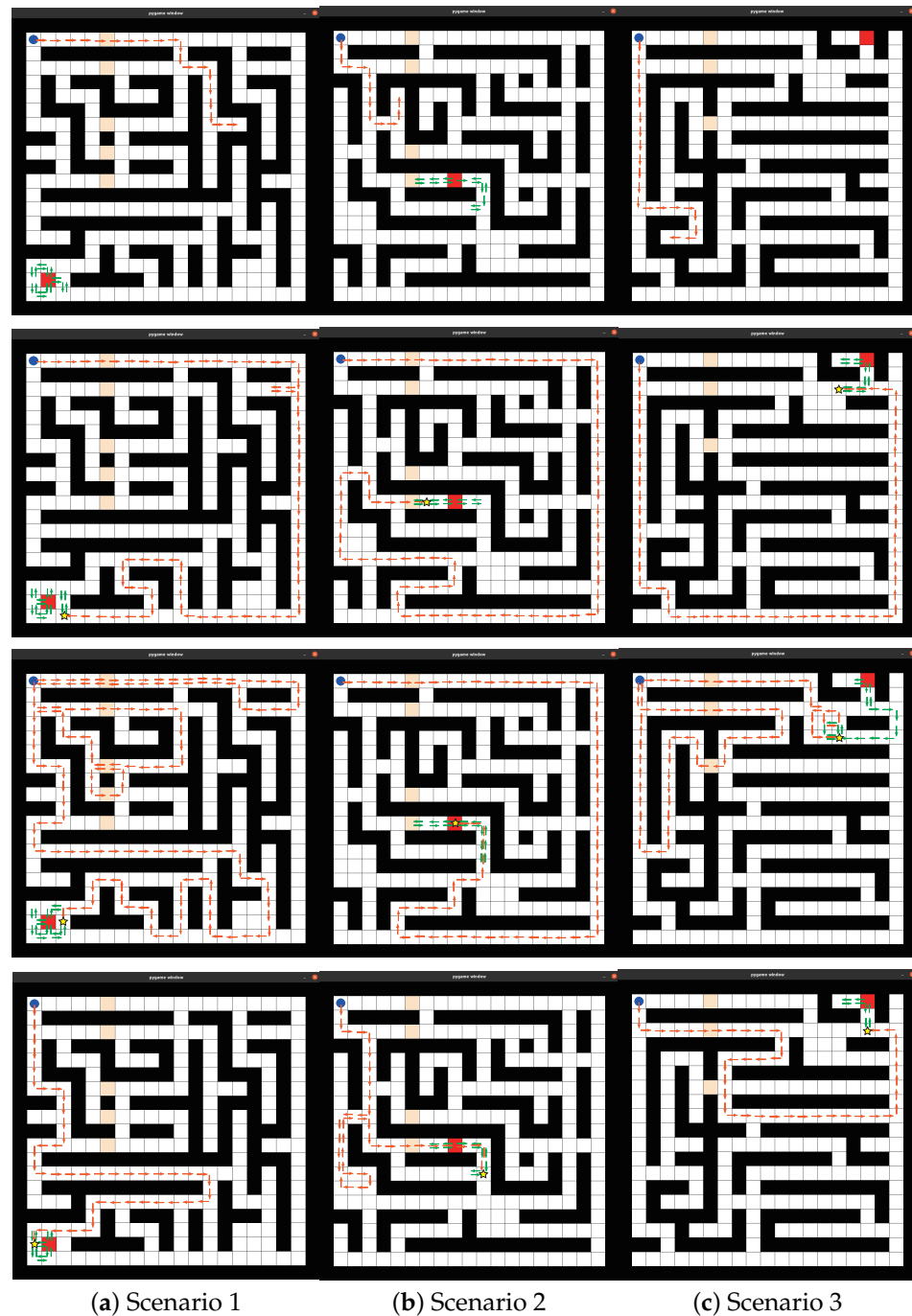
## 5. Experimental Verification and Result Analysis

### 5.1. Simulating Scenarios and Experimental Setup

The environment of the simulated case models and randomly selects three scenarios based on the mining SAR environment model. In order to verify the ability of the UGV agent to avoid obstacles online, the scene is randomly set by a map generation code (based on the code in [19]), and only kept consistent in the comparative experiment. Because the experiment involves online path-planning verification, the UGV does not conduct offline learning in advance. All algorithms in this paper are coded and implemented in Python version 3.9. The training environment for this model is Intel Core i7-4710MQ CPU @2.50 GHz four cores Chinese-made laptop.

### 5.2. Simulation and Comparison Experiment

The model is compared with the A\* search algorithm, RRT algorithm, and QL algorithm, while the rolling online algorithm is used to supplement the traditional algorithm. The comparison results are shown in Figure 3. In order to adapt to a four-way proxy, the A\* search algorithm combines a linear function and the Manhattan distance as the heuristic function. The RRT algorithm selects the nearest node to join the tree after obstacle judgment. The QL algorithm combines the average rolling algorithm to complete the test, which is different from the proposed method. All algorithms only search online on the map without being trained beforehand. From Figure 3, it can be seen that the A\* search algorithm cannot adapt to real-time scenarios, and the agent is easily trapped in deadlock situations, resulting in its inability to complete tasks. Due to the lack of a target decision process, the generalization ability of the RRT algorithm could be improved in different scenarios, and the algorithm often cannot continue due to stuttering during runtime. The QL algorithm and the proposed method can complete planning and produce coverage search paths, but the proposed method performs better when searching for the target.

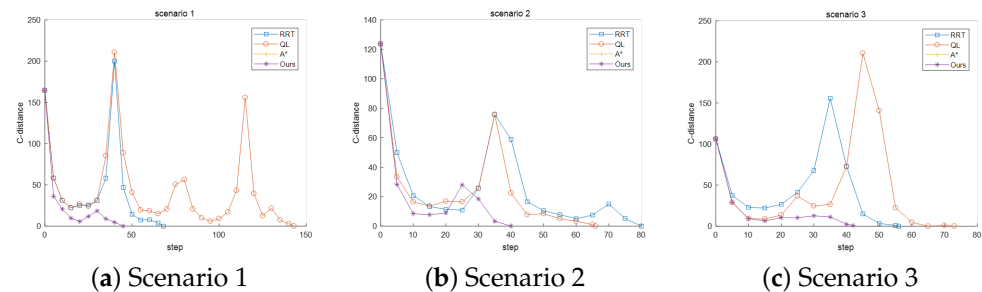


**Figure 3.** The planning results of different algorithms in three scenarios (from top to bottom: A\* algorithm, RRT algorithm, QL algorithm, and our algorithm). The orange arrow in the figure represents the search path. The green arrow represents the movement of the target. The yellow star represents the final capture position.

Statistical analyses are conducted on the path-planning results for final coverage, repeated coverage, and step size to evaluate the above algorithms. Table 1 shows the quantitative evaluation results. It shows that the search path produced by the proposed method performs better when searching for the target. In addition, the algorithm is evaluated from the perspective of searchability. As shown in Figure 4, the proposed method has the fastest convergence speed. This indicates that the agent in the proposed method catches the target preferentially. Traditional path-planning algorithms are prone to becoming stuck in local optima, especially when dealing with autonomous navigation tasks



in unknown maps. In contrast, reinforcement learning algorithms have more substantial autonomous planning capabilities. Our improved algorithm has a faster convergence speed in the autonomous planning process and can effectively handle searching for dynamic targets in unknown maps.



**Figure 4.** Comparison of changes in CD values between different algorithms. (a–c) Numerical statistical results of CD of four algorithms in Scenario 1, Scenario 2, and Scenario 3, measuring ability to search for target.

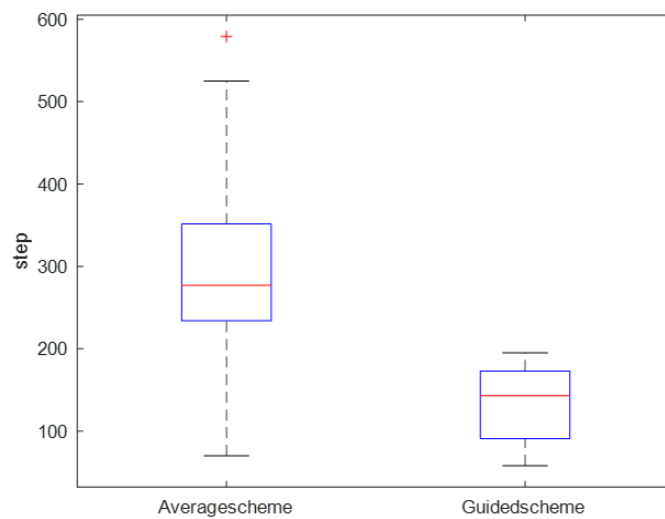
**Table 1.** Quantitative evaluation of planning results for different scenarios.

Scenario	Algorithms	Step	Coverage (%)	Repeated Coverage (%)
Scenario 1	A*	-	9.81	-
	RRT	68	31.31	1.40
	QL	143	57.94	9.35
	Ours	45	21.50	0
Scenario 2	A*	-	5.96	-
	RRT	80	37.16	0
	QL	66	30.73	0
	Ours	40	15.60	3.21
Scenario 3	A*	-	9.95	-
	RRT	56	27.01	0
	QL	73	32.23	2.37
	Ours	42	20.38	0

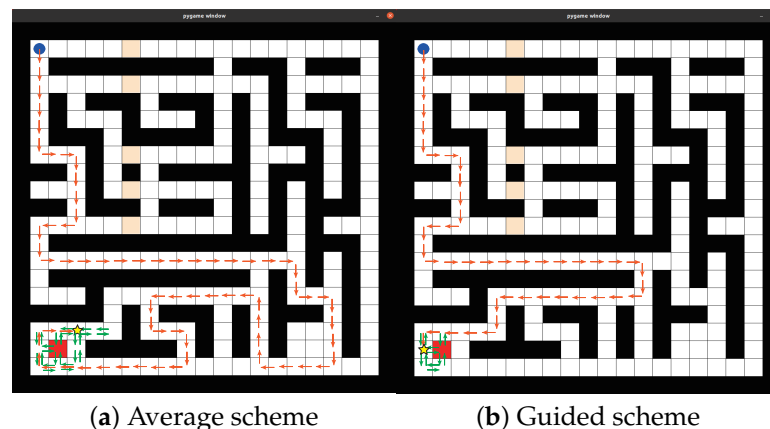
### 5.3. Policy Evaluation

According to the planning results of sub-section A, the improved algorithm is smaller than the conventional QL planning algorithm in terms of running step size. In addition, the QL algorithm has a high repetition coverage rate. The target decision algorithm can play a good role in assisting with decision making. To evaluate whether the corresponding parameters of the target decision algorithm can effectively assist in the decision problem when the target position is uncertain, 20 simulation experiments were carried out in the mine SAR scenario in Figure 5. Figure 5 shows that the median set of step size of the guidance scheme is smaller than that of the average scheme. In addition, the step sizes of guidance schemes is concentrated below 200, and the data fluctuation is not significant. In contrast, the step sizes of the average scheme are discretely distributed between 225 and 375, resulting in significant data fluctuation. The above results illustrate that the guided solution has a better adjustment effect on the target decision algorithm.

The search results under different schemes are shown in Figure 6. In the early and medium stages, all schemes remain explored and developed to obtain the best search path. In the later stage, however, the average scheme will still maintain a high probability of exploration, reducing data use. The above results indicate that improved policy has advantages from the perspective of action selection.



**Figure 5.** A comparison of the step size values of UGV agents successfully searching for targets under different scheme settings. The black line in the figure represents the edge value, the blue line represents the quartile value, and the red line represents the median value.



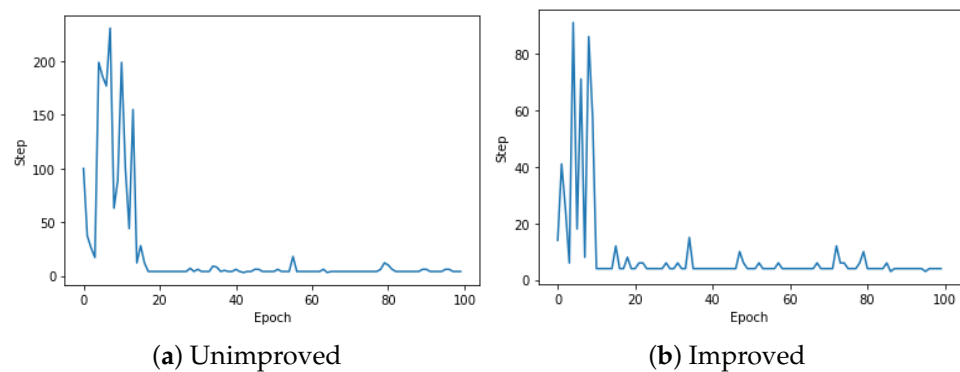
**Figure 6.** The results of path planning for different schemes. The orange arrow in the figure represents the search path. The green arrow represents the movement of the target. The yellow star represents the final capture position.

#### 5.4. Reward Function Evaluation

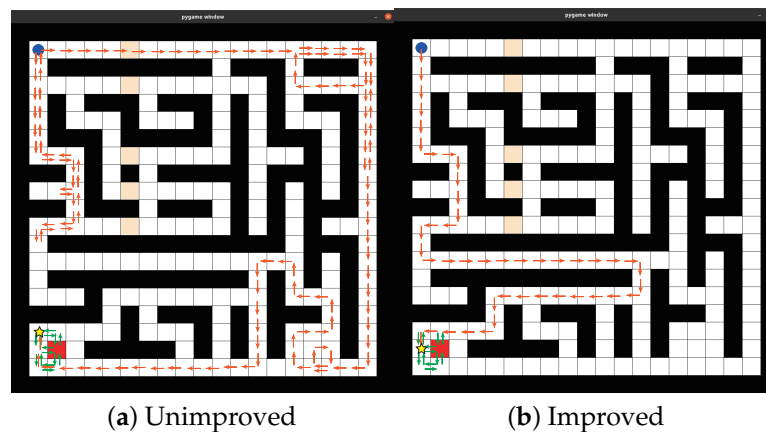
The data for different solutions are shown in Table 2. Figure 7 shows the performance of planning operation steps under different reward function solutions. The same curve trend can be observed between the unimproved and improved solutions, indicating that these two solutions play a role in solving obstacle avoidance and goal orientation. However, the convergence speed of the improved solution is faster. Meanwhile, the search path-planning results under different schemes are shown in Figure 8. During the learning process, the learning speed of the unimproved solution is slower, resulting in inaccurate planning multiple times. The improved solution provides more accurate results than the unimproved one during the learning process. It reduces the repetitive path rate of the agent and improves the efficiency of online planning. This proves that the reward function is goal-oriented for the intelligent agent, enabling it to learn gradually.

**Table 2.** The principle parameters of the target-tracking algorithm.

Solution	Parameter	Value
Improved Solution	learning rate	0.8
	reward decay	0.9
	$\epsilon$ greedy	0.9
	$\Phi(t)$	5
	$\beta(t)$	0
Unimproved Solution	learning rate	0.4
	reward decay	0.8
	$\epsilon$ greedy	0.9
	$\Phi(t)$	1
	$\beta(t)$	0



**Figure 7.** A comparison of different solutions in Table 2 in terms of training effectiveness. (a) The change process of training steps in the unimproved scheme. (b) The change process of training steps in the improvement plan.



**Figure 8.** The results of path planning for different solutions. The orange arrow in the figure represents the search path. The green arrow represents the movement of the target. The yellow star represents the final capture position.

**6. Conclusions and Future Work**

This paper studies the online path planning of mine SAR. Based on simulating an SAR environment using the characteristics of a mining cave field, a path-planning model for POMDP-based online search areas is proposed. The gray objective decision function guides this model and plans the optimal search path through information analysis, making it more flexible and intelligent. The comparative experimental results show that the model can complete target search tasks online in different maps of the same region and generate search paths to achieve faster convergence to dynamic target positions.

The future potential trends that can be integrated with the planning issues of UGVs will include, but are not limited to, the following:

- The interaction of UGVs with Cyber–Physical systems could enhance UGV’s adaptability to dynamic environments and validate the effectiveness and feasibility of path-planning algorithms through the use of 3D virtual reality models [39].
- The strategic application of human–machine interaction technology is not just a possibility, but a promising avenue that is poised to play a pivotal role in the rapid advancement of UGV intelligence [40].
- Swarm Intelligence technology can expand UGVs’ application scope and significantly improve its efficiency in completing tasks [41].
- Another important research direction to enhance the autonomy of UGVs is to combine them with Internet of Things technology to improve their flexibility in different locations [42,43].

Given the aforementioned topics, further in-depth research will be carried out within the above-specified fields.

**Author Contributions:** Conceptualization, S.Z.; Methodology, S.Z.; Software, S.Z.; Validation, S.Z.; Writing—original draft, S.Z.; Writing—review & editing, Q.Z.; Supervision, Q.Z.; Project administration, Q.Z. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded in part by the National Natural Science Foundation of China under Grant 62188101 and 62303135, and in part by the Heilongjiang Touyan Team Program, and in part by the Harbin Institute of Technology.

**Institutional Review Board Statement:** Not applicable.

**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.

## References

1. Hester, G.; Smith, C.; Day, P.; Waldock, A. The Next Generation of Unmanned Ground Vehicles. *Meas. Control* **2012**, *45*, 117–121. [[CrossRef](#)]
2. Hu, J.; Niu, H.; Carrasco, J.; Lennox, B.; Arvin, F. Voronoi-Based Multi-Robot Autonomous Exploration in Unknown Environments via Deep Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2020**, *69*, 14413–14423. [[CrossRef](#)]
3. Niroui, F.; Zhang, K.; Kashino, Z.; Nejat, G. Deep Reinforcement Learning Robot for Search and Rescue Applications: Exploration in Unknown Cluttered Environments. *IEEE Robot. Autom. Lett.* **2019**, *4*, 610–617. [[CrossRef](#)]
4. Ai, B.; Li, B.; Gao, S.; Xu, J.; Shang, H. An Intelligent Decision Algorithm for the Generation of Maritime Search and Rescue Emergency Response Plans. *IEEE Access* **2019**, *7*, 155835–155850. [[CrossRef](#)]
5. Tao, X.; Lang, N.; Li, H.; Xu, D. Path Planning in Uncertain Environment with Moving Obstacles Using Warm Start Cross Entropy. *IEEE/ASME Trans. Mechatron.* **2022**, *27*, 800–810. [[CrossRef](#)]
6. Wang, C.; Zhang, X.; Li, R.; Dong, P. Path Planning of Maritime Autonomous Surface Ships in Unknown Environment with Reinforcement Learning. In *Communications in Computer and Information Science, Proceedings of the Cognitive Systems and Signal Processing, ICCSIP, Beijing, China, 29 November–1 December 2018*; Sun, F., Liu, H., Hu, D., Eds.; Springer: Singapore, 2018; Volume 1006. [[CrossRef](#)]
7. Zhang, X.; Wang, C.; Liu, Y.; Chen, X. Decision-Making for the Autonomous Navigation of Maritime Autonomous Surface Ships Based on Scene Division and Deep Reinforcement Learning. *Sensors* **2019**, *19*, 4055. [[CrossRef](#)]
8. Tatsch, C.; Bredu, J.A.; Covell, D.; Tulu, I.B.; Gu, Y. Rhino: An Autonomous Robot for Mapping Underground Mine Environments. In *Proceedings of the 2023 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM), Seattle, WA, USA, 28–30 June 2023*; pp. 1166–1173. [[CrossRef](#)]
9. Cao, Y.; Zhao, R.; Wang, Y.; Xiang, B.; Sartoretti, G. Deep Reinforcement Learning-Based Large-Scale Robot Exploration. *IEEE Robot. Autom. Lett.* **2024**, *9*, 4631–4638. [[CrossRef](#)]
10. Vlahov, B.; Gibson, J.; Fan, D.D.; Spieler, P.; Agha-mohammadi, A.-A.; Theodorou, E.A. Low Frequency Sampling in Model Predictive Path Integral Control. *IEEE Robot. Autom. Lett.* **2024**, *9*, 4543–4550. [[CrossRef](#)]
11. Luo, Y.; Zhuang, Z.; Pan, N.; Feng, C.; Shen, S.; Gao, F.; Cheng, H.; Zhou, B. Star-Searcher: A Complete and Efficient Aerial System for Autonomous Target Search in Complex Unknown Environments. *IEEE Robot. Autom. Lett.* **2024**, *9*, 4329–4336. [[CrossRef](#)]

12. Cheng, C.X.; Sha, Q.X.; He, B.; Li, G.L. Path planning and obstacle avoidance for AUV: A review. *Ocean. Eng.* **2021**, *235*, 109355. [[CrossRef](#)]
13. Peake, A.; McCalmon, J.; Zhang, Y.; Raiford, B.; Alqahtani, S. Wilderness Search and Rescue Missions using Deep Reinforcement Learning. In Proceedings of the 2020 IEEE International Symposium on Safety, Security, and Rescue Robotics (SSRR), Abu Dhabi, United Arab Emirates, 4–6 November 2020; pp. 102–107. [[CrossRef](#)]
14. Liu, C.; Zhao, J.; Sun, N. A Review of Collaborative Air-Ground Robots Research. *J. Intell. Robot. Syst.* **2022**, *106*, 60. [[CrossRef](#)]
15. Palacin, J.; Palleja, T.; Valganon, I.; Pernia, R.; Roca, J. Measuring Coverage Performances of a Floor Cleaning Mobile Robot Using a Vision System. In Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, 18–22 April 2005; pp. 4236–4241. [[CrossRef](#)]
16. Ai, B.; Jia, M.X.; Xu, H.W.; Xu, J.L.; Wen, Z.; Li, B.S.; Zhang, D. Coverage path planning for maritime search and rescue using reinforcement learning. *Ocean. Eng.* **2021**, *241*, 110098. [[CrossRef](#)]
17. Sun, Y.; Fang, Z. Research on Projection Gray Target Model Based on FANP-QFD for Weapon System of Systems Capability Evaluation. *IEEE Syst. J.* **2021**, *15*, 4126–4136. [[CrossRef](#)]
18. Ross, S.; Pineau, J.; Paquet, S.; Chaib-Draa, B. Online planning algorithms for POMDPs. *IEEE Robot. Autom. Lett.* **2008**, *32*, 663–704. [[CrossRef](#)]
19. Sartoretti, G.; Kerr, J.; Shi, Y.; Wagner, G.; Kumar, T.K.S.; Koenig, S.; Choset, H. PRIMAL: Pathfinding via Reinforcement and Imitation Multi-Agent Learning. *IEEE Robot. Autom. Lett.* **2019**, *4*, 2378–2385. [[CrossRef](#)]
20. Wang, C.; Cheng, J.; Wang, J.; Li, X.; Meng, M.Q.-H. Efficient Object Search With Belief Road Map Using Mobile Robot. *IEEE Syst. J.* **2018**, *15*, 3081–3088. [[CrossRef](#)]
21. Agha-mohammadi, A.-A.; Agarwal, S.; Kim, S.-K.; Chakravorty, S.; Amato, N.M. SLAP: Simultaneous Localization and Planning Under Uncertainty via Dynamic Replanning in Belief Space. *IEEE Trans. Robot.* **2018**, *34*, 1195–1214. [[CrossRef](#)]
22. Hubmann, C.; Schulz, J.; Xu, G.; Althoff, D.; Stiller, C. A Belief State Planner for Interactive Merge Maneuvers in Congested Traffic. In Proceedings of the 2018 21st International Conference on Intelligent Transportation Systems (ITSC), Maui, HI, USA, 4–7 November 2018; Volume 1, pp. 1617–1624. [[CrossRef](#)]
23. Hubmann, C.; Becker, M.; Althoff, D.; Lenz, D.; Stiller, C. Decision making for autonomous driving considering interaction and uncertain prediction of surrounding vehicles. In Proceedings of the 2017 IEEE Intelligent Vehicles Symposium (IV), Los Angeles, CA, USA, 11–14 June 2017; Volume 1, pp. 1671–1678. [[CrossRef](#)]
24. Bai, A.; Wu, F.; Chen, X. Posterior sampling for Monte Carlo planning under uncertainty. *Appl. Intell.* **2018**, *48*, 4998–5018. [[CrossRef](#)]
25. Liu, P.; Chen, J.; Liu, H. An improved Monte Carlo POMDPs online planning algorithm combined with RAVE heuristic. In Proceedings of the 2015 6th IEEE International Conference on Software Engineering and Service Science (ICSESS), Beijing, China, 23–25 September 2015; Volume 1, pp. 511–515. [[CrossRef](#)]
26. Xiao, Y.; Katt, S.; ten Pas, A.; Chen, S.; Amato, C. Online Planning for Target Object Search in Clutter under Partial Observability. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA), Montreal, QC, Canada, 20–24 May 2019; Volume 1, pp. 8241–8247. [[CrossRef](#)]
27. Bayerlein, H.; Theile, M.; Caccamo, M.; Gesbert, D. Multi-UAV Path Planning for Wireless Data Harvesting With Deep Reinforcement Learning. *IEEE Robot. Autom. Lett.* **2021**, *2*, 1171–1187. [[CrossRef](#)]
28. Bhattacharya, S.; Badyal, S.; Wheeler, T.; Gil, S.; Bertsekas, D. Reinforcement Learning for POMDP: Partitioned Rollout and Policy Iteration with Application to Autonomous Sequential Repair Problems. *IEEE Open J. Commun. Soc.* **2020**, *5*, 3967–3974. [[CrossRef](#)]
29. Yan, P.; Jia, T.; Bai, C.; Fravolini, M.L. Searching and Tracking an Unknown Number of Targets: A Learning-Based Method Enhanced with Maps Merging. *Sensors* **2021**, *21*, 1076. [[CrossRef](#)]
30. Amato, C.; Konidaris, G.; Cruz, G.; Maynor, C.A.; How, J.P.; Kaelbling, L.P. Planning for decentralized control of multiple robots under uncertainty. In Proceedings of the 2015 IEEE International Conference on Robotics and Automation (ICRA), Seattle, WA, USA, 26–30 May 2015; Volume 1, pp. 1241–1248. [[CrossRef](#)]
31. MacDonald, R.A.; Smith, S.L. Active sensing for motion planning in uncertain environments via mutual information policies. *Int. J. Robot. Res.* **2019**, *38*, 146–161. [[CrossRef](#)]
32. He, Y.; Chong, K.P. Sensor scheduling for target tracking in sensor networks. In Proceedings of the 2004 43rd IEEE Conference on Decision and Control (CDC) (IEEE Cat. No.04CH37601), Nassau, Bahamas, 14–17 December 2004; Volume 1, pp. 743–748. [[CrossRef](#)]
33. Gerrig, R.J.; Zimbardo, P.G. *Psychology and Life*; People's Posts and Telecommunications Press: Beijing, China, 2011; ch. 5, ses. 3, pp. 114–117.
34. Duchoň, F.; Babinec, A.; Kajan, M.; Beňo, P.; Florek, M.; Fico, T.; Jurišica, L. Path Planning with Modified a Star Algorithm for a Mobile Robot. *Procedia Eng.* **2014**, *96*, 59–69. [[CrossRef](#)]
35. Rodriguez, S.; Tang, X.Y.; Lien, J.M.; Amato, N.M. An obstacle-based rapidly-exploring random tree. In Proceedings of the 2006 IEEE International Conference on Robotics and Automation, Orlando, FL, USA, 15–19 May 2006; Volume 1, pp. 895–900. [[CrossRef](#)]
36. Konar, A.; Chakraborty, I.G.; Singh, S.J.; Jain, L.C.; Nagar, A.K. A Deterministic Improved Q-Learning for Path Planning of a Mobile Robot. *IEEE Trans. Syst. Man Cybern. Syst.* **2013**, *43*, 1141–1153. [[CrossRef](#)]
37. Liu, S.F. The Three Axioms of Buffer Operator and Their Application. *J. Grey Syst.* **1991**, *1*, 178–185.

38. Wei, Y.; Kong, X.H.; Hu, D.H. A kind of universal constructor method for buffer operators. *Grey Syst. Theory Appl.* **2011**, *3*, 39–48.
39. Cecil, J. A conceptual framework for supporting UAV based cyber physical weather monitoring activities. In Proceedings of the 2018 Annual IEEE International Systems Conference (SysCon), Vancouver, BC, Canada, 23–26 April 2018; pp. 1–8. [[CrossRef](#)]
40. Zhu, S.; Xiong, G.; Chen, H. Unmanned Ground Vehicle Control System Design Based on Hybrid Architecture. In Proceedings of the 2019 IEEE 8th Joint International Information Technology and Artificial Intelligence Conference (ITAIC), Chongqing, China, 24–26 May 2019; pp. 948–951. [[CrossRef](#)]
41. AlShabi, M.; Ballous, K.A.; Nassif, A.B.; Bettayeb, M.; Obaideen, K.; Gadsden, S.A. Path planning for a UGV using Salp Swarm Algorithm. In Proceedings of the SPIE 13052, Autonomous Systems: Sensors, Processing, and Security for Ground, Air, Sea, and Space Vehicles and Infrastructure 2024, 130520L, National Harbor, MD, USA, 7 June 2024. [[CrossRef](#)]
42. Romeo, L.; Petitti, A.; Colella, R.; Valecce, G.; Boccadoro, P.; Milella, A.; Grieco, L.A. Automated Deployment of IoT Networks in Outdoor Scenarios using an Unmanned Ground Vehicle. In Proceedings of the 2020 IEEE International Conference on Industrial Technology (ICIT), Buenos Aires, Argentina, 26–28 February 2020; pp. 369–374. [[CrossRef](#)]
43. Chang, B.R.; Tsai, H.-F.; Lyu, J.-L.; Huang, C.-F. IoT-connected Group Deployment of Unmanned Vehicles with Sensing Units: IUAGV System. *Sens. Mater.* **2021**, *33*, 1485–1499. [[CrossRef](#)]

**Disclaimer/Publisher’s Note:** The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.