SCIENTIFIC Publishing Limited

Journal of Intelligent Communication

Article

Multi UAV Cooperative Reconnaissance based on Dynamic Programming VDN Algorithm

Jingyi Huang, Ziyi Yang, Jiarui Li, Shuying Wu, Xinyu Zhang and Bo Li *

School of Electronics and Information, Northwestern Polytechnical University, Xi'an 710072, China
* Correspondence: libo803@nwpu.edu.cn

Received: 19 March 2024; **Revised**: 30 April 2024; **Accepted**: 17 May 2024; **Published**: 31 May 2024

Abstract: This paper proposes a multi agent value decomposition network (VDN) based multi UAV collaborative reconnaissance and control method to address the issue of insufficient strategies for multi UAV collaborative reconnaissance and control. By designing corresponding algorithm networks and training processes, the goal of autonomy, collaboration, and intelligence among multiple unmanned aerial vehicle systems has been achieved, assisting unmanned aerial vehicle combat forces in achieving collaborative operations and decision-making. This article uses AirSim as the simulation verification environment to verify the effectiveness of the proposed algorithm. The experimental results show that the algorithm proposed in this paper can achieve multi UAV collaborative reconnaissance tasks in complex environments, providing an intelligent solution for UAV collaborative control.

Keywords: reinforcement learning; dynamic programming; UAV collaborative reconnaissance; VDN algorithm

1. Introduction

With the development of technology, drones can be seen in various aspects of civil and military applications. Drones have shown excellent performance in aerial photography, distributed positioning, 3D reconstruction, as well as military early warning, regional search [1], electronic countermeasures [2], etc., completing various tasks at a lower cost. Drone technology has become one of the core technologies for the development of aviation industry both domestically and internationally [3]. Drones play a more important role in air combat, not only for adversarial purposes, but also for reconnaissance, maneuvering tracking, and decision-making games [4]. In modern warfare, drones have become an indispensable reconnaissance tool on the battlefield. But with the complexity of warfare and the demand for higher reconnaissance efficiency, single drone systems are no longer able to meet the needs of complex warfare reconnaissance tasks. However, with the emergence of multi drone system collaborative control technology, multi drone systems have gradually replaced single drone systems to take over reconnaissance tasks. Multi drone mission planning technology is a technology for reconnaissance cooperation among multiple drones. Currently, multi drone collaborative reconnaissance mission planning technology mainly includes the following two fields according to the different reconnaissance objects: "point-to-point" collaborative reconnaissance and "point-to-point" collaborative reconnaissance [5,6]. "Point to point" collaborative reconnaissance refers to unmanned aerial vehicles (UAVs) conducting collaborative reconnaissance on target points, while "point-to-point" collaborative reconnaissance refers to UAVs conducting collaborative reconnaissance on large target areas. During the process of target reconnaissance, multiple UAV systems try to minimize the cost and efficiency of reconnaissance. In addition, in complex battlefield environments, multiple unmanned aerial vehicle systems need to continuously perceive the environment and

https://doi.org/10.54963/jic.v4i1.238

complete wireless communication and collaborative positioning [7]. At the same time, enemy targets may have better maneuverability, which places higher demands on the autonomy, collaboration, and intelligence of drone maneuvering decisions. To complete the collaborative reconnaissance mission of unmanned aerial vehicles, it is necessary to break through three key technologies: collaborative perception, collaborative task allocation, and collaborative trajectory planning.

1.1. Collaborative Perception Technology

Collaborative perception technology is an important foundation for drone clusters to complete various tasks. Drones use perception technology to "understand" the environment, and the core functions of drone collaborative perception are target detection, recognition, positioning, and tracking [8]. For example, in response to rejection environments, in order to meet the optimal maneuvering needs of unmanned aerial vehicles in various mission environments, the Defense Advanced Research Projects Agency (DARPA) of the United States has proposed intelligent situational awareness and target detection, recognition, and tracking technology based on aerial vision [9]; Li Congcong et al. studied the applicability of various sensors in the degraded visual environment and designed the data frame format transmitted by UAV to improve the mutual fusion rate of images from different sensors [10]. The experimental results show that the mutual fusion image technology based on the data frame format transmitted by UAV can effectively improve the target recognition ability; The Merino L team proposed a collaborative perception system suitable for various heterogeneous drone clusters, which can automatically detect and locate targets [11].

The collaborative perception process can be divided into three stages: information acquisition, information fusion, and intelligence analysis [12]. In the process of multi drone combat, each drone selects one or more sensors, such as cameras, radars, etc., based on the battlefield environment to obtain information from the target. After obtaining the raw data of the target, drones need to process information. Multi drone systems integrate data and information from multiple sources to obtain accurate location and identity information of the target. This process is called information fusion, which is a process of refining the obtained information. The multi drone system uses artificial intelligence and other technologies to perform intelligence analysis on refined data, interpret the knowledge contained in the data, and complete the collaborative perception process. However, there are still some unresolved issues, first of which is the inefficient utilization of combined sensor information. Secondly, there is a lack of collaborative trajectory optimization across multiple platforms. In addition, further research is needed on efficiency evaluation techniques.

1.2. Collaborative Task Allocation Technology

Multi drone collaborative task allocation is guided by target value, taking into account the number of drones, flight performance, and types of resources carried, and reasonably allocating the targets to be executed to multiple drones, achieving reasonable scheduling of combat tasks and optimizing task execution [13].

Currently, many collaborative task allocation methods for unmanned aerial vehicles have been developed. Traditional algorithms such as branch and bound, dynamic programming, deep search, etc. [14,15], but these algorithms face problems such as long solving time and difficulty in solving multi constraint tasks when facing large-scale unmanned aerial vehicle task allocation problems. Therefore, researchers currently mostly use intelligent task allocation algorithms, such as using the traveling salesman problem solving method to solve the multi drone collaborative task allocation problem, treating the multi task problem as a multi traveling salesman problem, adding virtual target locations in the calculation process, and then decomposing the multi traveling salesman problem into single traveling salesman problems to solve them one by one, ultimately solving the feasible flight paths of each drone [16]. Kang Xuchao et al. proposed the discrete firefly algorithm to solve the task allocation problem of unmanned aerial vehicles [17]. By improving the firefly's movement mechanism, the convergence speed of the algorithm was improved, enabling the drone to quickly reach the target position. Other trajectory planning algorithms such as multi-layer coding genetic algorithm [18], improved genetic algorithm [19], etc.

1.3. Collaborative Trajectory Planning Technology

In terms of trajectory planning, mainstream methods both domestically and internationally include polygon region decomposition and efficient convergence algorithms [20], parallel region search methods [21], graphic

algorithms, and intelligent biomimetic algorithms. Graphics algorithms include A * algorithm [22], RRT algorithm [23], Voronoi graph algorithm [24], etc. Intelligent biomimetic algorithms include particle swarm optimization [25], genetic algorithm [26], ant colony algorithm [27], and reinforcement learning algorithm [28,29], among others. The decomposition based on polygons and efficient convergence methods are also applied to the trajectory calculation of small-scale heterogeneous drone clusters. Intelligent biomimetic algorithms have advantages over graphic computing in areas such as global search and parallel evolution, and can even handle some larger scale problems. In intelligent biomimetic algorithms, particle swarm optimization has the advantages of fast convergence, good robustness, and high efficiency. When based on particle swarm optimization, it is easier to combine with other algorithms [30].

In complex reconnaissance environments, drone operators require years of training to master excellent control capabilities over drones, which invisibly increases the difficulty and efficiency of executing drone flight missions in complex environments. Throughout the entire mission, each drone will face a complex external environment that may result in various malfunctions. The reconnaissance and flight decision-making of unmanned aerial vehicle systems in complex environments will be the focus of future research on unmanned aerial vehicles. This requires autonomous control technology to control unmanned aerial vehicles, achieve precise reconnaissance and perception in the environment, and achieve autonomous decision-making in reconnaissance. In complex environments, drones use image information, position information, and attitude information collected by sensors to make flight decisions. When the surrounding environment changes, drones need to identify obstacles, avoid external risks, and continue to complete flight missions [31].

Based on the advantages of using reinforcement learning to handle sequential decision-making problems, more and more researchers are incorporating reinforcement learning algorithms combined with deep learning into unmanned aerial vehicle reconnaissance decision-making problems. On the one hand, by enhancing the interaction between intelligent agents and the environment, unmanned aerial vehicles can perceive the environment and achieve navigation tasks such as path planning for unmanned systems. On the other hand, for different task backgrounds and environments, suitable reward functions can be designed as incentive signals for drone decision-making, helping drones to complete reconnaissance decision-making tasks autonomously through training. Zhao Yu et al. used deep reinforcement learning algorithms to construct a control model and a coordination mechanism between drones to control the collaborative flight of multiple fixed wing drones, and verified the effectiveness of collision avoidance during collaborative flight of multiple fixed wing drones; Fan Longtao et al. proposed a reinforcement learning method based on attention mechanism [32,33]. Firstly, a task allocation solution model was constructed through attention mechanism, and then the model was continuously optimized using reinforcement algorithm to obtain an approximate optimal solution.

At present, there are still shortcomings in the research of multi UAV collaborative reconnaissance and encirclement decision-making based on deep reinforcement learning:

a. In the process of research on unmanned aerial vehicle (UAV) collaborative reconnaissance decision-making based on deep reinforcement learning, there is a problem of relatively simple UAV modeling, which is relatively different from real UAV reconnaissance flight modeling. Most of them are implemented in two-dimensional space, with small decision action space and low decision-making difficulty.
b. In the research of reconnaissance decision-making, the task scenario is relatively simple, and the analysis of the perception process, task allocation process, navigation process, etc. in the reconnaissance process is insufficient. The environment and model are simplified, making it difficult to apply to complex and dynamic real reconnaissance scenarios.

Based on the above analysis, this paper proposes a cooperative control algorithm for multi UAV reconnaissance based on dynamic programming reinforcement learning multi-agent value decomposition networks (VDN) algorithm in complex environment. The tight coupling of drone target allocation tasks and trajectory planning is more in line with the demand for collaborative reconnaissance decision-making among multiple drones in complex environments, and is of great significance for realizing future multi-UAV, multi-manned/UAV collaborative operations and territorial defense.

2. Modeling of Multi UAV Collaborative Reconnaissance Tasks

2.1. Description of Multi UAV Collaborative Reconnaissance Tasks

The main question of the co-reconnaissance of Drum-Machinery is how to make multiple drones cooperate with each other to complete the value target reconnaissance tasks in the regional space. In this process, the drone needs to be decided and shared. To achieve the purpose of collaboration.

When studying from the top of the top to study the co-reconnaissance of the drone, you need to consider the coordination between each drone system in the multi-drone system. The form of mathematical modeling describes the interaction between the coordination between drones and the environment to obtain the feasible solution of the reconnaissance model and meet the requirements of many drones to collaborate with reconnaissance. As shown in Figure 1, the hierarchical description diagram is described from the top of the top of the top.

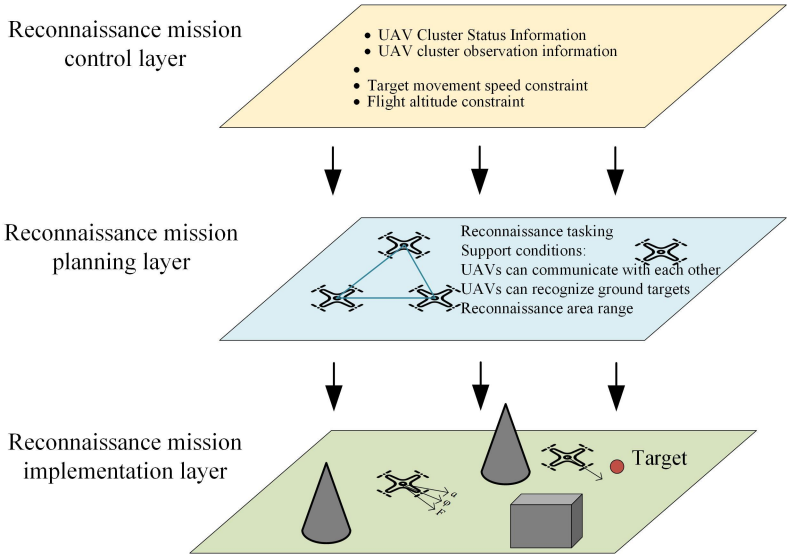


Figure 1. Top-down multi-UAV collaborative reconnaissance mission hierarchical description diagram.

2.2. Analysis of Multi-UAV Cooperative Reconnaissance Mission

2.2.1. UAV Flight Platform

When conducting regional collaborative reconnaissance, the drone that obtains the reconnaissance target after the reconnaissance mission is assigned needs to fly to the reconnaissance area within a certain period of time to complete the reconnaissance work. When the drone is not assigned a reconnaissance mission, the drone maintains the cruising altitude and cruising speed.

2.2.2. Detection Target

In the battlefield, there are two main types according to the value classification, the first is high-value targets, such as enemy combat command centers, ammunition depots, and cluster targets, and the second is low-value targets, such as man-portable units, isolated combat vehicles, and so on. In the reconnaissance process, the UAV may be reconnaissance target counter reconnaissance, in general, the threat level of high-value targets is higher than that of low-value targets, so the UAV needs to consider the threat level of different value targets in the process of reconnaissance, to avoid enemy radar counter reconnaissance, and to improve the probability of survival of its own side in the process of reconnaissance.

2.2.3. Reconnaissance Constraints

In the process of multi-UAV cooperative reconnaissance, only a single UAV can appear on the same spatial location point at the same moment, and the positional synergy of multiple UAVs in the whole combat space can

satisfy the requirements of multi-UAV cooperative reconnaissance in order to maximize the spatial utility of the multi-UAV system, which is known as the spatial constraint. In addition, other constraints, such as the UAV flight altitude constraint, i.e., the UAV cannot be lower than the flight safety altitude, the on-board sensor detection range constraint, energy constraints, etc., constitute the control conditions for multi-UAV cooperative reconnaissance missions.

2.3. Multi-UAV Coordinated Reconnaissance Mission Model

The position of the UAV and the reconnaissance target is defined in the Earth coordinate system $O_eX_eY_eZ_e$, which is used to study the process of relative changes in the positions of the UAV and the target. The airframe system $O_uX_uY_uZ_u$ is defined on the UAV airframe, the origin O_u is the position of the center of gravity of the UAV; the O_uX_u direction is the direction pointing to the nose in the plane of symmetry of the UAV; the O_uZ_u and O_uX_u axes are in the same plane perpendicularly O_uX_u downward, and the O_uY_u axis is determined according to the right-hand rule. The relationship between the Earth coordinate system and the airframe coordinate system is illustrated in Figure 2.

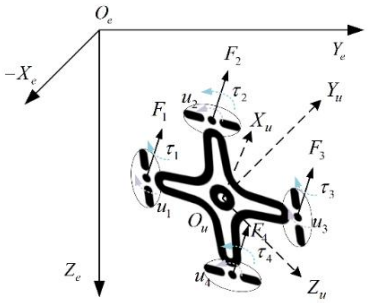


Figure 2. Relationship between the Earth coordinate system and the body coordinate system.

The multi-UAV cooperative reconnaissance process assumes that there are N UAVs and M reconnaissance targets, and there is a formal description of the reconnaissance target $T^j, j \in M$ for UAV $U^i, i \in N$, as shown in Equation (1).

$$U^i = \{v^e, a^e, \phi, \varphi, \psi, P^e, I_F\} \tag{1}$$

where, in the above equation, v^e, a^e denotes the Earth coordinate system velocity and acceleration of the UAV U^i, respectively, ϕ, φ, ψ denotes the roll angle $\phi \in [-\pi, \pi]$, pitch angle. $\varphi \in [-\pi/2, \pi/2]$., and yaw angle $\psi \in [-\pi, \pi]$, $\mathbf{P}^e$ denotes the Earth coordinate system position of the UAV, and $\mathbf{I}_F$ denotes the inter-aircraft datalink communication data.

The description of the reconnaissance target is shown in Equation (2).

$$T^j = \{\boldsymbol{v}^e, \mathbf{P}^e\} \tag{2}$$

where, in the above equation, $\boldsymbol{v}^e, \mathbf{P}^e$ denotes the target T^j Earth coordinate system velocity and position, respectively.

2.3.1. Flight Control Model

1. Physical modeling of drones

For computational convenience, the UAV model is considered as a rigid body with motor control inputs $\{\mu_1, \mu_2, \mu_3, \mu_4\}$, and the force $\{F_1, F_2, F_3, F_4\}$ and torque $\{\tau_1, \tau_2, \tau_3, \tau_4\}$ at the rotor vertices are generated according to the direction normal to its plane of rotation. As the physical model of the UAV shown in the body coordinate system in Figure 2, The tension and torque generated by each rotor of the drone can be calculated according to Equation (3).

$$F_n = C_{Fg}\rho\omega_{\max}^2 D_{pro}^4 u_n$$
$$\tau_n = \frac{1}{2\pi} C_{pow}\rho\omega_{\max}^2 D_{pro}^5 \mu_n \tag{3}$$

where C_{Fg} and C_{pow} are the thrust and power coefficients, respectively, ρ is the air density, D_{pro} is the propeller diameter, and $\omega_{\max}$ is the maximum angular velocity per minute, $n \in \{1,2,3,4\}$.

In order to solve the position and attitude information of the UAV in real time, this paper adopts the UAV flight control rigid body model, which includes the UAV kinematics and dynamics model.

2. UAV kinematic modeling

UAV kinematics model: including position kinematics model and attitude kinematics model, by inputting the values of velocity and angular velocity, the UAV kinematics model can solve the position and attitude of the UAV.

The position coordinate of the center of gravity of the drone U^i in the Earth coordinate system is $\mathbf{P}_e \in \mathrm{R}^3$, and the kinematic model of the drone's position is shown in the following Equation (4).

$$\dot{\mathbf{P}}^e = \boldsymbol{v}^e \tag{4}$$

The UAV attitude kinematics is shown in the following Equation (5).

$$\begin{cases} \dot{h}_0 = -\frac{1}{2}\,\mathbf{h}'_v\omega^u \\ \dot{\mathbf{h}}_v = \frac{1}{2}(q_0 I_3 \mathbf{h}_v)\omega^u \end{cases} \tag{5}$$

where $\omega^u \in \mathrm{R}^3$ is the angular velocity. $h_0 \in \mathrm{R}$ is the scalar part of the UAV quaternion, $\mathbf{h}_v \in \mathrm{R}^3$ is the vector part of the UAV quaternion, and h'_v denotes its transpose matrix.

3. Drone dynamics modeling

The UAV position dynamics model is shown in the following Equation (6).

$$\dot{v}^u = g\boldsymbol{e_3} - \frac{F}{m}\mathbf{R}e_3 \tag{6}$$

where m denotes the mass of the UAV, F denotes the magnitude of the total propeller tension, g is the gravitational acceleration. $\mathbf{R} \in \mathrm{R}^{3\times3}$ is the transformation matrix from the airframe coordinate system to the Earth coordinate system.

The attitude dynamics model of unmanned aerial vehicles in the aircraft system is shown in the following Equation (7).

$$I \cdot \dot{\omega}^u = -\omega^u \times (I \cdot \omega^u) + \mathrm{G}_a + \tau \tag{7}$$

where, $\tau \triangleq [\tau_{x_u} \quad \tau_{y_u} \quad \tau_{z_u}]' \in \mathrm{R}^3$ represents the rotor torque of the drone. $I \in \mathrm{R}^3$ is the rotational inertia of the drone itself. $\mathrm{G}_a \triangleq [\mathrm{G}_{a,\phi} \quad \mathrm{G}_{a,\theta} \quad \mathrm{G}_{a,\psi}]' \in \mathrm{R}^3$ represents the gyroscopic moment.

Combined, the above lead to the following rigid body model for UAV flight control, as shown in Equation (8).

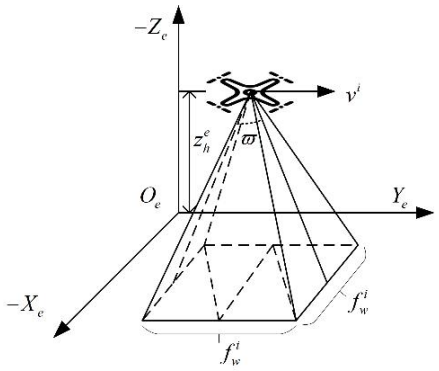
$$\begin{cases} \dot{\mathbf{P}}^e = \boldsymbol{v}^e \\ \dot{v}^u = g\boldsymbol{e_3} - \frac{F}{m}\mathbf{R}e_3 \\ \dot{h}_0 = -\frac{1}{2}\,\mathbf{h}'_v\omega^u \\ \dot{\mathbf{h}}_v = \frac{1}{2}(q_0 I_3 \mathbf{h}_v)\omega^u \\ I \cdot \dot{\omega}^u = -\omega^u \times (I \cdot \omega^u) + \mathrm{G}_a + \tau \end{cases} \tag{8}$$

2.3.2. Airborne Sensor Model

UAVs acquire information and localize targets through the sensors they carry, and the performance of the sensors is the basis for UAVs to carry out reconnaissance missions. When the image sensor carried by the drone captures a square reconnaissance area, its field of view width is shown in Equation (9).

$$f_w^i = 2z_h^e \tan(\varpi/2) \tag{9}$$

where ϖ is the reconnaissance angle and z_h^e is the flight altitude in the Earth coordinate system, as shown in the



UAV image sensor reconnaissance schematic shown in Figure 3.

Figure 3. Schematic of drone image sensor reconnaissance.

In addition, a set of distance sensors are used in this paper to help the UAV detect possible obstacle threats within a certain range around it. Specific UAV detection information is described in subsequent sections.

2.3.3. Flight Trajectory Calculation Model

Defaulting to point target reconnaissance, a single UAV accomplishes the reconnaissance mission with two main components:

Trajectory distance to the target: when calculating the length of the segment, the safety and feasibility of the UAV flight process should be fully considered, common threats such as missiles launched by the enemy, enemy radar scanning, so the UAV is not flying in a straight line, and need to go around the bends, then the segment of the trajectory distance $L^i_{O_eT_e^j}$ is shown in the following Equation (10).

$$L^i_{O_eT_e^j} = \int_{t^i_{O_e}}^{t^i_{T_e^j}} \left(v_0^i + a_t^i \cdot t\right) dt \tag{10}$$

where, $t^i_{T_e^j}, t^i_{O_e}$ denotes the moment when UAV U^i reconnoiters the target T^j and the moment when it takes off from the starting point O_e, respectively, and v_0^i, a_t^i denotes the initial speed when UAV U^i takes off and the acceleration at the moment t, respectively.

Returning trajectory distance: the trajectory length of the segment is recorded as $L^i_{T_e^j O_e'}$, the UAV U^i reconnaissance target T^j end position to return, the same method of calculating the length of the trajectory to arrive at the target, as shown in Equation (11).

$$L^i_{T_e^j O_e} = \int_{t^i_{T_e^j}}^{t^i_{O_e}} \left(\dot{v}_0^i + a_t^i \cdot t\right) dt \tag{11}$$

As a result, the total trajectory length $L_{T^j}^{U^i}$ for the UAV U^i to accomplish the reconnaissance mission to the target can be expressed as Equation (12).

$$L_{T^j}^{U^i} = L^i_{O_eT_e^j} + L^i_{T_e^j O_e} \tag{12}$$

2.3.4. Reconnaissance Tasking Model

In this paper, high-value targets and low-value targets in the reconnaissance environment are abstracted as point targets, and UAV clusters detect the target location by covering the reconnaissance mission area with scanning and searching. In the reconnaissance process, in order to maximize the combat effectiveness and complete the detection and strike integrated combat mission, it is necessary for the UAV to move toward the high -value target, but due to the limitation of UAV resources, usually a single UAV cannot complete the detection and strike mission of the high-value target, and in terms of the overall situation of the combat, the low-value target also needs to be reconnaissance and strike. Therefore, in the multi-UAV reconnaissance task allocation, it is

necessary to reasonably allocate single or multiple UAVs to simultaneously accomplish the detection and strike tasks for low-value and high-value targets under a certain number of UAVs. In order to highlight the performance of the proposed algorithm for real-time task assignment and obstacle avoidance for UAVs, the reconnaissance targets are set as static or low-slow speed targets in this paper. In addition, a target-oriented mechanism is used in the reconnaissance process to improve the efficiency of reconnaissance search.

Multi-UAV coordinated reconnaissance mission allocation needs to achieve two purposes, namely, to minimize the cost of a single UAV and maximize the benefit of a multi-UAV cluster, so as to achieve an overall allocation, complete coordinated reconnaissance of high-value and low-value targets, and improve the efficiency of mission execution.

The UAV reconnaissance gain consists of two parts. The first is the reconnaissance target gain, when the target T^j appears in the field of view of UAV U^i, it is regarded as reconnaissance of the target T^j, then the matrix element $E_{T^j}^{U^i}$ in the target reconnaissance gain matrix $\mathbf{E}_T^U$ is shown in the following Equation (13).

$$E_{T^j}^{U^i} = \begin{cases} 1 \; if \; \mathbf{P}_{T^j}^e \in \{f_w^i, f_w^i\} \\ 0, else \end{cases} \tag{13}$$

where $\mathbf{P}_{T^j}^e$ denotes the target T^j position and f_w^i is the field of view width of the UAV U^i.

Next is the reconnaissance area gain, UAV U^i in the process of reconnaissance target in collaboration with other teammates UAV as much as possible with less repetitive trajectory reconnaissance to more areas, reconnaissance area gain matrix for $\mathbf{E}_T^S$, then the total gain matrix is shown in the following Equation (14).

$$\widehat{\mathbf{E}} = \mathbf{E}_T^S + \mathbf{E}_T^U \tag{14}$$

A single UAV pays an energy cost during reconnaissance, the cost matrix is $\hat{\mathbf{L}}$, as shown in the following Equation (15).

$$\hat{\mathbf{L}} = \begin{bmatrix} L_{T^1}^{U^1} & \cdots & L_{T^M}^{U^1} \\ \vdots & \ddots & \vdots \\ L_{T^1}^{U^N} & \cdots & L_{T^M}^{U^N} \end{bmatrix} \tag{15}$$

Then the task assignment optimization model can be defined as the maximum indicator function $K(\mathrm{x,y})$, i.e., making the reconnaissance gain y optimal with x as the constraint, then the system task assignment optimization model O^*, as shown in the following Equation (16).

$$O^*(L, E) = \underset{E}{\operatorname{argmax}} \min_{L} K\left(\hat{L}, \widehat{E}\right) \tag{16}$$

3. VDN Based Collaborative Reconnaissance Algorithm for Multiple UAV

3.1. VDN Algorithm

VDN is a multi-agent reinforcement learning algorithm based on Deep Q-learning Network (DQN) [34], which automatically decomposes complex learning problems into local and easier to learn subproblems. It adopts the CTDE architecture and solves the false reward problem and lazy agent problem in multi-agent reinforcement learning through a value function-based method. These two problems are essentially credit assignment problems.

The core of this value function is to approximately decompose the team joint Q value function $\widetilde{Q}(h, a)$ into the sum of the sub functions Q_i of N agents, and the sub functions Q_i serve as the basis for each agent to execute actions, as shown in Equation (17).

$$\begin{aligned}\widetilde{Q}^{\kappa}(h,a)&=\widetilde{Q}^{\kappa}\left((h^1,\cdots h^i,\cdots h^N),\ (a^1,\cdots a^i,\cdots a^N)\right)=\mathbb{E}_{\kappa}\left[\sum_{t=0}^{\infty}\gamma^t R(s_t,a_t)\,|s_0=s,\ a_0=a\right]\\&=\mathbb{E}_{\kappa}\left[\sum_{t=0}^{\infty}\gamma^t R_1(o_t^1,a_t^1)\,|s_0=s,\ a_0=a\right]+,\cdots,+\mathbb{E}_{\kappa}\left[\sum_{t=0}^{\infty}\gamma^t R_N(o_t^N,a_t^N)\,|s_0=s,\ a_0=a\right]\\&=:Q_1^{\kappa}(s,a)+,\cdots,Q_i^{\kappa}(s,a)+,\cdots,+Q_N^{\kappa}(s,a)\approx Q_1^{\kappa}(h^1,a^1)+,\cdots,Q_i^{\kappa}(h^i,a^i)+,\cdots,+Q_N^{\kappa}(h^N,a^N)\\&\approx\sum_{i=1}^{n}Q_i^{\kappa}\,(h^i,a^i)\end{aligned}\tag{17}$$

Among them, $\boldsymbol{s},\boldsymbol{a}$ is the joint state and joint action of the system, $h^i$is the historical sequence information of the drone intelligent agent i, including observation information and other additional information, $a^i$is its action, and $o^i$is its observation value of the environment. The VDN network structure is shown in Figure 4.

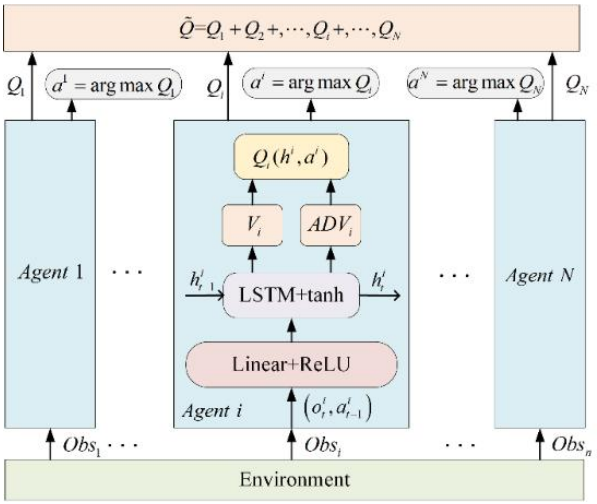


Figure 4. VDN network structure.

VDN uses a backpropagation mechanism to decompose the joint reward signal onto various agents, and then iterates the reward value to fit the joint Q value, learning the optimal linear value decomposition. During the training process of VDN, a parameter sharing mechanism is used to further solve the problem of lazy agents. An end-to-end training method is adopted. In the system, agent i obtains the observation value o^i through a local observation environment at time t combines it with the previous action a_{t-1}^i to obtain Q^i value, and selects the action a_t^i according to the greedy strategy to generate a decentralized strategy κ^i. Then, using the Qvalue update rule of DQN, $Q(s,a|\theta)$ is replaced with $\widetilde{Q}(\boldsymbol{h},\boldsymbol{a})$, the joint $\widetilde{Q}$ value TD error function $\tilde{L}(\theta)$ is calculated, and then the error is backpropagated to each sub Q function to learn the optimal strategy, and the network parameters are updated by minimizing the error function, as shown in the following Equations (18) and (19).

$$\tilde{L}(\theta)=\underset{h_t^i,a_t^i,r_t^i,s_{t+1}^i}{\mathbb{E}}\left[\widetilde{Q}(\boldsymbol{h}_t,\boldsymbol{a}_t)-\widetilde{Y}_t\right]^2\tag{18}$$

$$\widetilde{Y}_t=\tilde{r}_{t+1}+\gamma\max_{a_{t+1}}\widetilde{Q}\left(\boldsymbol{s}_{t+1},\boldsymbol{a}_{t+1}|\theta^{'}\right)\tag{19}$$

Among them, h_t,a_t,r_t,h_{t+1} represents the system t time joint historical sequence information, joint action, joint reward, and t+1time sequence information. $\widetilde{Y}_t$represents the joint target value.

At the same time, optimize the current value network based on the minimization strategy gradient. The gradient optimization of value networks can be expressed as shown in the following Equation (20).

$$\nabla_{\theta}\tilde{L}(\theta)=\underset{h_t^i,a_t^i,r_t^i,s_{t+1}^i}{\mathbb{E}}\left[\left(\widetilde{Q}(h_t,a_t)-\widetilde{Y}_t\right)\nabla_{\theta}\widetilde{Q}(h_t,a_t)\right]\tag{20}$$

Like the DQN algorithm, the VDN algorithm also adopts a soft update strategy to update each target network. For the target network of agent, its update method is represented, as shown in the following Equation (21).

$$\theta^i = \tau\theta^i + (1-\tau)\theta^{\dot{i}} \tag{21}$$

Among them, i represents the agent number, and is the soft update coefficient.

The process of multi drone collaborative reconnaissance algorithm based on VDN is shown in Figure 5. In the training process of a multi drone collaborative reconnaissance model based on VDN, each drone agent interacts with the environment, takes actions based on the observation values obtained, and obtains reward values and the next moment observation values. After all intelligent agents execute decisions, all experience samples $[h, \{a^i, i \in N\}, \{r^i, i \in N\}, h^{'}]$ are retained and stored in the experience replay pool for centralized training. When network updates are required, M_{batch} samples are randomly selected from the experience replay pool in batches, as shown in the sampling process in Figure 5. During the testing process, after deploying the trained network model to the drone, each drone agent can execute actions based on the current observation values, achieving distributed execution of the model.

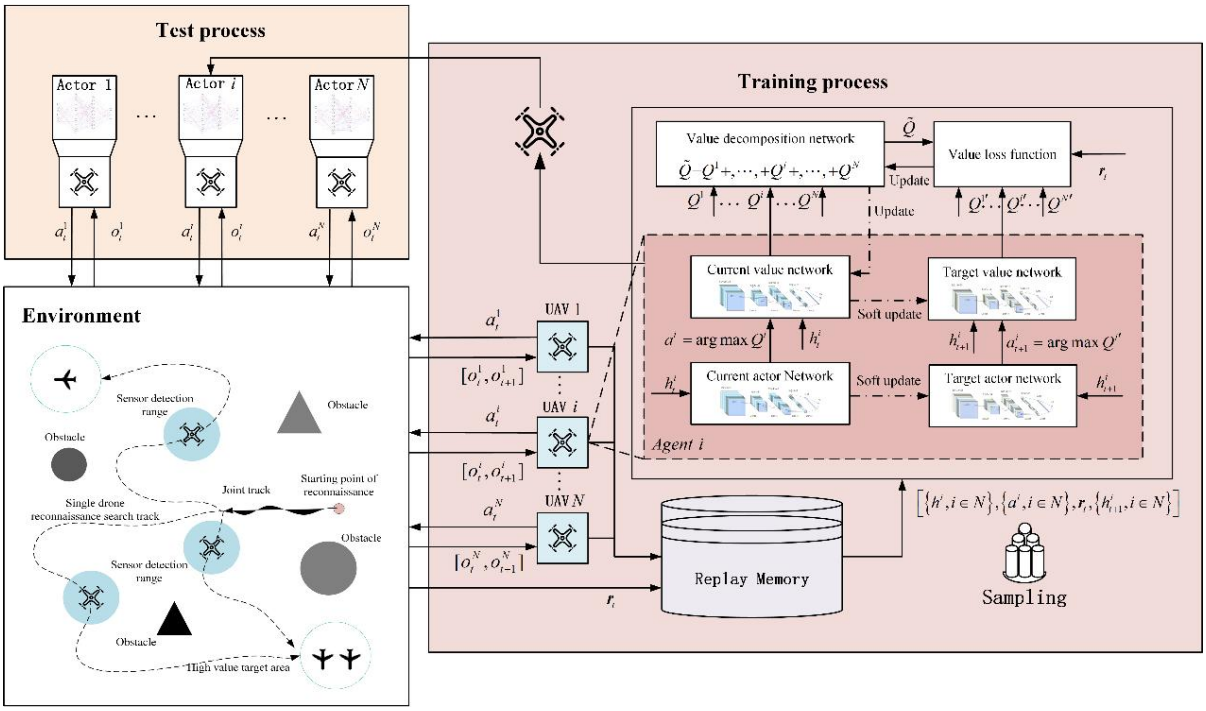


Figure 5. Schematic diagram of multi drone training based on VDN algorithm.

3.2. Design of Multiple Unmanned Aerial Vehicle Collaborative Reconnaissance Model Based on VDN Algorithm

3.2.1. Status Space

In the process of multi-agent algorithm training, the reconnaissance drone intelligent agent obtains environmental observation information and its own state through interaction with the environment. The drone state space designed in this paper can be represented shown in the following Equation (22).

$$\mathbf{S} = \left[\mathbf{S}_{uav}, \mathbf{S}_{teamer}, \mathbf{S}_{task}, \mathbf{S}_{obser}, \mathbf{S}_{history}, \mathbf{S}_{finish}\right] \tag{22}$$

Among them, $\mathbf{S}_{uav_i}$ and $\mathbf{S}_{teamer_i}$ obtain their own information and observation information of their teammates for the drone, including position, velocity, and angular velocity.

To meet the resource allocation of reconnaissance drones for high value and low value targets in reconnaissance missions, there is task information $\mathbf{S}_{task_i}$, as shown in the following Equation (23).

$$\mathbf{S}_{task}^i = \begin{cases} 0, L_{value} \\ 1, H_{value} \end{cases} \tag{23}$$

Among them, L_{value} represents that drone i is assigned to the low value target reconnaissance formation, while H_{value} represents that it is assigned to the high value target reconnaissance formation.

In order to observe the environment, an environmental observation status $\mathbf{S}_{obser_i}$ is set up to store distance information observed by drones, such as targets and obstacles.

In order to avoid UAV reconnaissance of repeated areas in the reconnaissance process, there is reconnaissance history information state $\mathbf{S}_{history}^{i}$ which is represented by matrix representation, where the matrix element 1 represents the mark of the position that has been scouted, as shown in the following Equation (24).

$$\mathbf{S}_{history}^{i} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \tag{24}$$

At the same time, in order to assist the drone in effectively completing reconnaissance of the target, the state variable $\mathbf{S}_{finish}^{i}$ is set to indicate whether the reconnaissance information of the target has been completed, as shown in the following Equation (25).

$$\mathbf{S}_{finish}^{i} = \begin{cases} 1, detection\ target \\ 0, other \end{cases} \tag{25}$$

3.2.2. Action Space

In the process of simulation experiment, because of the nonlinear characteristics of the kinematics and dynamics model of the rotor UAV, it is difficult to directly use the unmanned model training to realize the end-to-end control of reinforcement learning. Therefore, this article adopts a reinforcement learning model scheme for a single unmanned aerial vehicle, as shown in the schematic diagram of the hierarchical decision-making model structure in Figure 6.

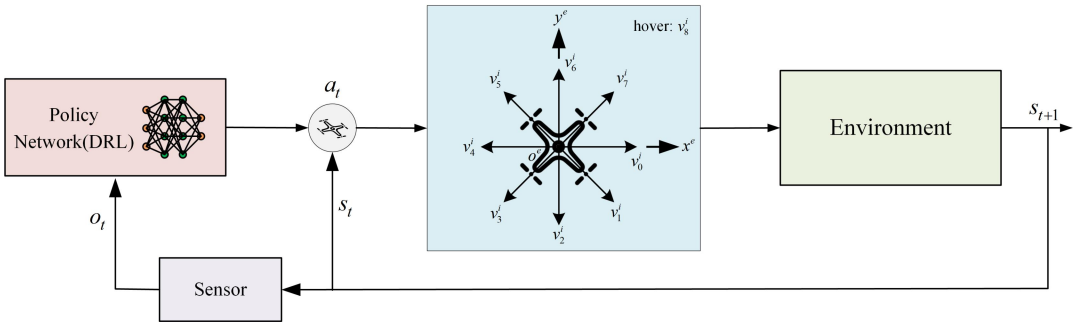


Figure 6. Schematic diagram of hierarchical maneuvering decision structure.

During the reconnaissance process of drones, when the drones take off, they only move on a certain plane. Therefore, this article designs the flight actions of drones as shown in Figure 7, and the action space on the horizontal plane of multiple drone systems is $9 \times N$.

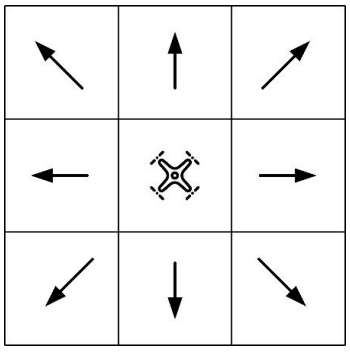


Figure 7. Unmanned aerial vehicle movement space on the horizontal plane.

3.2.3. Reward Function

In order to ensure the safe flight of each drone, search the target area and ultimately achieve collaborative reconnaissance of high value and low value targets. Considering the decision-making processes of task allocation, area search, maneuvering approach to reconnaissance targets, and autonomous obstacle avoidance in the collaborative reconnaissance process of multiple unmanned aerial vehicles, a reward function is designed for each reconnaissance drone individual in the multi drone system as shown in the following Equation (26).

$$r^i = \mathrm{e}_1 \cdot r^i_{safe} + \mathrm{e}_2 \cdot r^i_{priority} + \mathrm{e}_3 \cdot r^i_{speed} + \mathrm{e}_4 \cdot r^i_{guide} + \mathrm{e}_5 \cdot r^i_{finish} \tag{26}$$

During the reconnaissance process, our ground personnel usually provide estimates of the target area reconnaissance position before dispatching drones to conduct precise reconnaissance, which can improve reconnaissance efficiency. Therefore, design guidance rewards r^i_{guide}, as shown in the following Equations (27) and (28).

$$r^i_{guide} = \begin{cases} c_{guide}, c_{guide} > 0 \\ 0, else \end{cases} \tag{27}$$

$$c_{guide} = M - \frac{1}{10 \times M} \times \sum_{j=1}^{M} \min\left(\left|\mathbf{P}^e_i - \dot{\mathbf{P}}^e_j\right|\right) \tag{28}$$

Among them, $\mathbf{P}^e_i$ is the location of drone i, $\dot{\mathbf{P}}^e_j$ is the estimated center position of the ground report target j in the area, M is the number of targets in the task environment, and m is the number of high-value targets. $r^i_{priority}$ is a target priority reward used to distinguish high and low value goals.

In order to achieve area coverage as much as possible for drones, it is necessary to design a position continuous reward to drive drones to move towards undetected areas as much as possible, reducing repetitive flights to previously detected areas. Therefore, the forward speed reward and penalty r^i_{speed} for drones are set, as shown in the following Equation (29).

$$r^i_{speed} \begin{cases} \left|v_i^{e_{xy}}\right| \times 0.03, else \\ -0.1, if\ v_i^{e_x} < 0\ and\ v_i^{e_y} < 0 \end{cases} \tag{29}$$

Among them, $v_i^{e_{xy}}$ represents the velocity of drone i in the Earth coordinate system $o^e x^e y^e$ plane, and $v_i^{e_x}, v_i^{e_y}$ is its component. When $v_i^{e_x}, v_i^{e_y}$ is not both less than 0, the drone receives a positive reward, otherwise it receives a negative penalty. The drone that receives a positive reward has a speed in the first quadrant of the $o^e x^e y^e$ coordinate system as shown in Figure 8, other quadrants are similar to the first quadrant.

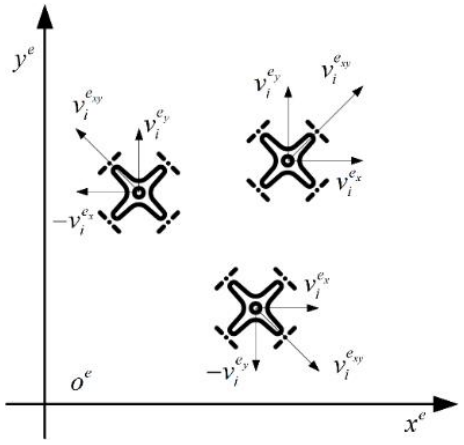


Figure 8. Schematic diagram of obtaining forward speed reward speed.

In order to enable unmanned aerial vehicles to navigate independently and bypass obstacles, design a reward r^i_{safe} for safe flight of unmanned aerial vehicles, as shown in the following Equations (30)–(32).

$$r_{safe}^{i} = r_{collision}^{i} + r_{out}^{i} \tag{30}$$

$$r_{collision}^{i} = \begin{cases} -5, & d_{io}^{e_{xy}} \le d_{obstacle} \ or \ d_{ii'}^{e} \le 0.1 \cdot d_{safe} \ or \ d_{ij}^{e} \le 1 \\ 0, other \end{cases} \tag{31}$$

$$r_{out}^{i} = \begin{cases} -10, & \mathbf{P}_{i}^{e} \notin Scenario \\ 0, other \end{cases} \tag{32}$$

Among them, r_{safe}^{i} consists of collision penalty $r_{collision}^{i}$ and out of bounds penalty r_{out}^{i}. $d_{io}^{e_{xy}}$ represents the distance between the unmanned aerial vehicle i and the obstacle on the $o^{e}x^{e}y^{e}$ plane in the Earth coordinate system; $d_{ii'}^{e}$ is the distance between the drone i, i', which can be obtained from the $\mathbf{P}_{i}^{e}, \mathbf{P}_{i'}^{e}$ coordinate in the Earth coordinate system; d_{ij}^{e} is the distance between the drone $i$and the target j. To prevent the reconnaissance target from counter reconnaissance, when the drone i is too close to the reconnaissance target, it will also receive a penalty, which is set to -1 in numerical terms.

When the drone detects the target, it will receive a completion reward r_{finish}^{i}. In addition, corresponding weights $\epsilon_{1\sim5}$ will be set for each sub reward to ensure that our reconnaissance drone receives effective reward rewards. Based on the constructed state input and action output models of reconnaissance drones, and using the set reward function to complete signal feedback, adaptive state perception and collaborative decision-making model training of multiple reconnaissance drones can be completed.

4. Experimental Design and Result Analysis

4.1. Network Structure and Parameter Design

The VDN network consists of a current value network, an action value network, and an experience pool. The action value network inputs the state space of the unmanned aerial vehicle intelligent agent, while the current value network inputs the output of the action value network and the state information of the unmanned aerial vehicle intelligent agent. The experience pool stores the relevant information of the intelligent agent during the training process.

In the algorithm section of this simulation experiment, the hyperparameter settings of IDQN and VDN algorithms are consistent. In order to ensure the gradient descent of the drone intelligent agent, the network learning rate is set to 0.01, and it decays once per round as the training progresses. When it decays to 0.0005, the learning rate no longer decays, and the training continues at this time. In addition, for this task, the maximum simulation step size for each round is set to 400. When 400 simulation steps are reached or a drone Done is used, the task for that round will automatically terminate and the environment will be reset for the next round of training. The other hyperparameters are specifically shown in Table 1.

Table 1. Network training hyperparameter setting.

Parameter Name	Parameter Value	Parameter Name	Parameter Value
Experience Pool Size $\boldsymbol{M}$	100,000	Training frequency $\boldsymbol{K}$	1000
Learning rate $\boldsymbol{lr}$	0.01~0.0005	Sample size $\boldsymbol{M_{batch}}$	3200
Pre training step size $\boldsymbol{P_{step}}$	3200	Maximum number of rounds $\boldsymbol{E}$	2000
Discount factor $\boldsymbol{\gamma}$	0.99	Greedy strategy decay rate $\boldsymbol{\varepsilon_{dec}}$	0.00001
Greedy strategy minimum $\boldsymbol{\varepsilon_{\min}}$	0.1	Maximum value of greedy strategy $\boldsymbol{\varepsilon_{\max}}$	1.0
Training step size $\boldsymbol{t_{step}}$	10,000	Soft update rate $\boldsymbol{\tau}$	0.01

In the experiment, the parameters of each drone are shown in Table 2.

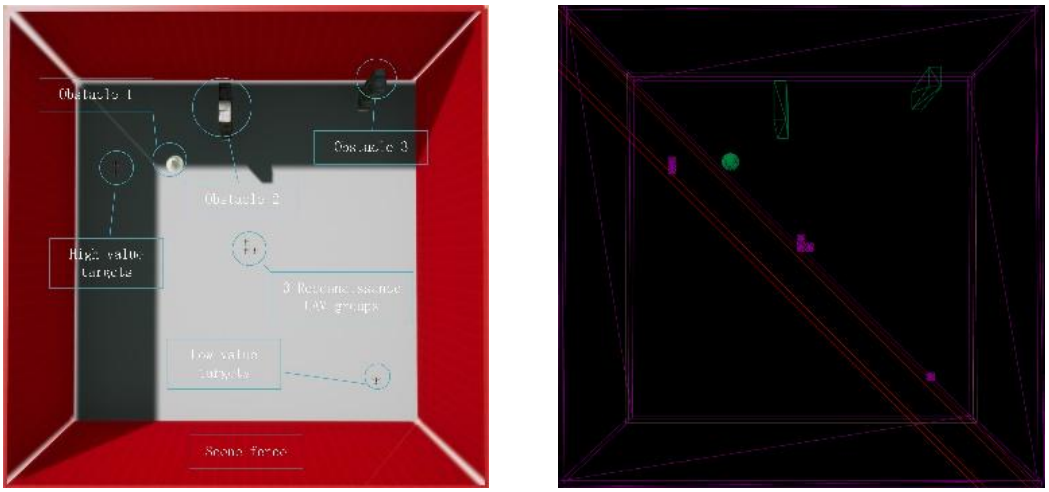
Table 2. UAV training parameters.

Parameter Name	Parameter Value	Parameter Name	Parameter Value
Initial speed of drone v	$(\mathbf{0}, \mathbf{0}, \mathbf{0})$ **m/s**	The cruising speed of unmanned aerial vehicles	**5 m/s**
UAV reconnaissance perspective $\boldsymbol{\varpi}$	**8°**	UAV reconnaissance cruise altitude	**0~5 m**
Image sensor detection range	$\mathbf{0 \sim 0.49\ m^2}$	Image sensor field of view width	**0~0.7 m**

4.2. Simulation Experiment and Result Analysis

4.1.1. Design of Simulation Experimental Environment

This article uses the AirSim project Air Learning as the simulation environment to validate the proposed algorithm. In this experiment, the number of drone intelligent agents is 3, and the environment contains 2 targets with different values and 3 obstacles, as shown in Figure 9a, shows the top view of the environment, and Figure 9b shows the ambient lighting view [35].



(**a**) Top view of the environment. (**b**) Environment lighting view.

Figure 9. Environmental top view and lighting map.

4.1.2. Result Analysis

After 2000 rounds of training, the accumulated reward curve collected by the VDN algorithm is shown in Figure 10, where the solid line represents the smoothed value and the shadow represents the actual value. In this environment, the maximum sum of rewards obtained by the VDN algorithm during the training process is 402.8. After smoothing the reward curve throughout the entire training process, the VDN algorithm's reward curve is relatively stable and ultimately converges to 362.3.

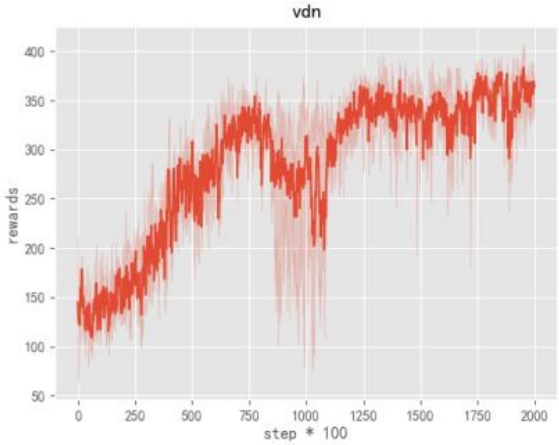


Figure 10. The sum curve of rewards obtained by multiple drone collaborations in the environment.

Select a model that has been successfully trained based on the VDN algorithm and load it into the testing environment. The reconnaissance flight trajectory searched by the sensors carried by the VDN algorithm in the testing environment is shown in Figure11. In the figure, the purple circle represents the high value target area, the yellow triangle represents the low target value area, and the black square represents the obstacle position.

As shown in Figure 11, the flight trajectories of three reconnaissance drones were trained using the VDN algorithm. UAV1, UAV2, and UAV3 took off simultaneously from the reconnaissance starting point. In the first half of the flight, the three drones flew together. In the second half of the flight, reinforcement learning strategies were used to assign reconnaissance targets to the drone swarm. Among them, UAV1 is assigned to scout low value target areas, while UAV2 and UAV3 are synergistically assigned to scout high-value target areas, achieving effective matching of reconnaissance forces and reconnaissance targets. Additionally, all three drones approach the estimated target area and identify the target through obstacle avoidance, completing reconnaissance tasks and improving collaborative reconnaissance efficiency. The three-dimensional trajectory of the drone demonstrates the effectiveness of the VDN algorithm strategy.

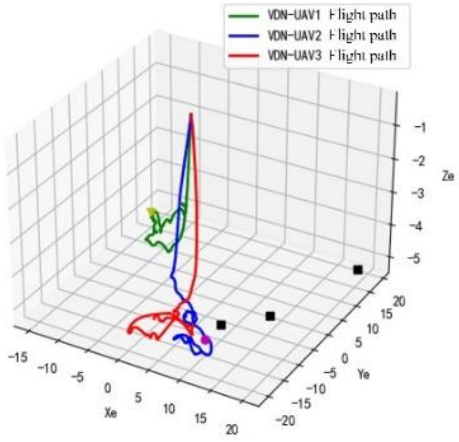
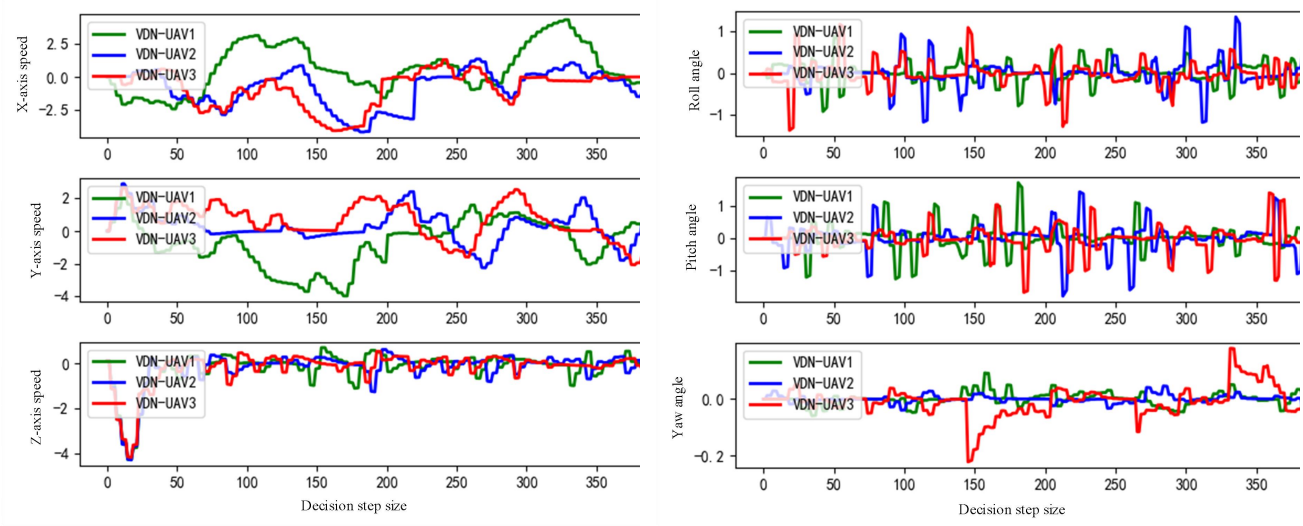


Figure 11. VDN algorithm strategy decision for UAV 3D trajectory in environment.

The velocity and angular velocity curves of the drone in this environment are shown in Figure 12. In Figure 12a, the VDN algorithm's strategy decision for the three drones resulted in sustained positive or negative three-axis velocities for a period of time, which allowed the drones to maintain forward reconnaissance and reduce turnaround flights; In Figure 12b, the roll angle, pitch angle, and yaw angle of the three drones in the VDN algorithm strategy decision vary more frequently within their respective positive and negative axes, indicating that the flight attitude of the drones is not stable enough.



(**a**) UAV three-axis speed. (**b**) UAV three-axis angular velocity.

Figure 12. UAV speed and angular velocity curves.

From the above experiments, it can be seen that the VDN algorithm has convergence performance during the training process, and has the characteristics of large search coverage area, fast and safe flight when implementing reconnaissance flight tasks.

5. Discussion

The VDN algorithm has a simple structure, and the Q value obtained through its decomposition allows agents to choose greedy actions based on their local observations, thereby executing distributed strategies. Its centralized training method can to some extent ensure the optimality of the overall Q function. In addition, the end-to-end training and parameter sharing of VDN make the algorithm converge very quickly. For some simple tasks, the algorithm can be said to be both fast and effective. The multi UAV collaborative reconnaissance decision-making method proposed based on this algorithm provides an intelligent solution for future multi UAV collaborative decision-making applications.

6. Conclusions

This article proposes a multi UAV collaborative reconnaissance method based on multi-agent value decomposition network (VDN) algorithm to address the shortcomings of multi UAV collaborative reconnaissance strategies and dynamic planning. Through experiments, it has been verified that this method can effectively handle complex environments in UAV reconnaissance, achieve task allocation and trajectory planning in collaborative decision-making, and complete multi UAV collaborative reconnaissance tasks. The contributions of this paper are as follows:

a. In response to the problem of relatively simple drone modeling and task environment in deep reinforcement learning based unmanned aerial vehicle (UAV) collaborative reconnaissance process, which differs significantly from real UAV reconnaissance flight modeling, this study establishes a detailed multi UAV collaborative reconnaissance task model, including UAV flight control model, airborne sensor model, flight trajectory calculation model, and reconnaissance task allocation model. And the Airsim platform, which has a realistic fidelity to the real environment, is used as the simulation environment to meet the needs of unmanned aerial vehicle autonomous collaborative reconnaissance decision-making in complex environments.

b. This study adopts the multi-agent reinforcement learning VDN algorithm as the problem-solving method, which automatically decomposes complex learning problems into local and easier to learn sub problems,

solves the problems of false rewards and lazy agents in multi-agent reinforcement learning, and promotes unmanned aerial vehicle intelligent agents to scout unknown environments.

c. This study closely couples drone target allocation tasks with trajectory planning, which is more in line with the needs of collaborative reconnaissance decision-making among multiple drones in complex environments.

Author Contributions

J.H. is responsible for designing research directions and developing experimental plans. Z.Y. participated in the collection and analysis of experimental data. J.L. is responsible for literature review and background research. S.W. provides technical support and experimental equipment. X.Z. participated in the outcome discussion and conclusion summary. B.L. is responsible for editing and proofreading the manuscript.

Funding

This work received no external funding.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Data Availability Statement

Not applicable.

Acknowledgments

The authors would like to acknowledge Project Supported by the National Nature Science Foundation of China (Grant No.62003267), Supported by the Key Research and Development Program of Shaanxi Province (Grant No.2023-GHZD-33), Supported by the Fundamental Research Funds for the Central Universities (Grant No.G2022KY0602) Project Supported by Technology on Electromagnetic Space Operations and Applications Laboratory (Grant No.2022ZX0090) and Open Project of the State Key Laboratory of Intelligent Game(Grant No. ZBKF-23-05) to provide fund for conducting experiments.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Li, B.; Yang, Z.Y.; Jia, Z.R.; Ma, H. A. Unsupervised Learning Neural Network for UAV Regional Reconnaissance Path Planning. *J. Northwestern Polytechnical Univ.* **2021**, *39*, 77–84. [CrossRef]
2. Zhou, X.Q. The Development Status and Trends of Foreign Electronic Warfare Drones. *Ship Electron. Countermeasures* **2003**, *26*, 6–9+19. [CrossRef]
3. Chen, Z.J.; Wei, J.Z.; Wang, Y.X.; Zhou, R. UAV Autonomous Control Levels and System Structure. *Acta Aeronautica et Astronautica Sinica* **2011**, *32*, 1075–1083. [CrossRef]
4. Cao, J.H.; Gao, X.G. Intelligent Command and Control System for Multi UAV Collaborative Operations. *Firepower Command Control.* **2003**, *28*, 4. [CrossRef]
5. Chen, H.; Wang, X.M.; Jiao, Y.S.; Li, Y.A. UAV Coverage Trajectory Planning Algorithm for Convex Polygonal Regions. *J. Aeronaut.* **2010**, *31,* 1802–1808.
6. Fu, X.W.; Wei, G.W.; Gao, X.G. Multi UAV Collaborative Area Search Algorithm in Uncertain Environments. *Syst. Eng. Electron. Technol.* **2016**, *38,* 821–827.
7. Wang, D.; Zhang, G.Z.; Mu, W.D. Multi UAV collaborative combat communication self-organizing network technology. *Aviat. Missile* **2012,** *1*, 59–63.

8. Merino, L.; Caballero, F.; Ferruz, J.; Wiklund, J.; Forssén, P.; Ollero, A. Multi UAV Cooperative Perception Techniques. In *Multiple Heterogeneous Unmanned Aerial Vehicles*; Ollero, A., Maza, I., Eds.; Springer Tracts in Advanced Robotics: Berlin, Germany; Volume 37, pp. 67–110. [CrossRef]
9. Hu, P.L.; Zhao, C.H.; Hu, J.W. Reject Collaborative Perception and Autonomous Control of Unmanned Aerial Vehicle Clusters in the Environment. In Proceedings of the 40th China Control Conference, Shanghai, China, 26–28 July 2021.
10. Li, C.C. Collaborative Perception and Visualization of Three-Dimensional Complex Environments. Master Dissertation, Xi'an University of Electronic Science and Technology, Xi'an, China, 1 May 2020. [CrossRef]
11. Merino, L.; Caballero, F.; Dios, J.R.M. A Cooperative Perception System for Multiple UAVs: Application to Automatic Detection of Forest Fires. *J. Field Rob.* **2006**, *23*, 165–184. [CrossRef]
12. Zhong, S.B.; Zhu, W.; Peng, L.; Huang, X.B. Research on the Key Technology System of Collaborative Perception. *CN Emerg. Manag.* **2021**, *12*, 52–55. [CrossRef]
13. Chen, Y. Research on Planning and Simulation of Collaborative Reconnaissance Tasks for Drone Clusters. Master Dissertation, Nanjing University of Aeronautics and Astronautics, Nanjing, China, 1 March 2021. [CrossRef]
14. Rasmussen, S.J.; Shima, T. Branch and Bound Tree Search for Assigning Cooperative UAVs to Multiple Tasks. In Proceedings of the 2006 American Control Conference, Minneapolis, MN, USA, 14–16 June 2006. [CrossRef]
15. Azam, Md A.; Shankarachary, R. Decentralized Formation Shape Control of UAV Swarm Using Dynamic Programming. *Signal Process., Sens./Inf. Fusion Target Recogn. XXIX* **2020**, *11423*, 69–76. [CrossRef]
16. Pang, Q.W.; Li, W.G.; Li, Y.K.; Hu, Y.J.; Jia, H.X. Multi UAV Collaborative Reconnaissance Trajectory Planning Based on Improved Genetic Algorithm. *CN J. Inertial Technol.* **2020**, *28*, 248–255. [CrossRef]
17. Kang, X.C.; He, G.J.; Chen, F.; Li. X.G. A Discrete Firefly Algorithm for Solving the Task Allocation Problem of Unmanned Aerial Vehicle ISR. *J. Missile Guidance* **2019**, *39*, 131–134+138. [CrossRef]
18. Lin, J.C.; Jia, G.W.; Hou, Z.X. Research on Task Assignment of Heterogeneous UAV Formation in the Anti-radar Combat. *Sys. Eng. Electron.* **2018**, *40*, 1986–1992. [CrossRef]
19. Tian, Z.; Wang, X.F. Cooperative Multiple Task Assignment for Heterogeneous Multi-UAVs with Multi-Chromosome Genetic Algorithm. *Flight Dyn.* **2020**, *9*, 687. [CrossRef]
20. Maza, I.; Ollero, A. Multiple UAV Cooperative Searching Operation Using Polygon Area Decomposition and Efficient Coverage Algorithms. In *Distributed Autonomous Robotic Systems 6*, 1st ed.; Alami, R., Chatila, R., Asama, H., Eds.; Springer Tokyo: Tokyo, Japan, 2007; Volume 1, pp. 221–230. [CrossRef]
21. Agarwal, A.; Hiot, L.M.; Nghia, N.T. Parallel Region Coverage Using Multiple UAVs. In Proceedings of the 2006 IEEE Aerospace Conference, Big Sky, MT, USA, 4–11 March 2006. [CrossRef]
22. Yao, Y.; Li, Q.; Chen, X. Optimization of the Application of A* Algorithm in Path Planning. *Microelectron. Comput.* **2017**, *34*, 51–55.
23. Chen, J.Y.; Hu, K.K.; Li, Y.W. Research on UAV Multi-point Navigation Algorithm Based on MBRRT*. *Comput. Sci.* **2018**, *45*, 85–90.
24. Pehlivanoglu, Y.V. A New Vibrational Genetic Algorithm Enhanced with a Voronoi Diagram for Path Planning of Autonomous UAV. *Aerosp. Sci. Technol.* **2012**, *16*, 47–55. [CrossRef]
25. Wang, C.; Dong, H.L.; Gu, X.S. Improved Particle Swarm Optimization Algorithm and its Application Path Planning. *Control Eng. CN* **2019**, *26*, 1466–1471. [CrossRef]
26. Zhou, Q.; Zhang, R.; Suo, X.J. Genetic Algorithm for UAV Trajectory Planning with Timing Constraints. *Aeronautical Comput. Tech.* **2016**, *46*, 93–96.
27. Li X.G.; Cai, Y.L. Unmanned Aerial Vehicle Path Planning Based on Improved Ant Colony Algorithm. *Flight Mech.* 2017, *35*, 52–56. [CrossRef]
28. Li, Y.Q. Route Planning for Multi UAV Collaborative Area Surveillance Based on Genetic Algorithm and Deep Reinforcement Learning. Master Dissertation, Xi'an University of Electronic Science and Technology, Xi'an, China, 1 June 2018.
29. Xu, T.H. Research on Deep Reinforcement Learning Method for Autonomous Collaborative Reconnaissance of Drone Clusters. Master Dissertation, National University of Defense Technology, Changsha, China, 1 October 2019.
30. Zhang, F.Z.; Zhu, Y. Task Allocation Method for Collaborative Reconnaissance of Multiple Unmanned Aerial Vehicles in Complex Environments. *J. Sys. Simul.* **2022**, *34*, 2293–2302.
31. Li, B.; Huang, J.Y.; Wan, K.F.; Song, C. A Review of Research on the Application of UAV System Based on Deep Reinforcement Learning. *Tacti-cal Missile Technol.* 2023, *1*, 58–68.
32. Zhao, Y.; Guo, J.F.; Zheng, H.X.; Bai, C.C. A Reinforcement Learning Based Collision Avoidance Computational Guidance Method for Multiple Unmanned Aerial Vehicles. *Navigation Positioning Timing* **2021**, *8*, 31–40. [CrossRef]
33. Fan, L.T. Research on Multi UAV Collaborative Task Planning Algorithm Based on Reinforcement Learning. Master Dissertation, Henan University of Science and Technology, Luoyang, China, 1 May 2019.

34. Value-decomposition Networks for Cooperative Multi-Agent Learning. Available online: https://arxiv.org/abs/1706.05296 (accessed on 1 March 2024).
35. Zhou, Y. Research on 3D Obstacle Avoidance Algorithm for Unmanned Aerial Vehicles Based on Airsim Simulation Platform. Master Dissertation, University of Electronic Science and Technology, Sichuan, China, 15 March 2020.