applied sciences

MDPI

Article

Reinforcement Learning-Based Anti-Jamming in Networked UAV Radar Systems

Qinhao Wu [1], Hongqiang Wang [1], Xiang Li [1], Bo Zhang [2] and Jinlin Peng [2,*]

[1] College of Electronic Science, National University of Defense Technology, Changsha 410073, China; qinhaowu@hotmail.com (Q.W.); oliverwhq@tom.com (H.W.); lixiang01@vip.sina.com (X.L.)

[2] Artificial Intelligence Research Center, National Innovation Institute of Defense Technology, Beijing 100010, China; bo.zhang.airc@outlook.com

* Correspondence: peng_jinlin@126.com

Received: 20 October 2019; Accepted: 25 November 2019; Published: 28 November 2019

Abstract: The networked unmanned aerial vehicle (UAV) radar system may exploit inter-UAV cooperation for enhancing information acquisition capabilities, meanwhile its inter-UAV communications may be interfered with by external jammers. This paper is devoted to quantifying and optimizing the anti-jamming performance of networked UAV radar systems in adversarial electromagnetic environments. Firstly, instead of using the conventional metric of signal-to-interference ratio (SIR), this paper explores use of the theory of radar information representation as the basis of evaluating the information acquisition capabilities of the networked UAV radar systems. Secondly, this paper proposes a modified Q-Learning method based on double greedy algorithm to optimize the anti-jamming performance of the networked UAV radar systems, through joint programming in the frequency-motion-antenna domain. Simulation results prove the effectiveness of the algorithm in two different networking scenarios.

Keywords: radar anti-jamming; UAV radar networking; low-bit beam synthesis; radar information representation; double-greedy algorithm

1. Introduction

A radar countermeasure is an important part of an electronic countermeasure [1]. An effective radar countermeasures can deal with the jamming of various electronic equipment and signal modes [2]. At present, the mainstream countermeasure methods are as follows: 1. Game theory: establish a network model of multiple agents and use game theory to model the adjustment of policies between jammer and radar as Stackelberg game. Through analyzing and solving the game equilibrium solution (such as the Nash equilibrium), we can obtain the optimal transmission power and other policies for users to realize anti-jamming [3–5]. 2. Physical layer security: this is a supplement to traditional cryptography-based methods. Users usually generate noise signals to confuse potential eavesdroppers, which significantly improves the quality and reliability of secure communication between legitimate terminals [6,7]. 3. Machine learning: intelligent countermeasures represented by reinforcement learning have been widely used in recent years. Reinforcement learning realizes the policy iteration through the interaction between agent and environment and finally generates a reasonable policy set [8,9]. With the progress of electronic jamming, the traditional jammer with single function and fixed working mode is promoted to intelligent jamming of multi-function, multi-parameter and variable working mode [10]. Therefore, the intelligent radar system becomes the focus of anti-jamming research [11]. Because the traditional radar system often uses passive methods such as a waveform library to identify the jamming form, this results in poor real-time policy [12]. When it deals with new

forms of jamming, the utility of this approach is limited. Therefore, an intelligent radar system needs to generate real-time policies under the condition of flexible jamming forms.

Radar countermeasure of an intelligent unmanned aerial vehicle (UAV) swarm is an effective form of intelligent radar countermeasure. Unlike the mission planning of general multi-radar systems, swarm intelligence UAV technology pays more attention to the individual UAV with a simple structure and limited energy. Through internal interaction and swarm control, the detection capability of a single UAV in a swarm is greatly promoted, and it is more flexible than traditional radar. A UAV has the characteristics of good concealment, flexible motion and rapid networking. An intelligent UAV cannot only perform common tasks such as surveillance of unknown areas, protection of key targets, territorial investigation and so on, and more flexible operations such as radar networking can be used to achieve radar countermeasures by airborne small radar [13]. Radar networking based on a UAV swarm greatly improves the ability of radar information acquisition by properly distributing radar stations with different systems, different frequency bands, different working modes, and linking them into networks.

However, the following design considerations should be taken into account in intelligent networked UAV radar systems:

1. Intelligent UAVs may be energy-limited. The energy modules on most UAVs may pose a major constraint on their motion ability. Meanwhile, when the UAV carries a radar payload to perform detection tasks, its motion and adaptation of the radar parameters may introduce considerable energy consumption. Therefore, a UAV may need to strike a beneficial trade-off between the motion energy and energy consumed by the radar.
2. The antenna performance of smart UAVs may be limited. Due to the limited volume and load capacity of the UAV, the number of array elements may be limited, resulting in poor spatial resolution. However, the beam width adopted by radar in certain tasks may demand high spatial resolutions, e.g., when measuring angle by beam scanning method. Therefore, the networked UAV radar system should exploit the distributed UAV cooperation to overcome the antenna constraint of a single UAV.
3. The topology of the networked UAV radar system may be complex. In order to comprehensively analyse the received information, the target data need to be transmitted to a data fusion center. Therefore, secondary forwarding devices such as relays may be applied during long-range communications. The networked UAV radar system should control complexity of information transmission process, and control the risks of information interception and jamming.

In order to address the aforementioned design concerns, this paper proposes the use of reinforcement learning for training the behavior of the networked UAV radar systems. Although reinforcement learning is being adopted as an effective method in radar anti-jamming, there is a lack of research on multi-parameter programming in the literature. Meanwhile, most researchers only regard signal-to-interference ratio (SIR) as the criterion of the reward, which may not fully reflect the ability of the radar system to acquire target information. Therefore, this paper applies information representation theory to evaluate the capability of radar information acquisition.

Specifically, this paper investigates the radar countermeasures under the conditions of malicious jamming, constrained UAV motion, fewer airborne radar elements, and complex topology of UAV networking. We construct two radar networking methods, introduce the theory of radar information representation, and use modified Q-Learning based on a double greedy algorithm to generate anti-jamming policies. The simulation results will prove the effectiveness of the algorithm.

The contributions of this paper are as follows:

1. A radar countermeasure system based on intelligent networked UAV swarms is established. Compared with the existing radar countermeasure models, our model introduces joint programming of UAV-borne radar to make full use of various degrees of freedom, where the policies can be generated from three dimensions of antenna, frequency and position.

2. Two UAV radar networking modes are proposed and implemented. The simulation results show that the proposed algorithm can realize the effective programming of radar system parameters in both modes.
3. This paper introduces a novel reward design by the information representation method to quantify the information received by the networked UAV radar systems.

The structure of this paper is organized as follows: Section 2 introduces the related works of UAV networking and anti-jamming based on reinforcement learning. Section 3 is the system modeling. We will introduce the intelligent UAV swarm networking model and the phased array beam synthesis method with limited phase coding bits. Section 4 introduces DGQL (Double-Greedy Reinforcement Learning) algorithm based on information representation. Section 5 gives the simulation results and analysis, which proves the effectiveness of the DGQL. Section 6 summarizes the contribution of this paper.

2. Related Works

UAVs have been used widely in the anti-jamming field. A UAV is designed in [14] to avoid the jamming of the aerial jammer to the channel by differential game method. This method could generate the optimal policy to avoid jamming. A UAV was used in [15] to send onboard units information, which improved the communication performance of vehicular ad-hoc networks in the presence of intelligent jammer. As for swarm UAV, Li Haitao et al. proposed a countermeasure model based on airborne cognitive radio and used an improved energy detection method to detect jamming [16]. Li Zhiwei et al. presented a method of using a software-defined network (SDN) to counter software-defined radio (SDR) [17]. A new type of network using a dual-controller generation policy was proposed. Rahmes Mark et al. designed a multi-UAV countermeasure system and used Q-Learning algorithm to generate radar transmitting frequency [18]. Cevik Polat et al. reduced the resource allocation problem in electronic countermeasures to an optimal allocation problem, and proposed a resource allocation algorithm based on particle swarm optimization theory [19].

At present, there is some research on radar reinforcement learning, mainly in the field of deep reinforcement learning. An intelligent radar countermeasure method was studied in [20] under the condition where the number of working modes of multi-functional radar was unknown. The jamming receiver recognized the working state of the radar by processing the information, and then generated a new effective jamming pattern. Liu Pengfei et al. combined reinforcement learning with cognitive radar in [21], so that transmitting and receiving could adapt to an unknown dynamic environment. It explained how to realize spectrum allocation of automotive radar in reinforcement learning-based cognitive radar system. You Shixun et al. simulated an electronic attacker as a UAV and a defender as an observatory in [22], where the UAV needed to detect targets. A deep deterministic policy gradient (DDPG) algorithm based on variational Bayesian estimation was designed to train the motion policy of the UAV. Wang Li et al. proved that some basic reinforcement learning algorithms could be successfully applied to dynamic detection of multiple-input-multiple-output (MIMO) radar with unknown environment and unknown number of targets [23]. Deep learning was used in [24] to discuss channel selection when wideband radar signals coexist with narrowband communication signals in the communication systems.

Some progress has also been made in the research of the combination of jamming and physical layer security. For example, Furqan Jameel et al. reviewed the latest work of cooperative relaying and jamming in detail [6]. These technologies are used to protect wireless transmission from eavesdropping nodes. They explained that using cooperative security can improve the physical layer security, and proposed some methods and conclusions to improve the wireless communication performance based on physical layer security. For example, relay positioning can be further explored, multi-cellular design can be further studied, etc. This paper is of great significance to the design a security scheme in the 5G communication network. At the same time, Huo Yan et al. summarized the jamming policies in the

physical layer of wireless communication and classified the jamming policies in [25], which provided a theoretical basis for the design of jamming policy in a wireless communication system.

From the literature, it may be concluded that frequency-domain anti-jamming techniques have been widely adopted in the field of anti-jamming of UAV radar networking, but the degrees of freedom exhibited in motion and antenna arrays of networked-UAV radar systems are not fully exploited. Secondly, the radar anti-jamming techniques proposed in the literature generally adopted SIR or SNR as the reward in reinforcement learning, while it may not fully capture the information acquisition ability of the radar. In this paper, the beam synthesis, networking and autonomous motion in networked-UAV radar systems are jointly optimized, with the aid of radar information representation theory and reinforcement learning.

3. System Model

3.1. Modeling of Radar Unmanned Aerial Vehicle (UAV) Swarm Networking

Spectrum resource allocation is a common problem in radar countermeasures. In the actual scenario, the spectrum resources are very limited because there are numerous electronic devices working at various modes. The setting of parameters in a radar system, such as position and antenna, needs to be tuned to avoid jamming effectively.

Traditional ground-based radar has the advantages of high power, long detection distance and narrow beam width. However, its mobility poses a threat for real-time detection in case of jamming. In comparison, small array radars carried by networked UAVs may realize swift position shift, phased array element combination, relatively flexible beam scanning and conversion between multiple detecting modes. Not only can UAVs in one single swarm cooperate with each other to synthesize beams, but also cooperate among multiple swarms. Thus, many parameters of targets can be observed simultaneously, which greatly improves the detection efficiency of the radar system.

However, due to the energy constraints, the power and moving range of UAVs should be considered for effective applications. When multiple swarms work at the same time, each swarm needs to send the data back to the control center. The multi-channel data can be analyzed synthetically by the control center to obtain multi-dimensional information. In case the UAV performs tasks in remote areas, communication UAV may serve as a mobile relay to assist the data transmission. As an effective radar countermeasure, jamming systems may be used to decrease the efficiency of the radar system, where both the UAV swarm and the relay may be jammed.

Against the above background, a radar countermeasure system composed of multi-agents is established in this paper. The system framework is shown in Figure 1a. The system includes 5 kinds of agents: jammer, radar UAV swarms, communication UAV, base station and target. Among them, radar UAV swarms, communication UAV and base station form the radar systems together. The objective of is to maximize the target information acquisition in different networking modes by adjusting the parameters of the radar system. In Figure 1b, the internal schematic diagram of a radar UAV swarm is given. A radar UAV swarm consists of a number of UAVs, and each UAV carries an antenna array. Radar UAV swarms are responsible for transmitting electromagnetic waves to detect targets and receiving signals. This process is called the transmission link of the radar UAV swarm. It is assumed that the distance between UAVs is much smaller than the distance from the whole UAV swarm to the target.

Specifically, the UAVs in a swarm can form a uniform linear array (ULA) or uniform plane array (UPA) by changing their arrangement to synthesize beams. However, due to the limited array volume of the UAV, its phase shifter cannot have a high coding bit like a large ground-based phased array radar. Therefore, this paper explores the beam synthesis method under low-bit phase coding, which will be introduced in Section 3.2.

The UAVs in a swarm transmit linear frequency modulation (LFM) signals with the same waveform parameters. Therefore, the UAVs in a swarm are strictly synchronized, where each UAV swarm can

play a similar role to a complete array radar. In order to simplify the signal modeling process, we set up two UAV swarms in frequency division multiplexing mode. When transmitting signals, the frequency bands occupied by two radar UAV swarms only need to be expressed by carrier frequency f_c and bandwidth b. Radar UAV swarm 1 occupies the frequency range of $[f_c - b/2, f_c)$, and radar UAV swarm 2 occupies the frequency range of $(f_c, f_c + b/2]$. When the mission is determined, the position of the UAV swarm remains unchanged, and the transmitting frequency, forwarding frequency and antenna pattern are tunable.

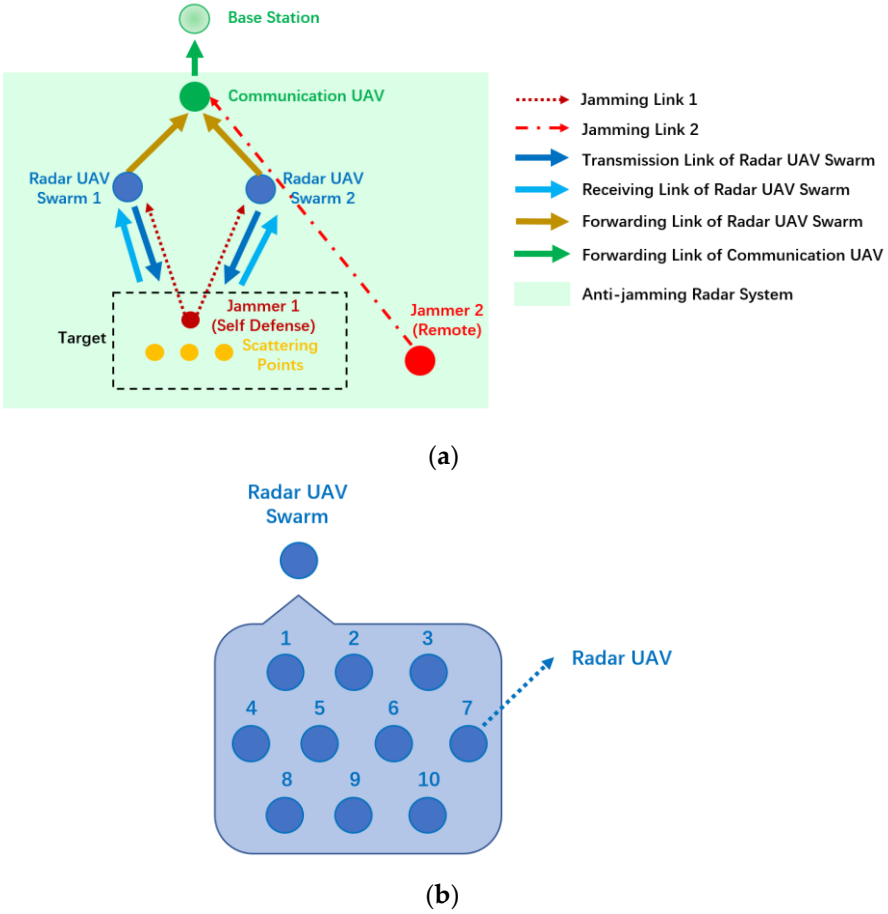


Figure 1. (**a**) Framework of radar swarm networking system; (**b**) internal structure of radar unmanned aerial vehicle (UAV) swarm.

Since the two radar UAV swarms need to transmit data back to the base station for comprehensive analysis, after receiving the target echo the two radar UAV swarms need to convert part of the received signal within the passband into the communication UAV using the frequency difference of Δf. This process is called forwarding link of radar UAV swarm. The reason for frequency conversion is to reduce the interference between transmission link and forwarding link. After frequency conversion, the signal still has two main parameters: f_c^* and b^*. f_c^* is the carrier frequency after frequency conversion, and b^* remains unchanged. Frequency division multiplexing mode is still used for both signals.

The reason for setting up communication UAV is that due to the energy constraints of UAV, the two radar UAV swarms need to allocate most of the power on the transmission link and minimize the energy cost of the forwarding link. The forwarding power cannot help send data directly back to a distant base station. Therefore, the communication UAV is responsible for collecting data from two signals and forwarding them to the base station. In this paper, the transmission link from communication UAV to base station is not considered. In order to avoid jamming better, the position of communication UAV is tunable. System parameters are shown in Table 1.

Table 1. System parameters.

Parameters	Variables
Transmitting Power of Radar Unmanned Aerial Vehicle (UAV) Swarm 1	P_R
Target Scattering Coefficient	$\sigma_1, \sigma_2, \sigma_3, \sigma_4, \sigma_5$
Antenna Gain of Radar UAV Swarm 1 for Each Scattering Point	$G_{R1}, G_{R2}, G_{R3}, G_{R4}, G_{R5}$
Transmission Link Path Loss of Radar UAV Swarm 1	L_{R1}
Forwarding Gain of Radar UAV Swarm 1	F_R
Forwarding Link Path Loss of Radar UAV Swarm 1	L_{R2}
Transmitting Power of Radar UAV Swarm 2	P_r
Antenna Gain of Radar UAV Swarm 2 for Each Target Scattering Point	$G_{r1}, G_{r2}, G_{r3}, G_{r4}, G_{r5}$
Transmission Link Path Loss of Radar UAV Swarm 2	L_{r1}
Forwarding Gain of Radar UAV Swarm 2	F_r
Forwarding Link Path Loss of Radar UAV Swarm 2	L_{r2}
Transmitting Power of Jammer 1	P_{j1}
Antenna Gain of Jammer 1	G_{j1}
Path Loss of Jamming Link 1	L_{j1}
Transmitting Power of Jammer 2	P_{j2}
Antenna Gain of Jammer 2	G_{j2}
Path Loss of Jamming Link 2	L_{j2}

3.2. Beam Synthesis Method under Low-Bit Phase Array

This section introduces the beam synthesis method used in this paper. For phased array antennas, the phase shifter determines the precision of beam pointing. The higher the bit number of the phase shifter, the smaller the step of the phase difference between adjacent elements and the smaller the step of the synthesized beam pointing. Thus, the phase shifter used in most ground-based radars is large and expensive. Due to the limited size and load capacity of UAVs, the performance of a phase shifter is also very limited. There are many ways to achieve phased array miniaturization and beam synthesis with low-bit phase difference. The development of metamaterial is just an example [26]. In this paper, the antenna material of the UAV radar is not discussed. We only focus on the relatively flexible beam synthesis method for a UAV under low bit (2–3 bit) conditions.

A phased array antenna can control the beam pointing by controlling the phase of the array elements, which is a common method in an electric scanning antenna system. At present, one-dimensional phased array and two-dimensional phased array are widely used. They are composed of uniform linear array (ULA) and uniform plane array (UPA). ULA can realize beam scanning in a certain plane, while UPA can realize beam scanning in two-dimensional space. In order to simplify the model, without loss of generality, we first consider the low-bit beam synthesis method of ULA. Assuming that the UAVs in a swarm are arranged to form a ULA, the elements carried by each UAV may independently switch each element on or off.

It is assumed that the phase shifter can uniformly take values between 0°–360° at intervals of $360°/2^M$ when the coding number of phase shifters M is determined. To facilitate the representation of the discrete phase, we replace the phase values of 0°–360° with the code values of $0, 1, 2, \ldots, 2^{M-1}$, respectively. According to [27], two methods of repetitive coding and convolution can be applied. The repetitive coding method is performed by repeating the coded values in the ULA elements, and the convolution method is to perform the addition of the repeated codes, which corresponds to the convolution of the antenna pattern.

According to the convolution principle of the antenna pattern, if the beam pointings corresponding to the two different coding modes are θ_1 and θ_2 respectively, then the beam pointing after the code modular addition is pointed to θ_0.

$$\sin\theta_0 = \sin\theta_1 + \sin\theta_2 \tag{1}$$

This process is shown in Figure 2.

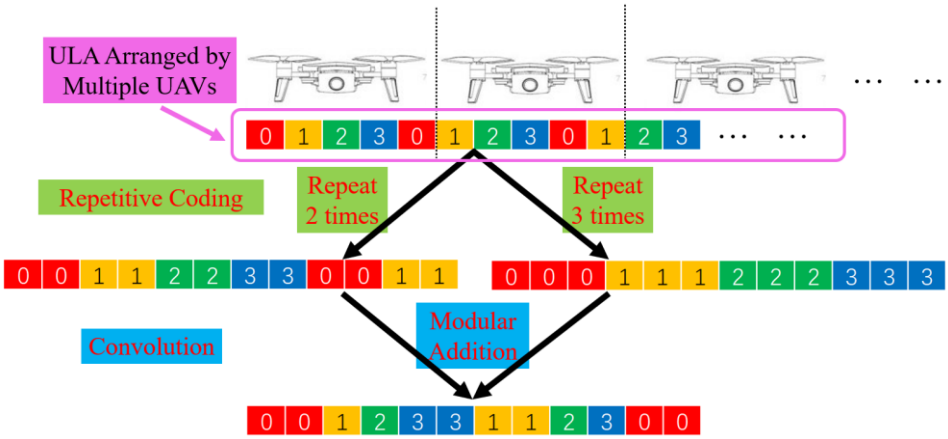


Figure 2. Schematic diagram of repetitive coding and convolution methods.

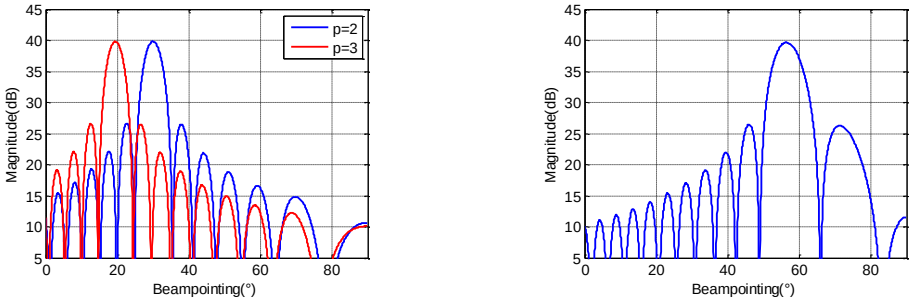
The spacing of array elements is d, the number of coded bits is M, the operating wavelength is $\lambda = 2^M d$, and the number of coding repetitions is p. In the real scenarios, the operating wavelength is related to the frequency. It is only set here to facilitate the interpretation of the beamforming principle.

Consider a case of 10 UAVs in a swarm and each UAV carries five elements, the number of ULA elements is $N = 50$. The spacing of array elements is $d = 1$cm and $M = 2$, Figure 3a shows the antenna pattern when $p = 2$ and $p = 3$, respectively. Figure 3b shows the antenna pattern after the two coding modes are added in Figure 3a. Since the beam pointing of a phased array antenna is:

$$\sin\theta = \frac{\varphi\lambda}{2\pi d} \tag{2}$$

where $\varphi = 2\pi/2^M$, and $\lambda = 2^M d$, thus:

$$\theta = \arcsin(1/p) \tag{3}$$



(**a**) (**b**)

Figure 3. (**a**) Antenna pattern for $p = 2$ and $p = 3$ (**b**) Antenna pattern after the two coding modes $p = 2$ and $p = 3$ are added.

The synthesized beam is now pointed to:

$$\theta_0 = \arcsin(\frac{1}{2} + \frac{1}{3}) = 56.44^{\circ} \tag{4}$$

From Figure 3b, the effectiveness of (1) can be verified. Therefore, this paper applies the antenna pattern generated by repetitive coding and convolution method under 3-bit condition to simulate. According to the basic theory of phased array antennas, the beam width can be controlled by the number of elements. When a wider beam needs to be generated, the UAV swarm can choose to

turn off some of the array elements to reduce the number of elements. There is a trade-off for the antenna pattern policy. For a radar UAV swarm, both the antenna pointing and the mainlobe width can be tuned. When a pattern with a wide mainlobe width is selected, its antenna gain also decreases. Therefore, the algorithm does not always choose the policy of a wide mainlobe antenna pattern.

In order to make the beam synthesis method more universal, we continue to consider the low bit beamforming method when the UAVs form a uniform plane array. Suppose the UAV swarm is arranged in an equivalent UPA as shown in Figure 4.

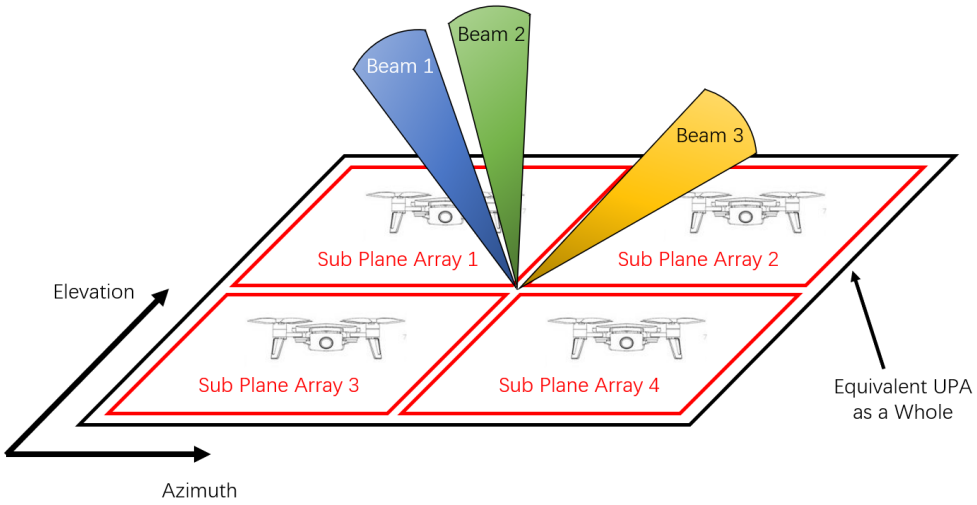


Figure 4. Equivalent uniform plane array (UPA) formed by 4 Sub-UAVs.

In Figure 4, the red box represents the UPA carried by each UAV, and they constitute the equivalent UPA as a whole. The equivalent UPA is represented by a black box. Because the element number of an equivalent UPA is much more than that of a single UAV, it has better beam performance, such as higher gain and narrower mainlobe width. The equivalent UPA can adjust the azimuth and elevation of the beam independently to realize the beam synthesis in two-dimensional space.

According to [28], the low-bit beam synthesis method of ULA can be extended to UPA. We assume that elevation is θ, azimuth is ϕ, the equivalent UPA is of m rows and n columns and the number of coded bits is M. Thus, the two-dimensional radiation pattern can be written as:

$$\left|E_1(\theta,\phi)\right| \approx \frac{\sin[\frac{n}{2}(\frac{2\pi p d_1}{\lambda}\cos\theta\sin\varphi - \frac{2\pi}{2^M})]}{\sin[\frac{1}{2}(\frac{2\pi p d_1}{\lambda}\cos\theta\sin\varphi - \frac{2\pi}{2^M})]} \tag{5}$$

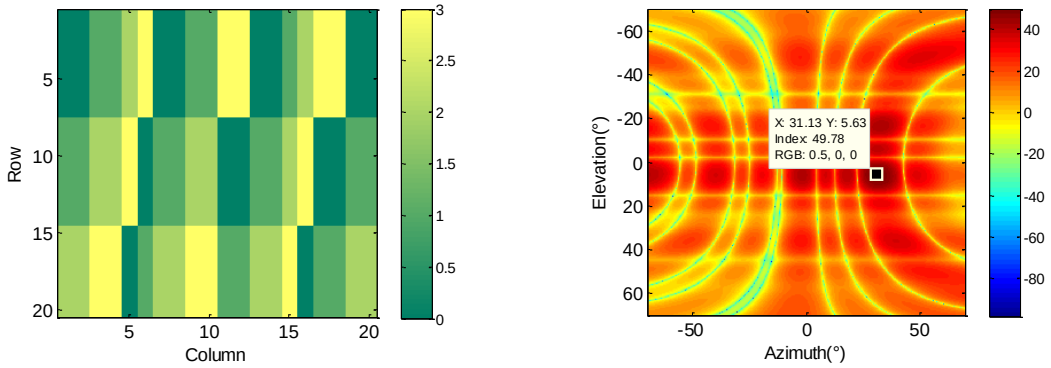
$$\left|E_2(\theta)\right| \approx \frac{\sin[\frac{m}{2}(\frac{2\pi q d_2}{\lambda}\sin\theta - \frac{2\pi}{2^M})]}{\sin[\frac{1}{2}(\frac{2\pi q d_2}{\lambda}\sin\theta - \frac{2\pi}{2^M})]} \tag{6}$$

We assume that $\mathbf{p} = [p_1, p_2, \cdots, p_a]$ represents the column coding (control azimuth) with the repetition times of $p_1, p_2, \cdots, p_a$ respectively, and $\mathbf{q} = [q_1, q_2, \cdots, q_b]$ represents the row coding (control elevation) with the repetition times of $q_1, q_2, \cdots, q_b$, respectively. According to the equivalent substitution theory of Figure 4 in [28], the elevation and azimuth of the beam can be designed by the following formula:

$$\sin\theta_0 = \frac{\lambda}{d_2 \times 2^M}\sum_{i=1}^{b}\frac{1}{q_i} \tag{7}$$

$$\sin\varphi_0 = \frac{\lambda}{d_1\cos\theta \times 2^M}\sum_{k=1}^{a}\frac{1}{p_k} \tag{8}$$

In order to show the above process more vividly, it is proved by the following simulation. The simulation parameters are as follows: $m = n = 2$, $p = [5]$, $q = [3, 2]$, $M = 2$, frequency is 10 GHz, element space $d = 1$ cm . Based on (7) and (8), we can get $\theta_0 = 6.15°$, $\phi_0 = 31.87°$. The simulated coding scheme is shown in Figure 5a, and the two-dimensional beam is shown in Figure 5b. At this time, the simulated beam pointing of two-dimensional beam is 5.63° (elevation) and 31.13° (azimuth) which is basically in line with theoretical design. In this way, we use four phase values only to obtain relatively high beam synthesis accuracy. It is shown that the low-bit beam synthesis method can also be applied to the equivalent UPA.



(**a**) (**b**)

Figure 5. Simulation result of 2-D beam synthesis with low-bit coding. (**a**) Coding scheme (**b**) 2-D radiation pattern.

4. Algorithm Design

Based on the modelling of Section 3, this section aims to solve the anti-jamming problem of UAV radar networking by introducing information representation theory and a modified Q-Learning algorithm. The objective of system optimization is to generate a reasonable multi-parameter policy of a radar system through the algorithm, so as to achieve the maximum information acquisition capability of a radar in any given jamming form. In the design of reinforcement learning algorithm, the rationality of the reward model is critical to the evaluation of generated policy.

Taking the following radar angle measurement case as an example, the target has multiple scattering points distributed in the azimuth direction, while the radar needs to measure the receiving angle of a strong scattering point from the target. In order to improve the angle measurement accuracy, a narrow beam should be formed, but this may result in a low SIR as the beam may not cover the target area. Therefore, using SIR as the reward function may not capture the information acquisition capability of the radar system. Therefore, we introduce the theory of radar information representation to set the reward properly.

4.1. Information Representation Reward Design

The calculation of information quantity is based on SIR, which is related to the system topology and parameter setting.

The SIR η received by the communication UAV is:

$$\eta = \frac{[(\sum_{i=1}^{5} P_R G_{Ri}^2 \sigma_i - L_{R1})F_R - L_{R2}] + [(\sum_{j=1}^{5} P_r G_{rj}^2 \sigma_j - L_{r1})F_r - L_{r2}]}{(P_{j1}G_{j1} - L_{j1}) + (P_{j2}G_{j2} - L_{j2})} \tag{9}$$

Under different networking modes, the mapping relationship between SIR η and information quantity has different forms. In many cases, the information quantity obtained by the radar is not only

determined by SIR. Detailed theory will be introduced by the radar information representation method as follows.

Target detection can be regarded as a process of information acquisition. The amount of information directly affects the accuracy of target parameter estimation. In statistical signal processing theory, Cramer-Rao lower bound (CRLB) is the most commonly used parameter to measure the accuracy of parameter estimation. CRLB is related to Fisher information in data determined by unknown parameters. It is assumed that the relationship between radar measurements $\boldsymbol{x}$ and the target parameter vector $\boldsymbol{\theta}$ is as follows:

$$\boldsymbol{x} = h(\boldsymbol{\theta}) + \boldsymbol{n} \tag{10}$$

where h and $\boldsymbol{n}$ denote the measurement function and the measurement noise, respectively. Then its Fisher information matrix $\boldsymbol{I}(\boldsymbol{\theta})$ is defined as:

$$[\boldsymbol{I}(\boldsymbol{\theta})]_{ij} = -E\left[\frac{\partial^2 \ln p(\boldsymbol{y}|\boldsymbol{\theta})}{\partial x_i \partial x_j}\right] \tag{11}$$

where E represents the mathematical expectation of a random variable.

For radar detection, since the LFM signal occupies a much wider bandwidth than information bandwidth, it can also obtain a large processing gain. Thus, LFM signal is the most commonly used signal form in radar detection. Taking LFM signal as an example, this paper analyses its information acquisition ability in "range-Doppler" measurement. LFM baseband signal has two main parameters: pulse width T and frequency modulation rate K. The time-domain signal can be written as follows:

$$s(t) = \frac{1}{T} \cdot \mathrm{rect}\left(\frac{2t}{T}\right) \cdot \exp\left(j2\pi K t^2\right) \tag{12}$$

where rect is a rectangular window function defined as:

$$\mathrm{rect}\left(\frac{t}{T_p}\right) = \begin{cases} 1 & -T_p/2 \le t \le T_p/2 \\ 0 & t < -T_p/2 \quad \text{or} \quad t > T_p/2 \end{cases} \tag{13}$$

In general, ranging and measuring Doppler shifts are performed simultaneously. Therefore, we consider the Fisher information under the "range-Doppler" measurement first. According to [29,30], assuming that the delay of the echo signal is τ and the Doppler shift is v, then the ambiguity function of the signal is represented by $A(\tau, v)$. The Fisher information matrix of the "range-Doppler" estimation is:

$$I = \eta \begin{bmatrix} -\frac{\partial^2 A(\tau,v)}{\partial \tau^2}\Big|_{\tau=0,v=0} & -\frac{\partial^2 A(\tau,v)}{\partial \tau \partial v}\Big|_{\tau=0,v=0} \\ -\frac{\partial^2 A(\tau,v)}{\partial v \partial \tau}\Big|_{\tau=0,v=0} & -\frac{\partial^2 A(\tau,v)}{\partial v^2}\Big|_{\tau=0,v=0} \end{bmatrix} \tag{14}$$

If the SIR of the echo is assumed to be η, $\boldsymbol{\theta} = [T, K]^T$, then the Fisher information matrix is:

$$\boldsymbol{I}(\boldsymbol{\theta}) = \eta \begin{bmatrix} \frac{1}{2T^2} + 8\pi^2 K^2 T^2 & 2\pi K T^2 \\ 2\pi K T^2 & \frac{T^2}{2} \end{bmatrix} \tag{15}$$

Finding the determinant of the Fisher information matrix, we can get the total amount of information generated by the LFM waveform for "range-Doppler" measurement. The amount of information I_{rd} when measuring the "range-Doppler" by the radar swarm is:

$$I_{rd} = \frac{\eta^2}{4} \tag{16}$$

Next, the information of the radar UAV swarm in angle measurement is calculated. Beam scanning method and equal signal method are two commonly used angle measurement methods in radar. Due to the relatively small number of elements in the UAV swarm, it may be difficult to effectively form the dual beam needed by the equal signal method. Thus, we mainly consider the beam scanning method here. When the radar UAV swarm performs beam scanning, the direction corresponding to the maximum amplitude of the pulses is the direction of the target. According to the measurement accuracy theory of the angle gate-tracking technology, under the optimal integration processing condition, the accuracy of angle measurement is:

$$\sigma_\theta \approx \frac{\theta_{0.5}}{2\sqrt{\eta}} \tag{17}$$

where $\theta_{0.5}$ is the half power width of the beam. According to the root mean square value of the angle measurement error, the Fisher information of the radar angle measurement I_a can be obtained:

$$I_a = \frac{1}{\sigma_\theta^2} \approx \frac{4\eta}{\theta_{0.5}^2} \tag{18}$$

On this basis, we can set the reward for reinforcement learning algorithm according to different working modes of radar networking. Information fusion theory needs to be used here, which is to obtain a total estimate by processing the estimated value of target state of each radar. At this time, the total Fisher information is the sum of the Fisher information of each radar. For two simultaneous ranging radars, if their received SIRs are η_1, η_2 respectively, its joint reward should be:

$$r = \frac{{\eta_1}^2}{4} + \frac{{\eta_2}^2}{4} \tag{19}$$

For one ranging radar and the other angle-measuring radar, its joint reward should be:

$$r = \frac{{\eta_1}^2}{4} + \frac{4\eta_2}{\theta_{0.5}^2} \tag{20}$$

It can be seen from (20) that $\theta_{0.5}^2$ makes angle-measuring radar no longer pursue high SIR completely, but seeks a balance between SIR and angular resolution. In practical applications, we take the logarithm of reward and normalize it to avoid the value of the information being too high or too low for observation.

4.2. DGQL (Double-Greedy Reinforcement Learning) Algorithm Based on Information Representation

This section will further improve the Q-Learning algorithm based on the information representation theory and the greedy algorithm. The DGQL framework under the condition of networked UAV radar systems is given.

Q-Learning is a commonly used reinforcement learning method for anti-jamming. The policy selection for Q-Learning is based on the Q table. The Q table has two dimensions: state and action. We consider all parameters of jammer 1, jammer 2 and target as the environment. The algorithm is required to give the corresponding policy of a radar system when given any environment that changes with time, so as to maximize the target information received by the UAV. Specifically, we consider the combination of jamming parameters at each time as the state, and the combination of radar system policies at each time is considered as an action. This process can be modelled as a Markov decision process (MDP). Traditional Q-learning uses tables to store Q values corresponding to each state and action, hence it has two limitations, namely scalability and continuous action. Therefore, deep Q network (DQN) and DDPG (deep deterministic policy gradient) methods were proposed in the reinforcement learning family.

DQN improves the scalability of Q-learning by adopting neural network may generate Q value to avoid tabular representations. However, the learning effect of DQN is related to the memory stored previously in a database. Simulation experiments usually show that DQN has a better learning effect and convergence speed when the correct actions and wrong actions account for half of each in the memory, which is not the case in the model of this paper. DDPG is a reinforcement learning method for continuous action by integrating a deep convolutional neural network as the simulation of policy function and Q function, that is, policy network and Q network, and then use the method of deep learning to train the above neural network. However, fine tuning of parameters in DDPG is important, such as the learning rate of the actor network and critical network, regularization parameters, neural network structure and so on. If the parameters are not adjusted properly, the algorithm convergence behavior is poor.

Therefore, as an initial contribution to adopt reinforcement learning in UAV-networked radar systems, we select Q-learning algorithm of ensured stability. This helps to gain more insights in the system model, problem dimension reduction and performance analysis, rather than devoting to fine-tuning the "black-box" in DQN and DDPG for high state-action parameter dimensions.

Firstly, the algorithm calculates the SIR received by the communication UAV. This is based on the combination of various radar policies and jamming policies according to the network topology. According to the radar working mode, the algorithm judges the calculation method of information, and then stores the R value in reserve. Then the modified Q-Learning algorithm is used to iterate the policy. Considering that the moving position of the communication UAV will cause a large energy loss, if its action is not constrained, it will generate results that are not realistic. Therefore, we further constrain the action selection by the double greedy algorithm.

The greedy algorithm is embedded in the Q-Learning algorithms to allow randomized action selection. It enables the agents to take randomized actions determined by probability distribution rather than Q table. This probability may be controlled by the threshold ε of the greedy algorithm. In our system, most of the energy of the communication UAV is used for motion, which should be controlled. Therefore, a soft constraint is set on the communication UAV that allows 1 or 2 moving steps with a higher probability and only set a low probability for further movement. We set two thresholds ξ_1, ξ_2, where $0 < \xi_1 < \xi_2 < 1$. When the communication UAV chooses the action, a random number rnd' within the range of [0, 1] is generated. Then it is compared with ξ_1, ξ_2 to determine the action within three action subsets. If $0 < rnd' < \xi_1$, the UAV selects an action in the action subset a_1 which contains one motion step. If $\xi_1 < rnd' < \xi_2$, the UAV selects within the action subset a_2 containing two motion steps. If $\xi_2 < rnd' < 1$, the UAV will select the action in action subset a_3 of more than two motion steps.

In this way, the probability of choosing one step, two steps and more than two steps for the communication UAV is ξ_1, $\xi_2 - \xi_1$ and $1 - \xi_2$ respectively. The threshold values ξ_1, ξ_2 may be set according to the specific UAV setups and task requirements. When ξ_2 approaches 1, the motion of the communication UAV is strictly limited to two steps or less. This process is shown in Figure 6 and the pseudo-code of the DGQL is shown in Algorithm 1.

Algorithm 1. Double-Greedy Reinforcement Learning Algorithm

Initialize r table set $r(i, j)$
for (i, j) in M radar policies and N jamming policies **do**
  Generate η_1, η_2 via network topology
  if mode 1 **then**
    $r(i, j) = \frac{\eta_1^2+\eta_2^2}{4}$
  else if mode 2 **then**
    $r(i, j) = \frac{4(\eta_1+\eta_2)}{\theta_{0.5}^2}$
  end if
end for
Set $Q(s, a)$ arbitrarily
for each episode **do**
  Initialize s
  for each step of episode **do**
    Generate random number *rnd* and *rnd*′
    if $rnd<\varepsilon$ **then**
      Extract action subset a_1, a_2, a_3 from a
      Generate random number *rnd*′
      if $0 \leq rnd' < \xi_1$ **then**
        Choose a from (s, a_1) using policy derived from Q
      else if $\xi_1 < rnd' < \xi_2$ **then**
        Choose a from (s, a_2) using policy derived from Q
      else:
        Choose a from (s, a_3) using policy derived from Q
      end if
    else
      Choose a randomly from s
    end if
    Take action a, observe r, s'
    $Q(s, a) \leftarrow Q(s, a) + \alpha[r + \gamma \max_{a'} Q(s', a') - Q(s, a)]$
    $s \leftarrow s'$
    until s is terminal
  end for
end for

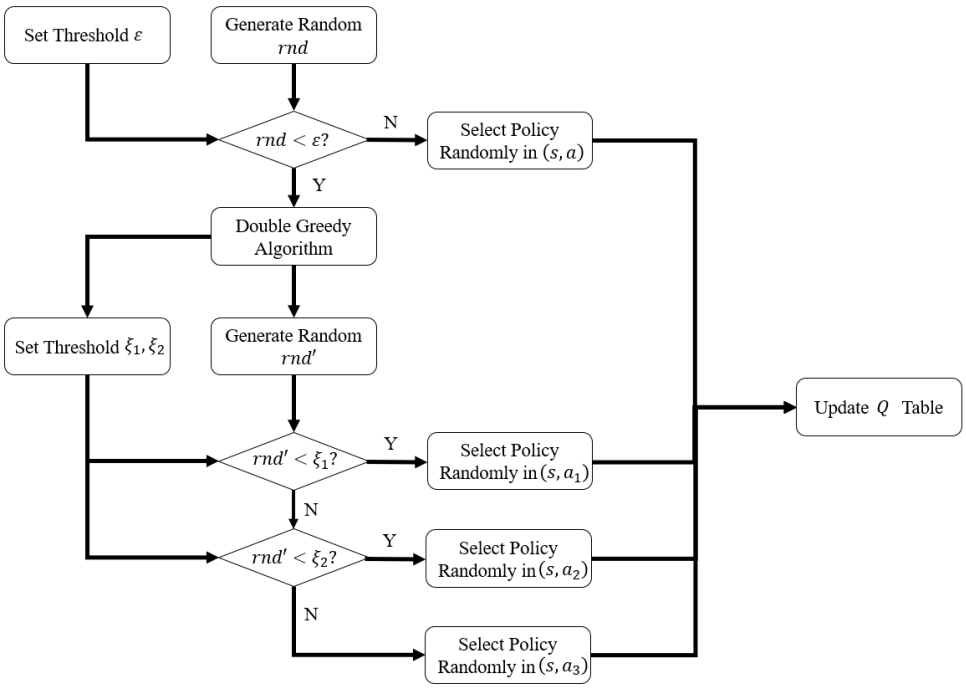


Figure 6. Flow chart of double greedy algorithm.

5. Simulation Results

5.1. Map Gridding

In order to quantify the location and beam pointing of each agent more conveniently, gridded map modeling is introduced. As shown in Figure 7. The two radar swarms are represented by two discrete points, which can be combined or separated according to the task requirements. The communication UAV is represented by a discrete point, which may move on optional positions in Figure 7. The target consists of a self-defense jammer (jammer 1) and five main scattering points, each of which is numbered. Considering the intensity difference of each scattering on actual targets, the intensity of scattering point 3 is set to be slightly stronger than the other four, and the other four are the same. The target may move within a certain range. Jammer 2 is represented by a fixed point. The azimuth definition is shown in Figure 7.

The tunable parameters of each agent are shown in Table 2. In order to show the angle more clearly, the azimuth is calculated based on the y-axis after 1% deformation of the actual distance. That is to say, 100m of y-axis and 1m of x-axis have the same length in Figure 7.

Table 2. Tunable parameters of agents.

Parameters	Range	Unit
Transmitting Frequency of Radar UAV Swarm	9–9.8	GHz
Forwarding Frequency of Radar UAV Swarm	10.2–11	GHz
Mainlobe Width of Radar UAV Swarm	5–70	degree
Position of Communication UAV	[0, 18], [2, 18], [4, 18], [6, 18], [8, 18], [10, 18], [12, 18]	[x (1 m),y (100 m)]
Transmitting Frequency of Jammer 1	9–9.8	GHz
Position of Jammer 1	[2, 4], [4, 4], [6, 4], [8, 4], [10, 4]	[x (1 m),y (100 m)]
Transmitting Frequency of Jammer 2	10.2–11	GHz
Beampointing of Jammer 2	60–83	degree

When the detection requirements are different, the networking topology of radar UAV swarms is different. Considering the problem of radar target location, this section presents two scenarios to

verify the effectiveness of the algorithm. In the simulation experiment, the algorithm will generate the corresponding radar multi-parameter policy according to the jamming policy of 10 time steps. In order to control the motion of the communication UAV, we set the greedy coefficient as follows: $\xi_1 = 0.6$, $\xi_2 = 0.95$. Thus, let the communication UAV take one step policy with 60% probability and two steps policy with 35% probability to prevent the communication UAV from consuming excessive energy.

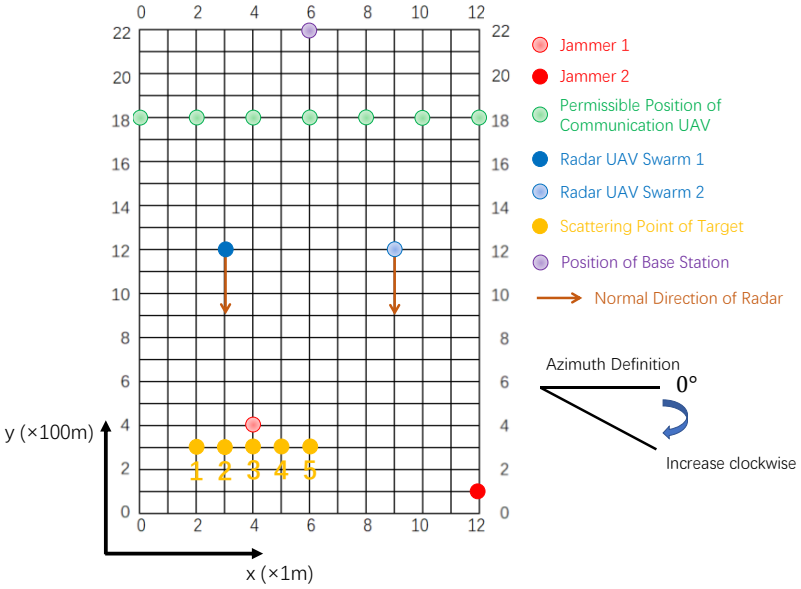


Figure 7. Schematic diagram of gridded modeling.

5.2. Collaborative Target Ranging

In this scenario, the radar aggregates the ranging measurements of multiple scattering points for target localization. Range measurement is carried out by using two radar UAV swarms. Two radar UAV swarms are separated from each other, and the target position is determined by the intersection area of beams form by the two swarms. In this scenario, because the distance spanning from the target to the radar is significantly larger than the horizontal size of the target, all scattering points of the target are in the same range profile and may be treated as a whole. Therefore, the radar UAV swarms tend to form wider beams to cover more scattering points, which results in improved SIRs and ranging accuracy.

In the default policy of this group of experiments, the radar transmitting frequency is maintained at 9.4 GHz. The forwarding frequency is always 10.6 GHz. The width of both mainlobe is 50°. The relay position is always at [6,18]. The scanning mode of beam pointing is from right to left, and then from left to right.

In order to display the results more vividly, we give the schematic diagram of time sequence (Figure 8), frequency chart (Figure 9) and normalized information chart (Figure 10). In the schematic diagram of the time sequence, the dot represents the position of each agent and the line represents the mainlobe range of each agent's 3 dB antenna pattern. Purple represents the communication UAV. Brown represents the radar UAV swarm (the distance between the UAV swarm is very small, so the whole swarm is replaced by a point). Red represents the jammer 1. Blue represents the target scattering point distribution. Pink represents the jammer 2. The legends of the schematic diagram of the time sequence are shown in Figure 11. In order to show the results more clearly, we exclude the base station shown in Figure 7. In order to illustrate the effectiveness of this discrete modelling provided by Q-Learning, we will give the results of normalized information quantity after enlarging the scale of the problem. It should be noted that the maximum and minimum information quantity of each experiment remain unchanged while increasing the scale of the problem. Take the scenario of collaborative target ranging for example, after taking the natural logarithm of each group's information

quantity, the maximum value is 6.2107, and the minimum value is −8.2248, so it is still meaningful to compare the normalized information quantity.

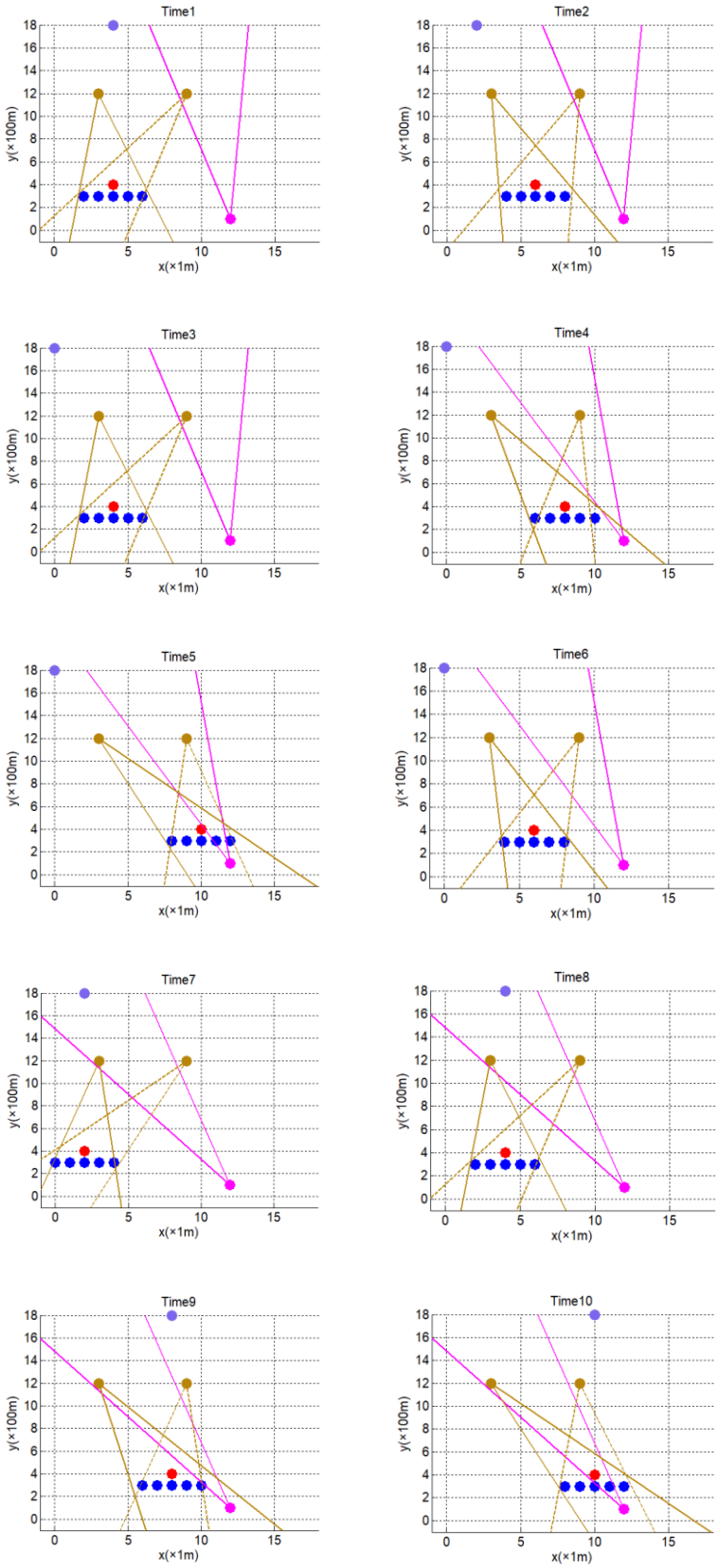


Figure 8. Schematic diagram of time sequence.

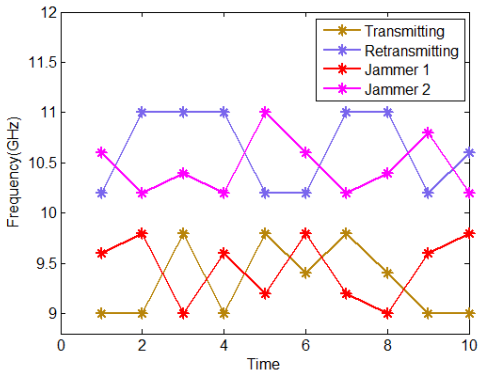


Figure 9. Frequency chart of collaborative target ranging.

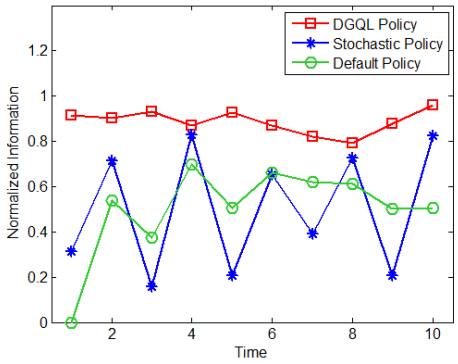


Figure 10. Normalized information chart of collaborative target ranging.

Figure 11. Legends of schematic diagram of time sequence.

Figure 8 illustrates the policy of UAV swarm radar beam synthesis and the communication UAV motion policy generated. According to the theory in Section 4, the ranging information is only related to SIR, so the beams tend to cover as many scattering echoes as possible. Simulations show that the beam pointing and mainlobe width fits the target area well. As the beam of jammer 2 scans from right to left, the communication UAV makes effective move to avoid jamming, while restricted its motion range to preserve energy as guided by the double greedy algorithm. Figure 9 illustrates the frequency policies adopted by the radar UAV swarms to avoid jamming.

In order to better illustrate the effectiveness of DGQL, we introduce the information of the stochastic policy and the default policy as benchmarks. Stochastic policy refers to the random selection of all radar policies at each time. Default policy refers to a policy choice when the radar UAV swarm does not have any prior information about the target. A common default policy is that the radar keeps the beam scanning back and forth in space, keeps the transmission frequency and mainlobe width unchanged, and does not change the position of the communication UAV (always in the center). Figure 10 shows that DGQL generates significantly more information than stochastic policies and default policies.

The normalized information curve shown in Figure 10 is based on 1125 actions, which is realized by 2.5×10^5 times of iteration. In order to analyze the influence of the number of optional actions on

information quantity, Figure 12 shows the normalized information curve when the number of actions is 1680, 2800, 4375, 6125 and 8575, respectively. The legends shown in Figure 12 are the number of actions in each group. Their corresponding iteration times are 3.7×10^5, 6.2×10^5, 9.7×10^5, 13.6×10^5 and 19.0×10^5, respectively. The adjusted actions are the transmitting frequency of the radar UAV swarm, the forwarding frequency of the radar UAV swarm and the mainlobe width of radar UAV swarm. Each parameter keeps uniform step in their own allowable range.

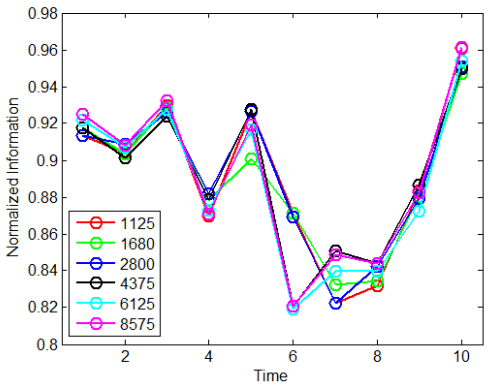


Figure 12. Normalized information chart of collaborative target ranging.

By comparing the curves of 1125 actions, the results show that in this scenario, increasing the number of actions cannot effectively improve the information acquisition ability of radar. This is because even when there are only 1125 actions, the effective policies (refers to the policy that enables the radar to obtain high information quantity) are still included in the action set. The algorithm can generate some of these policies by iterations.

Thus, this proves that if the problem scene can be modeled in a reasonable discrete way and the effective policy can be included in the action set, the computing time can be saved without losing the amount of information obtained by the radar, which is cost-effective in our model.

5.3. Collaborative Scattering Point Localization

In this scenario, the radar needs to locate a strong scattering point on the target. Let radar UAV swarm 1 measure angle and radar UAV swarm 2 measure range, because we assume that only radar UAV swarm 2 has the ability to generate narrow beams of angle measurement. When measuring distance, the target is regarded as a whole, which is the same case in collaborative target ranging. When measuring the angle, the target is regarded as 5 discrete scattering points. The strong scattering points are then located according to distance and angle. The angle measurement adopts a beam scanning method. At this time, two UAV swarms need to be in close positions. In order to highlight strong scattering points, 5 scattering points of the target are set to different intensities. The intensity of scattering point 3 in Figure 7 is slightly stronger than that of the other four scattering points. In this way, the angle measurement is concretized as the angle tracking of a key point on a wide azimuth target by radar UAV swarm 2.

The default policy of this experiment is the same as that of collaborative target ranging, except that the width of two mainlobes are 50° and 10°, respectively. This is because, for the default policy of angle measuring radar, its mainlobe is relatively narrow.

In particular, since the two radar UAV swarms are relatively close at this time, we use the same point to represent them in Figure 13. At this time, the antenna pattern of UAV swarm 2 shown in Figure 13 does not mean that the beam always points to scattering point 3. It represents the mainlobe width when the beam of UAV swarm 2 sweeps through scattering point 3 instead.

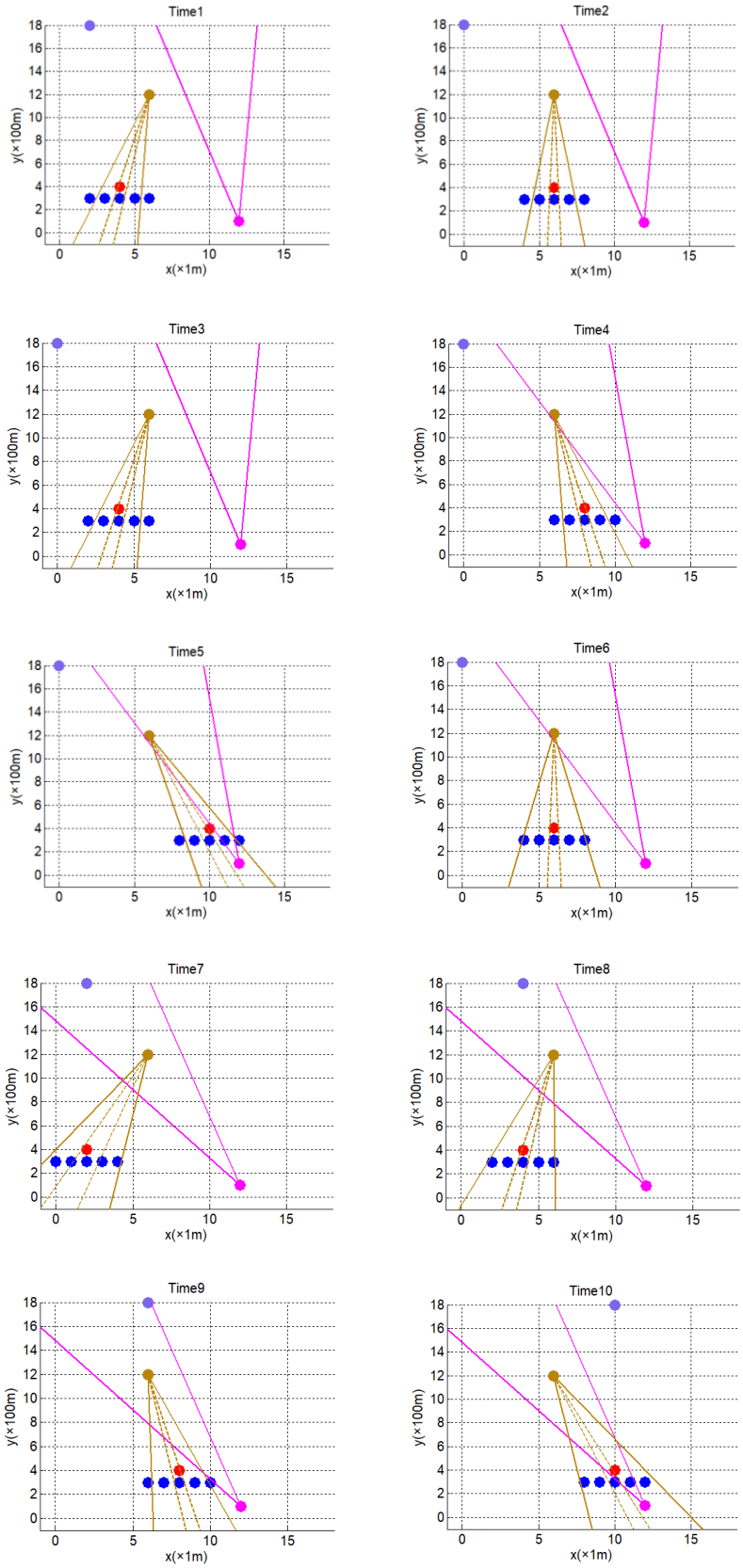


Figure 13. Schematic diagram of time sequence.

In Figure 13, the mainlobe of the angle measuring UAV swarm (UAV swarm 2) is very narrow, only 5°. This is because for the beam scanning method, the narrower the mainlobe width, the larger

the angle measuring information. At this point, the policy of the angle measuring UAV swarm no longer pursues the high SIR, but seeks a balance between the SIR and the mainlobe width. In this way, the angle-measuring UAV swarm realizes angle tracking of the key part (scattering point No. 3) of the target.

Figure 14 also proves that the frequency policy of the radar UAV swarm and the communication UAV can still be correctly generated when the ranging and the angle measuring modes coexist. Figure 15 shows the effectiveness of DGQL. In particular, we show the result of the policy based on the reward generated by SIR only. At this time, radar UAV swarm 2 adopts a mainlobe width of 20°. Although it covers more scattering points and gets higher SIR. The information quantity is still slightly lower than the DGQL policy. Wide mainlobe policy reduces the accuracy of radar angle measurement, which fully proves the validity of information representation theory. Figure 16 also proves that in the scenario of collaborative scattering point localization, increasing the number of action combinations does not significantly improve the information acquisition ability of the radar.

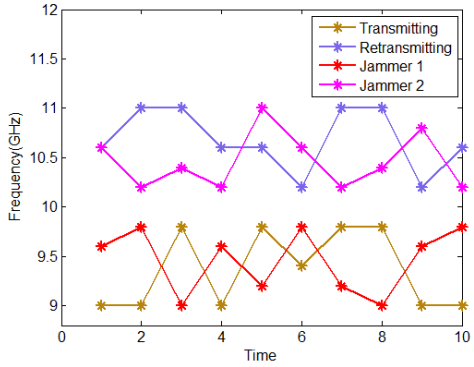


Figure 14. Frequency chart of collaborative scattering point localization.

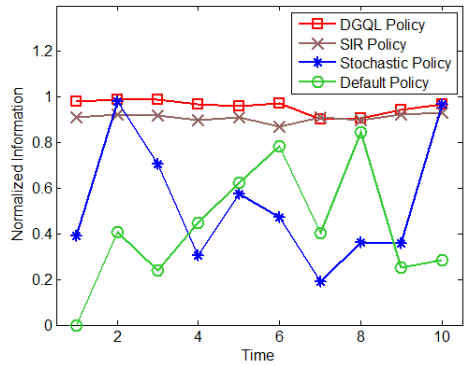


Figure 15. Normalized information chart of collaborative scattering point localization.

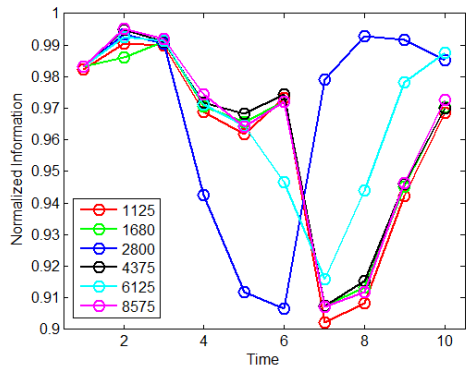


Figure 16. Normalized information chart of collaborative scattering point localization.

6. Conclusions

In this paper, a novel Q-Learning based anti-jamming algorithm for networked UAV radar systems is proposed for optimizing the information acquisition capabilities, which is captured by the radar information representation. Due to the payload size and energy constraints of UAVs, the beam synthesis method under low bit phase condition is adopted. Both the angle measurement and range measurement scenarios are constructed to evaluate the effectiveness of the proposed algorithms. Simulation results show that the algorithm may facilitate effective anti-jamming and adapt to different tasks. In this paper, the Q-Learning based algorithm is selected due to its stability and interpretability, while it also discretizes the state variables in space, frequency and antenna beamforming, which may be continuous in general. In our future work, we will consider the reinforcement learning algorithm to generate continuous policies, such as the deep deterministic policy gradient (DDPG).

Author Contributions: Methodology, Q.W.; Supervision, H.W., X.L. and J.P.; Writing—original draft, Q.W.; Writing—review and editing, B.Z.

Funding: This work is funded by the National Natural Science Foundation of China (No. 91648204, No. 61601486), Research Programs of National University of Defense Technology (No. ZDYYJCYJ140601), and State Key Laboratory of High Performance Computing Project Fund (No. 1502-02).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Xing, Q.; Zhu, W.-G.; Jia, X. Intelligent countermeasure design of radar working-modes unknown. In Proceedings of the IEEE International Conference on Signal Processing, Louisville, KY, USA, 6–8 December 2018; IEEE: Piscataway, NJ, USA, 2018.
2. Zhang, J. Study on wideband sparse spectrum waveform for anti-interception and anti-jamming countermeasure. In Proceedings of the CIE International Conference on Radar, Seattle, WA, USA, 8–12 May 2017; IEEE: Piscataway, NJ, USA, 2017.
3. Song, X.; Willett, P.; Zhou, S.; Luh, P.B. The MIMO radar and jammer games. *IEEE Trans. Signal Process.* **2011**, *60*, 687–699. [CrossRef]
4. Deligiannis, A.; Lambotharan, S. A Bayesian game theoretic framework for resource allocation in multistatic radar networks. In Proceedings of the 2017 IEEE Radar Conference (RadarConf), Seattle, WA, USA, 8–12 May 2017; IEEE: Piscataway, NJ, USA, 2017.
5. Lan, X.; Li, W.; Wang, X.; Yan, J.; Jiang, M. MIMO radar and target Stackelberg game in the presence of clutter. *IEEE Sens. J.* **2015**, *15*, 6912–6920. [CrossRef]
6. Jameel, F.; Wyne, S.; Kaddoum, G.; Duong, T.Q. A comprehensive survey on cooperative relaying and jamming strategies for physical layer security. *IEEE Commun. Surv. Tutor.* **2018**, *21*, 2734–2771. [CrossRef]
7. Xiong, J.; Cheng, L.; Ma, D.; Wei, J. Destination-Aided Cooperative Jamming for Dual-Hop Amplify-and-Forward MIMO Untrusted Relay Systems. *IEEE Trans. Veh. Technol.* **2016**, *65*, 7274–7284. [CrossRef]
8. Kang, L.; Bo, J.; Hongwei, L.; Siyuan, L. Reinforcement Learning based Anti-jamming Frequency Hopping Strategies Design for Cognitive Radar. In Proceedings of the 2018 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), Qingdao, China, 14–16 September 2018; IEEE: Piscataway, NJ, USA, 2018.
9. Wang, Y.; Zhang, T.; Xu, L.; Tian, T.; Kong, L.; Yang, X. Model-free Reinforcement Learning based Multi-stage Smart Noise Jamming. In Proceedings of the 2019 IEEE Radar Conference (RadarConf), Boston, MA, USA, 22–26 April 2019; IEEE: Piscataway, NJ, USA, 2019.
10. Li, H.; Ye, W. A Study on the Influence of Bit Error Ratio against Jamming Signal Ratio under Different Channel Jamming. In Proceedings of the International Symposium on Computational Intelligence & Design, Hangzhou, China, 9–10 December 2017; IEEE: Piscataway, NJ, USA, 2017.
11. Baker, C.J. Intelligence and radar systems. In Proceedings of the Radar Conference 2010, Arlington, VA, USA, 10–14 May 2010.

12. Xingyu, X.; Daoliang, H.; Li, Y.; Xiaoyang, W. Optimal Waveform Design for Smart Jamming Focused on CA-CFAR. In Proceedings of the 2017 International Conference on Computer Network, Electronic and Automation (ICCNEA), Xi'an, China, 23–25 September 2017; IEEE: Piscataway, NJ, USA, 2017.
13. Łabowski, M.; Kaniewski, P. A method of swath calculation for side-looking airborne radar. In Proceedings of the 2018 14th International Conference on Advanced Trends in Radioelecrtronics, Telecommunications and Computer Engineering (TCSET), Lviv, Ukraine, 20–24 February 2018; IEEE: Piscataway, NJ, USA, 2018.
14. Bhattacharya, S.; Başar, T. *Game-theoretic analysis of an aerial jamming attack on a UAV communication network. American Control Conference (ACC), Baltimore, MD, USA, 30 June–2 July 2010*; IEEE: Piscataway, NJ, USA, 2010.
15. Xiao, L.; Lu, X.; Xu, D.; Tang, Y.; Wang, L.; Zhuang, W. UAV Relay in VANETs Against Smart Jamming with Reinforcement Learning. *IEEE Trans. Veh. Technol.* **2018**, *67*, 4087–4097. [CrossRef]
16. Li, H.; Luo, J.; Liu, C. Selfish Bandit based Cognitive Anti-jamming Policy for Aeronautic Swarm network in Presence of Multiple Jammert. *IEEE Access* **2019**, *7*, 30234–30243. [CrossRef]
17. Li, Z.; Lu, Y.; Shi, Y.; Wang, Z.; Qiao, W.; Liu, Y. A Dyna-Q-Based Solution for UAV Networks Against Smart Jamming Attacks. *Symmetry* **2019**, *11*, 617. [CrossRef]
18. Rahmes, M.; Chester, D.; Clouse, R.; Hunt, J.; Ottoson, T. Cooperative cognitive electronic warfare UAV game modeling for frequency hopping radar. In *Unmanned Systems Technology XX*; International Society for Optics and Photonics: Orlando, FL, USA, 2018; Volume 10640.
19. Cevik, P.; Kocaman, I.; Akgul, A.S.; Akca, B. The Small and Silent Force Multiplier: A Swarm UAV—Electronic Attack. *J. Intell. Robot. Syst.* **2013**, *70*, 595–608. [CrossRef]
20. Xing, Q.; Zhu, W.-G.; Jia, X. Intelligent radar countermeasure based on Q-learning. *Syst. Eng. Electron.* **2018**, *40*, 1031–1035.
21. Liu, P.; Liu, Y.; Huang, T.; Lu, Y.; Wang, X. Cognitive Radar Using Reinforcement Learning in Automotive Applications. *arXiv* **2019**, arXiv:1904.10739.
22. You, S.; Diao, M.; Gao, L. Deep reinforcement learning for target searching in cognitive electronic warfare. *IEEE Access* **2019**, *7*, 37432–37447. [CrossRef]
23. Wang, L.; Fortunati, S.; Greco, M.S.; Gini, F. Reinforcement learning-based waveform optimization for MIMO multi-target detection. In Proceedings of the 2018 52nd Asilomar Conference on Signals, Systems, and Computers, Pacific Grove, CA, USA, 28–31 October 2018; IEEE: Piscataway, NJ, USA, 2018.
24. Alberge, F. Deep Learning Constellation Design for the AWGN Channel with Additive Radar Interference. *IEEE Trans. Commun.* **2018**, *67*, 1413–1423. [CrossRef]
25. Huo, Y.; Tian, Y.; Ma, L.; Cheng, X.; Jing, T. Jamming strategies for physical layer security. *IEEE Wirel. Commun.* **2017**, *25*, 148–153. [CrossRef]
26. Cui, T.J. Microwave metamaterials–From passive to digital and programmable controls of electromagnetic waves. *J. Opt.* **2017**, *19*, 084004. [CrossRef]
27. Liu, S.; Cui, T.J.; Zhang, L.; Xu, Q.; Wang, Q.; Wan, X.; Gu, J.Q.; Tang, W.X.; Qing, Q.M.; Han, J.G.; et al. Convolution operations on coding metasurface to reach flexible and continuous controls of terahertz beams. *Adv. Sci.* **2016**, *3*, 1600156. [CrossRef] [PubMed]
28. Wu, Q.; Cheng, Y.; Li, X.; Wang, H. Beam Synthesis with Low-Bit Reflective Coding Metamaterial Antenna: Theoretical and Experimental Results. *Int. J. Antennas Propag.* **2018**, *2018*, 1–9. [CrossRef]
29. Cheng, Y.; Wang, X.; Caelli, T.; Li, X.; Moran, B. On information resolution of radar systems. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 3084–3102. [CrossRef]
30. Cheng, Y.; Wang, X.; Morelande, M.; Moran, B. Information geometry of target tracking sensor networks. *Inf. Fusion* **2013**, *14*, 311–326. [CrossRef]