A Reinforcement Learning-based Intelligent Learning Method for Anti-active Jamming in Frequency Agility Radar

Jingjing WEI [1,2], Lei YU [1,2], Yinsheng WEI [1,2], Rongqing XU [1,2]

[1] School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China
[2] Key Laboratory of Marine Environmental Monitoring and Information Processing, Ministry of Industry and Information Technology, Harbin, China

weijingjing@stu.hit.edu.cn, yu.lei@hit.edu.cn*, hitweiysgroup@163.com, xurongqing@hit.edu.cn

Submitted June 12, 2024 / Accepted October 1, 2024 / Online first November 12, 2024

Abstract. *Active jamming's flexibility and variability pose significant challenges for frequency-agility radar (FAR) detection, as it can continuously intercept and retransmit radar signals to suppress or deceive the radar. To tackle this, we propose an intelligent learning method for FAR based on reinforcement learning (RL), integrating signal processing with compressed sensing (CS). We introduce an inter-pulse carrier-frequency hopping combined with intra-pulse sub-frequency coding (IPCFH-IPSFC) signal model to address time-domain discontinuities caused by active jamming, enabling effective mutual masking of pulses through agile waveform parameters. We develop jamming signal models and design four jamming strategies based on two common types of active jamming, providing essential data for the FAR intelligent learning method. To enhance FAR's adaptive anti-jamming and target detection performance, we propose an RL-based intelligent learning model. This model includes five submodules: signal processing, anti-jamming evaluation, target detection, optimization constraint design, and optimization algorithm design. We apply a proximal policy optimization combined with a generative pre-trained transformer (PPO-GPT) to solve this model, allowing FAR to adaptively learn jamming strategies and optimize IPCFH-IPSFC waveform parameters for effective anti-jamming. Simulation results confirm that our method achieves robust performance and rapid convergence, finding optimal anti-jamming strategies in just 215 training iterations. The FAR effectively counteracts jamming while accurately estimating target range and velocity.*

Keywords

Intelligent learning, anti-active jamming, frequency agility radar, reinforcement learning, compressed sensing

1. Introduction

As jamming technology advances, future electronic warfare (EW) equipment will learn and adapt to the environment. Radar workers encounter a more demanding and complicated work environment [1]. Digital radio frequency memory (DRFM) and active jamming modes are being developed quickly. This lets jammers make flexible and variable jamming signals that can precisely target the radar operating frequency, which makes it much harder for radar to find targets [2]. Jamming types include interrupted sampling repeater jamming (ISRJ), smart noise, etc. ISRJ switches sampling and modulating the radar-transmitted pulses, causing false targets [3]. Instead of traditional noise suppression jamming, smart noise jamming convolves or multiplies the received radar signal with narrowband noise [4]. It matches the radar signal and automatically targets the carrier frequency. With suppression and deception, these active jamming modes may construct several false targets and be coherent with radar signals. Researching radar anti-active jamming technology to increase radar detection is crucial.

Radar uses waveforms to perceive the environment, and anti-jamming and target detection performance are highly connected to waveforms [5]. Traditional waveform modulation is straightforward, and it is simple to identify by jammers in intelligent countermeasure scenarios [6]. Complex modulation techniques and parameter hopping technologies make the agile waveform less predictable, which makes it harder for jammers to intercept and easier for targets to tell apart [7], [8]. Frequency-agile technology usually randomly selects the carrier frequency or transmits the waveform in a fixed mode, and more attention is paid to waveform design and parameter estimation [9], [10]. The jammer uses DRFM to generate different jamming modes. If jamming learning techniques are not incorporated into FAR anti-jamming research, it will not fulfill

DOI: 10.13164/re.2024.0681

intelligent countermeasure demands [11]. Some researchers call the frequency agility design problem a deterministic optimization problem; however, it involves the artificial calculation of jamming and target properties to get the ideal waveform parameters [12], [13]. Active jamming varies rapidly, making real-time manual parameter estimations impossible for radars with limited resources. There is an urgent need for adaptive implementation methods.

Some researchers use RL in radar anti-jamming to boost radar's learning capabilities in jamming environments [14], [15]. Intelligent radar anti-jamming decision-making is solved by treating radar and jamming as a sequence decision-making issue [16], [17]. A new frequency-hopping strategy design method using Q-learning and deep Q network (DQN) algorithms is shown in case cognitive radar can't reliably find the smart jamming mode [18]. The reward function for finding something in a coherent processing interval (CPI) is made by processing the FAR echo signal in a way that isn't coherent, and a DQN-based approach for optimizing the frequency waveform is given [19]. To examine two FAR hopping techniques, a DQN with a long-short-term memory (LSTM) algorithm is developed to address jammer dynamics uncertainty [20], [21]. A proximal policy optimization (PPO) using LSTM is presented to autonomously develop efficient anti-jamming tactics for variable and intelligent jamming strategies [22]. The FAR error problem in the electromagnetic game is caused by wrong jamming state monitoring. To fix this, a strong anti-jamming strategy learning system was created using imitation learning and Wasserstein robust reinforcement learning (WR^2L) [23]. A FAR anti-jamming approach uses RL and supervised learning to tackle the problem of the RL algorithm not handling non-stationary jamming tactics [24]. One problem with the above studies is that the optimal design is based on the inter-pulse carrier frequency agility of FAR, which makes it difficult to suppress the ISRJ of intra-pulse sampling and repeaters. Therefore, frequency-agile waveform and parameter-intelligent learning methods to deal with active jamming created by DRFM must be researched quickly.

The FAR agile waveform design must also handle coherent processing of the echo signal to determine target parameters. Since FAR carrier frequency hopping causes phase discontinuity, fast Fourier transform (FFT) is no longer a viable approach for calculating target information [25]. Non-coherent processing is used to calculate target detection probability [26]. A signal-to-noise ratio (SNR) weighting technique coherently accumulates echoes of the same carrier frequency. Some researchers use CS for radar target detection [27]. CS transforms the detection challenge of moving targets into a sparse signal estimation problem, and signal sparse reconstruction in CPIs is utilized to estimate target information [28]. High-resolution range-Doppler reconstruction of random frequency hopping and pulse repetition frequency (PRF) agility for FAR was suggested using CS sparse optimization [9]. Two-dimensional sparse reconstruction using a conjugate gradient solver is presented to efficiently recreate high-resolution range-Doppler pictures from frequency and PRF agility waveforms [12]. The above studies support FAR target detection; however, they do not address the unique challenges of applying CS to the FAR in jamming environments. In the context of active jamming, FAR must research ways to suppress jamming and properly estimate target information.

The above mentioned studies did not consider how FAR, when confronting active jamming with dynamic jamming strategies, can learn jamming strategies through multi-round interactive learning and adaptively generate anti-jamming strategies. At the same time, they seldom consider the issue of ensuring efficient target detection and accurate information estimation while successfully countering jamming. Therefore, in response to the flexible and variable nature of active jamming patterns, there is an urgent need to develop intelligent anti-jamming methods for FAR, while also addressing the challenge of estimating target distance and velocity information due to the difficulty of directly accumulating phase-coherent FAR echo signals. To address these issues, we propose an RL-based intelligent learning method for FAR that integrates radar signal processing techniques with RL and CS technologies. This approach enables FAR to utilize adaptive waveform strategies to counteract the evolving patterns of active jamming. The main contributions of this work are summarized as follows:

- We have designed an IPCFH-IPSFC signal model for FAR, which facilitates pulse-to-pulse and sub-pulse mutual masking in complex and dynamic active jamming environments. Based on two typical types of active jamming—ISRJ and smart noise jamming—we have developed four jamming strategies, providing a data foundation for the intelligent learning method for FAR.
- We have developed an RL-based intelligent learning model for FAR. This model comprises five submodules: signal processing, anti-jamming evaluation, target detection, optimization constraint design, and optimization algorithm design. In complex jamming environments, this model enables effective signal accumulation of FAR target echoes and accurate estimation of target distance and velocity. By employing a PPO-GPT algorithm to solve the intelligent learning model, we not only enhance FAR's adaptive anti-jamming performance but also improve target detection capabilities. The proposed RL-based intelligent learning model for FAR allows continuous learning of jamming strategies, adaptive adjustment of IPCFH-IPSFC waveform parameters to generate optimal anti-jamming strategies, and precise estimation of target distance and velocity.
- Finally, we conducted simulations to validate the proposed RL-based intelligent anti-active jamming method for FAR and compared it with alternative approaches. The results demonstrate that this method

exhibits strong robustness, fast convergence, optimal anti-jamming capability, and the most accurate estimation of target range and velocity. It not only enables FAR to adaptively generate anti-jamming strategies and successfully block jamming signals from entering the FAR receiver, but also allows for precise estimation of target range and velocity.

The rest of this paper is organized as follows: The agile waveform and jamming strategies are explained in Sec. 2. The intelligent learning method design based on PPO-GPT algorithm is described in Sec. 3. Section 4 shows the simulation results, followed by the conclusions presented in Sec. 5.

2. Agile Waveform and Jamming Strategies

2.1 Agile Waveform of FAR

We created the IPCFH-IPSFC waveform with intelligently agile parameters to protect FAR from active jamming and accurately measure target range and velocity. In particular, the carrier frequency hopping provides mutual cover between pulses. Intra-pulse uses linear frequency modulation (LFM). Sub-frequency coding in each pulse ensures sub-pulse mutual cover. Inter-pulse carrier frequency hopping combined with intra-pulse sub-frequency coding improve radar anti-jamming. Therefore, we assume that the FAR transmits N pulses in a CPI, and the k-th sub-pulse of the n-th pulse is $s_{\text{T\_sub}}(\hat{t},n,k)$.

$$s_{\text{T\_sub}}(\hat{t},n,k)= \text{rect}\left(\frac{\hat{t}-T_{\text{sub}}/2-(k-1)T_{\text{sub}}}{T_{\text{sub}}}\right)\text{e}^{\text{j}\pi\gamma\left(\hat{t}-T_{\text{sub}}/2-(k-1)T_{\text{sub}}\right)^2}\text{e}^{\text{j}2\pi a_{n,k}\Delta f\hat{t}} \tag{1}$$

where $n = 1,2,\ldots,N$, $k = 1,2,\ldots,K$, $\text{rect}(\hat{t}/T_{\text{sub}})=\begin{cases}1, & |\hat{t}/T_{\text{sub}}|\le 1/2\\ 0, & \text{others}\end{cases}$ is the window function. $\gamma = B_{\text{sub}}/T_{\text{sub}}$ is the slope of LFM. B_{sub} is sub-pulse bandwidth. T_{sub} is sub-pulse width. $T_{\text{p}} = K\times T_{\text{sub}}$ is pulse width. $\hat{t}$ is fast time. $a_{k,n}$ is the k-th sub-frequency coding, $a_{k,n} \in \{0,1,2,\ldots,K-1\}$. Δf is a sub-frequency interval. The n-th transmitted pulse signal by FAR is represented as (2):

$$s_{\text{T}}\left(\hat{t},t_n\right)=\text{e}^{\text{j}2\pi f_n\left(\hat{t}+t_n\right)}\sum_{k=1}^{K}s_{\text{T\_sub}}(\hat{t},n,k)= \text{e}^{\text{j}2\pi f_n\left(\hat{t}+t_n\right)}\cdot\sum_{k=1}^{K}u\left(\hat{t}-T_{\text{sub}}/2-(k-1)T_{\text{sub}}\right)\text{e}^{\text{j}2\pi a_{n,k}\Delta f\hat{t}} \tag{2}$$

where $t_n = (n-1)T_{\text{r}}$ is slow time, $t=\hat{t}+t_n$. T_{r} is the pulse repetition interval (PRI). f_n is carrier frequency, $f_n = f_0 + a(n)\Delta F$. f_0 is the initial carrier frequency. ΔF is the minimum frequency interval. $a(n)$ is carrier frequency coding, $a(n) \in \{0,1,2,\ldots,M-1\}$. The diagram is shown in Fig. 1.

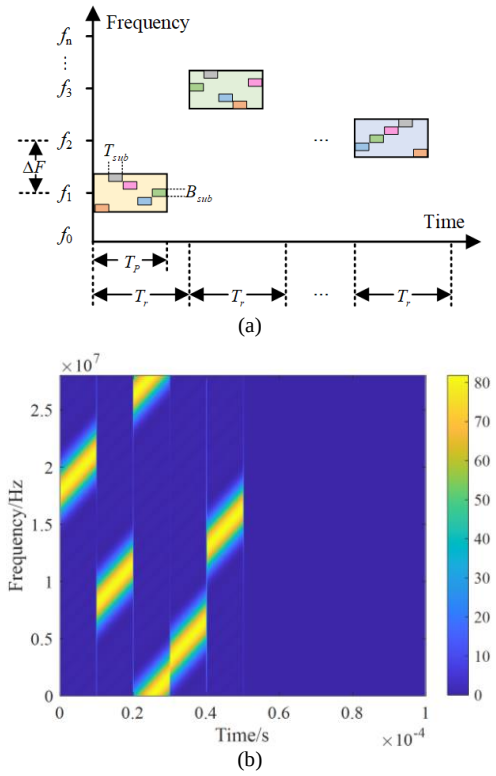

Fig. 1. Time-frequency diagram of the transmitted IPCFH-IPSFC waveform. (a) is the time-frequency diagram of pulses in a CPI. (b) is the time-frequency simulation result of sub-pulses in a PRI.

The FAR target echo in the observation scenario for an aircraft target is $s_{\text{r}}\left(\hat{t},t_n\right)$.

$$s_{\text{r}}\left(\hat{t},t_n\right)=s_{\text{T}}\left(\hat{t}-\tau_n,t_n\right)= \sigma_0\,\text{e}^{\text{j}2\pi f_n\left(\hat{t}+t_n-\tau_n\right)}\cdot\sum_{k=1}^{K}u\left(\hat{t}-T_{\text{sub}}/2-(k-1)T_{\text{sub}}-\tau_n\right)\text{e}^{\text{j}2\pi a_{n,k}\Delta f\left(\hat{t}-\tau_n\right)} \tag{3}$$

where σ_0 is the target backscatter coefficient. $\tau_n = 2(R_0 - v(n-1)T_{\text{r}})/C$ is the target delay. $C = 3\times10^8$ m/s is light speed. $u(\hat{t})=\text{rect}\left(\hat{t}/T_{\text{sub}}\right)\text{e}^{\text{j}\pi\hat{t}^2}$ is a complex envelope function. We assume that there is an aircraft target and it is a scatterer with radial velocity v. R_0 is the initial range of FAR and aircraft target.

2.2 Jamming Strategies for Active Jamming

We established signal models for ISRJ and smart noise jamming, two active jamming modes. The jammer of a time-sharing system with co-located transmitting and receiving antennas alternately samples and repeats intercepted radar signals during the current PRI. Deception jamming after pulse compression can seriously impair target detection. Different repeater methods divide ISRJ

into direct, repeat, and cyclic modes. We focus on direct repeater ISRJ. It samples a short radar signal segment. The modulated signal is quickly forwarded. Assume that $s_{\mathrm{T}}\left(\hat{t},t_n\right)$ is intercepted by the jammer. The directly forwarded sampling is rectangular pulse $p\left(\hat{t},t_n\right)$. Direct-repeater ISRJ can be expressed as (4) [29].

$$J_{\mathrm{ISRJ}}\left(\hat{t},t_n\right)=p\left(\hat{t},t_n\right)s_{\mathrm{T}}\left(\hat{t},t_n\right)= A_{\mathrm{j}}\sum_{z=0}^{Z-1}\mathrm{rect}\left(\frac{\hat{t}-T_{\mathrm{sub}}/2-T_{\mathrm{j}}-cT_{\mathrm{s}}-\tau_{\mathrm{j}}}{T_{\mathrm{j}}}\right)s_{\mathrm{T}}\left(\hat{t},t_n\right) \tag{4}$$

where A_{j} is the jamming amplitude. T_{j} is sample time. T_{s} is sample period, $T_{\mathrm{s}}=2T_{\mathrm{j}}$. $Z=\lceil T_{\mathrm{p}}/T_{\mathrm{s}}\rceil$ is a slice number. $\lceil\rceil$ is the rounding operation. τ_{j} is jamming delay. T_{j} determines the false target's delay from the real target after ISRJ pulse compression [29]. To produce different jamming effects, the jammer can adjust the sampling time to move the false target in the range domain.

Modulating the FAR-transmitted signal that the jammer has intercepted with narrow-band noise produces smart noise jamming [30]. It has convolution and product modulation modes. We primarily focus on smart noise jamming based on product modulation. Smart noise jamming utilizes intelligent algorithms to analyze radar signal characteristics, such as frequency, pulse width, and repetition frequency, in order to generate jamming signals that closely resemble target signal properties, with the aim of suppressing or deceiving the radar system. This type of jamming not only enhances the coherence between the jamming signal and the target echo but also exhibits randomness and non-stationarity, similar to noise-modulated jamming. Additionally, it automatically aligns with the radar's operating frequency, resulting in more concentrated energy.

The smart noise jamming first intercepts the target signal, and then modulates it using a noise sequence. Assuming the jammer samples the FAR-transmitted signal $s_{\mathrm{T}}\left(\hat{t},t_n\right)$ with full pulse storage and the narrow-band noise is $noise\left(\hat{t},t_n\right)$. The specific mathematical representation of the time domain of smart noise jamming is shown in (5)[30]:

$$J_{\mathrm{SNJ}}\left(\hat{t},t_n\right)=\mathrm{rect}\left(\frac{\hat{t}+t_n-\tau_{\mathrm{j}}}{T_{\mathrm{j}}}\right)\cdot \exp\left(\mathrm{j}2\pi\left(\begin{array}{l} f_0\left(\hat{t}+t_n-\tau_{\mathrm{j}}\right)+\\ \frac{\gamma}{2}\left(\hat{t}+t_n-\tau_{\mathrm{j}}\right)^2+\\ K_{\mathrm{FM}}\int_0^{\hat{t}'}noise\left(\hat{t}'\right)\mathrm{d}\hat{t}' \end{array}\right)\right) \tag{5}$$

where K_{FM} is the frequency modulation slope of the smart noise jamming, and $noise(t)$ is the instantaneous frequency noise function, which follows a zero-mean Gaussian stochastic process.

Smart noise jamming based on product modulation is the product output of the radar-transmitted signal and narrow-band noise signal. After applying the Fourier transform to the smart noise jamming, the output of the matched filter is obtained as follows [30]:

$$\begin{aligned} Y\left(t\right)&=\int_{-\infty}^{+\infty}J_{\mathrm{SNJ}}\left(\tau_{\mathrm{F}}\right)h\left(t-\tau_{\mathrm{F}}\right)\mathrm{d}\tau_{\mathrm{F}}\\ &=\int_{-\infty}^{+\infty}J_{\mathrm{SNJ}}\left(\tau_{\mathrm{F}}\right)\cdot J_{\mathrm{SNJ}}^{*}\left(\tau_{\mathrm{F}}-t\right)\mathrm{d}\tau_{\mathrm{F}} \end{aligned} \tag{6}$$

where τ_{F} is the filter delay. $h(t-\tau_{\mathrm{F}})$ is the conjugate time-reversed function, $h(t-\tau_{\mathrm{F}})=J^{*}_{\mathrm{SNJ}}(\tau_{\mathrm{F}}-t)$.

The jammer with smart noise jamming can automatically aim at the radar's working frequency without the frequency measurement system's guidance and concentrate the jamming signal's energy within the radar's working bandwidth. According to smart noise jamming after pulse compression, the range domain range depends on the narrow-band noise signal duration [4]. Therefore, by changing the narrow-band noise signal duration, the jammer can change the range-domain jamming coverage.

The ISRJ samples the sub-pulse signal because its sampling time is smaller than a PRI. We developed jamming strategies I and II, as depicted in Fig. 2. For jamming strategy I, the sampling time of ISRJ is equal to one sub-pulse width, resulting in a sample period of $T_{\mathrm{s}}=2T_{\mathrm{sub}}$. In jamming strategy II, ISRJ sampling time equals two sub-pulse widths, resulting in a sample period of $T_{\mathrm{s}}=4T_{\mathrm{sub}}$.

Smart noise jamming samples signals using full pulse storage. We developed jamming strategies III and IV, as depicted in Fig. 2. In jamming strategy III, smart noise jamming sampling time T_{width} is one PRI, and narrow-band noise signal duration is T_{p}. In jamming strategy IV, smart noise jamming sampling time T_{width} is two PRI, and narrow-band noise signal duration is $2T_{\mathrm{p}}$. Jamming strategies are assumed to remain unchanged in a CPI. In the FAR-jammer game, the jammer can use different jamming strategies in different CPIs, creating a dynamic environment with multiple jamming strategies.

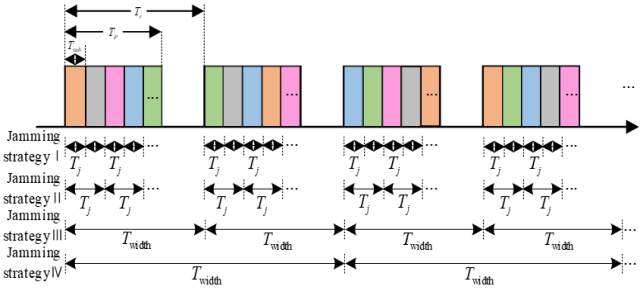

Fig. 2. Schematic diagram of the jamming strategies.

3. Intelligent Learning Method Design Based on PPO-GPT Algorithm

We consider an electronic countermeasure scenario with a FAR and an aircraft target, and the aircraft target is a scatterer with radial velocity v. The initial range of FAR and aircraft target is R_0. Airborne aircrafts use jammers with multiple strategies to interfere with FAR. FAR must continuously learn jamming strategies, adaptively optimize transmitted waveform parameters, and prevent jamming to accurately estimate target information. The FAR intelligent learning model is designed in Fig. 3.

The FAR sends IPCFH-IPSFC waveforms to the environment. The intelligent anti-jamming decision-making system receives echo signals and processes them into baseband signals. The intelligent anti-jamming decision-making system analyzes baseband signals, evaluates anti-jamming performance, and detects targets. Results of anti-jamming evaluation and target detection are sent to the learning algorithm module so that the best IPCFH-IPSFC waveform parameters can be made. Finally, the optimized IPCFH-IPSFC waveform is sent outside. The intelligent learning model includes an electronic countermeasure environment, radar receiver, intelligent anti-jamming decision-making system, and radar transmitter. We designed the FAR intelligent anti-jamming decision-making system. Five sub-modules make up the FAR intelligent anti-jamming decision-making system: signal processing module, anti-jamming evaluation module, target detection module, optimization constraint design module, and learning algorithm design module. We will provide a detailed introduction to each sub-module.

3.1 Signal Processing Module Based on Segmented Pulse Compression

IPCFH-IPSFC waveforms have dynamic parameters in the time-frequency domain, and FAR transmits signals with different sub-frequency coding, so we must distinguish sub-pulses by sub-frequency coding. The active jamming's time domain discontinuity allows it to suppress jamming signals by sorting out echo signals not sampled by the jammer. The signal processing method uses segmented pulse compression. To obtain sub-pulse signals at various sub-frequencies, a bandpass filter (BPF) processes the baseband signal in the frequency domain. The n-th pulse is fed into BPF for FFT with a bandwidth of $B_{\mathrm{BPF}}(n,k)$ to obtain the frequency domain signal, where $a_{n,k}\Delta f - B_{\mathrm{sub}}/2 \le B_{\mathrm{BPF}}(n,m) \le a_{n,k}\Delta f + B_{\mathrm{sub}}/2$. Inverse fast Fourier transform (IFFT) is used to obtain the time-domain signal. After obtaining sub-pulse signals for different sub-frequency encodings, we use a parallel processing structure to compress each sub-pulse. The sub-pulse compressed signal $y\left(\hat{t},t_n,k\right)$ is formulated as (7):

$$y\left(\hat{t},t_n,k\right)=s_{\mathrm{R\_sub}}\left(\hat{t},n,k\right)\otimes s^{*}_{\mathrm{T\_sub}}\left(-\hat{t},n,k\right) \tag{7}$$

where $s_{\mathrm{R\_sub}}\left(\hat{t},n,k\right)=s_{\mathrm{r}}\left(\hat{t},n,k\right)+J\left(\hat{t},n,k\right)+n_2\left(\hat{t},n,k\right)$, $n_2\left(\hat{t},n,k\right)$ is the noise signal.

3.2 Anti-jamming Evaluation Module Based on Adaptive Variance Threshold

After obtaining the sub-pulse compressed signal, we compare the energy variance of the jamming and the target to determine if the pulse signal contains jamming. The SINR measures whether the intelligent anti-jamming decision-making system chose the best strategy in the previous confrontation. The active jamming has time-domain discontinuities. Some sub-pulses have jamming after compression, while others have only noise and target. A much higher jamming energy than the target energy will distinguish jamming. After segmented pulse compression, jammed and unjammed sub-pulses have different amplitude fluctuation characteristics. Variance can reflect the amplitude fluctuation characteristics, and the jammed sub-pulse has a higher variance than the unjammed one. $\mathrm{var}\left(y\left(\hat{t},t_n,k\right)\right)$ is a symbol for the amplitude variance. The adaptive variance threshold is defined as $\Lambda_0=\frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{K}\sum_{k=1}^{K}\mathrm{var}\left(y\left(\hat{t},t_n,k\right)\right)\right)$ to classify the signals. Therefore, the jamming recognition result is expressed as (8)

$$\text{Jamming recognition}\begin{cases}H_0:\mathrm{var}\left(y\left(\hat{t},t_n,k\right)\right)\le\rho\times\Lambda_0,\text{ target signal}\\ H_1:\mathrm{var}\left(y\left(\hat{t},t_n,k\right)\right)>\rho\times\Lambda_0,\text{ jamming signal}\end{cases} \tag{8}$$

where ρ is the scaling factor, which is selected with respect to the amplitude distribution of the target, jamming, and noise. It can be set empirically, or the maximum output SINR criterion can be used.

Equation (8) determines if all sub-pulses after pulse compression jam. Sub-pulses with a variance less than the threshold $\rho\times\Lambda_0$ are target signals, and those with a variance greater than the threshold $\rho\times\Lambda_0$ are jamming signals. Thus, to obtain the pure sub-pulse $y'\left(\hat{t},t_n,k\right)$, we set the jammed sub-pulse amplitude to 0. After jamming suppression, all sub-pulses within a PRI are accumulated to obtain signal $y''\left(\hat{t},t_n\right)$.

$$y'\left(\hat{t},t_n,k\right)=\begin{cases}0 & \mathrm{var}\left(n,k\right)\ge g_{\eta}^{\ *},\\ y\left(\hat{t},t_n,k\right) & \mathrm{var}\left(n,k\right)<g_{\eta}^{\ *},\end{cases} \tag{9}$$

$$y''\left(\hat{t},t_n\right)=\sum_{k=1}^{K}y'\left(\hat{t},t_n,k\right)=\sum_{k=1}^{K}A_{n,m}\,\mathrm{e}^{-\mathrm{j}2\pi f_n\frac{2\tau_n}{C}}+n_3\left(\hat{t},t_n\right) \tag{10}$$

where $A_{n,k}$ is the amplitude of the target signal, and $n_3\left(\hat{t},t_n\right)$ is noise.

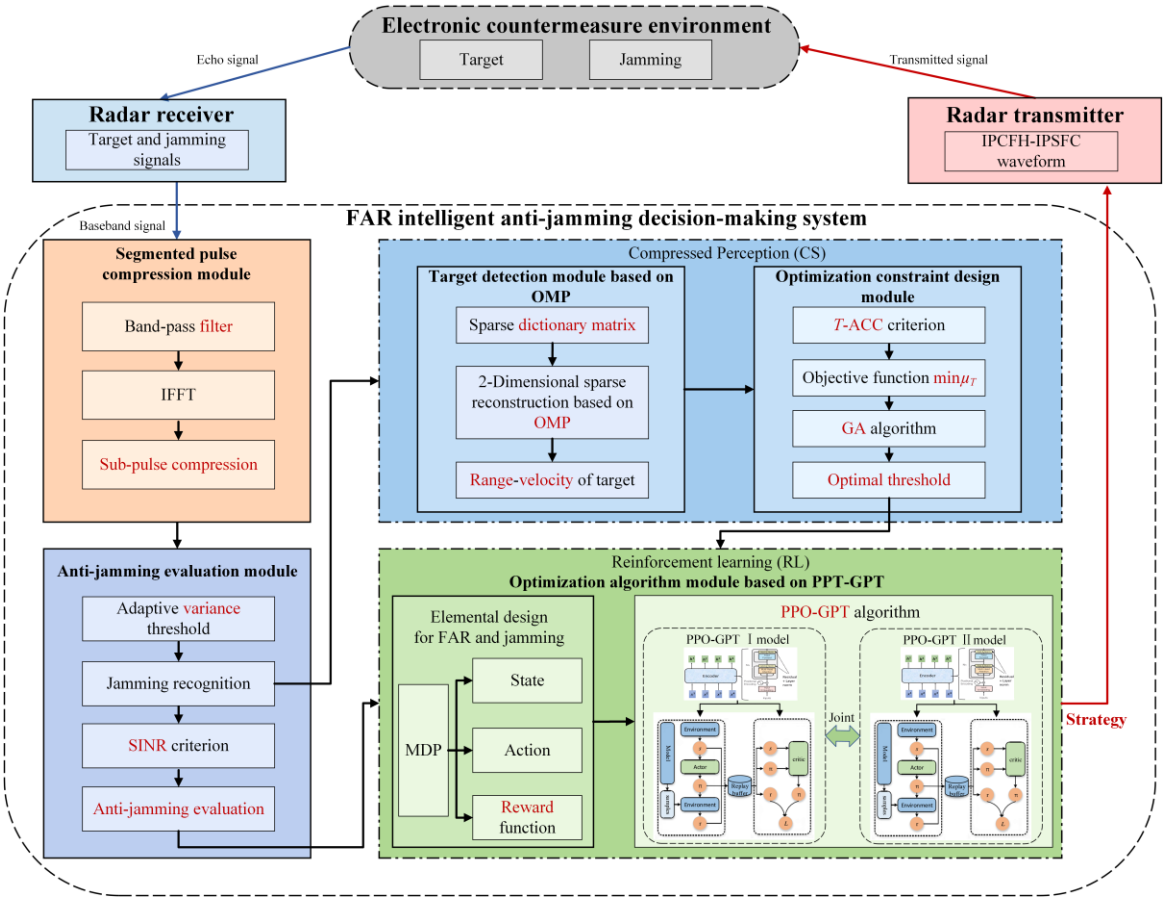

Fig. 3. FAR intelligent learning model.

We figure out the sub-pulse output SNR in the time domain based on the results of jamming recognition to see how well the previous confrontation worked at stopping jamming. This is defined as SNR_0:

$$\mathrm{SNR}_0 = \frac{\chi_\mathrm{s}}{\sigma^2_{\mathrm{noise}}} \tag{11}$$

where $\chi_\mathrm{s} = \sum \left| y'\left(\hat{t}, t_n, k\right)\right|^2$ is the total energy of the sub-pulse. $\sigma^2_{\mathrm{noise}}$ is the total energy of the noise.

According to (9), the jamming signal's amplitude is 0, so its SNR cannot be calculated. Sub-pulse recognition as a jamming signal indicates that anti-jamming failed in the previous confrontation. Thus, we use signal-to-jamming ratio (SJR) from expert experience to evaluate the sub-pulse containing the jamming signal in the time domain. The anti-jamming evaluation result is:

$$E_{\text{anti-jamming}} = \begin{cases} \mathrm{SJR}_0, & y'\left(\hat{t}, t_n, k\right) = 0, \\ \mathrm{SNR}_0, & y'\left(\hat{t}, t_n, k\right) = y\left(\hat{t}, t_n, k\right). \end{cases} \tag{12}$$

3.3 Target Detection Module Based on OMP

There are discontinuous carrier frequencies in a CPI because the IPCFH-IPSFC waveform changes dynamically in time-frequency. Segmented pulse compression yields the intra-pulse accumulation signal $y''\left(\hat{t}, t_n\right)$ within a PRI, leaving only target signals in the FAR observation scene. Targets are sparse within one range unit. Therefore, we use the 2-dimensional sparse reconstruction method based on CS theory to calculate target range and velocity [31]. A $L \times W$-unit 2-dimensional range and velocity plane represent the radar observation scene. There are L units in the range dimension and W units in the velocity dimension. We build a sparse dictionary matrix as $\boldsymbol{\Psi}$ [31].

$$\boldsymbol{\Psi} = \left\{ \underbrace{e_{1,1} \cdots e_{1,L}}_{L} \cdots \quad \cdots \underbrace{e_{W,1} \cdots e_{W,L}}_{L} \right\}_{N \times (L \times W)} \tag{13}$$

where $e_{l,w} = \varphi_l(n) \cdot \varphi_l(n)$, $n = 1,2,\ldots,N$. $\varphi_l(n)$ is the range phase, $\varphi_l(n) = \exp(-\mathrm{j}4\pi a_n \Delta F\, r_l / C)$. $1 \le l \le L$. $\varphi_w(n)$ is the velocity phase, $\varphi_l(n) = \exp[-\mathrm{j}4\pi(f_0 + a_n \Delta F)\, v_w n T_\mathrm{r} / C]$. $1 \le w \le W$. $\xi_{l,w}$ is the target backscatter coefficient, $\xi_{l,w} = A_{l,mw} \exp(-\mathrm{j}4\pi f_0\, r_l / C)$. The echo signal of the q-range unit is $\mathbf{y}'_q$ [31].

$$\mathbf{y}'_q = \boldsymbol{\Psi}\boldsymbol{\theta}_q + \boldsymbol{\delta}_q, \quad q = 1, 2, \cdots, Q \tag{14}$$

where Q is the range unit number. $\boldsymbol{\theta}_q$ is the sparse reconstruction vector. $\boldsymbol{\delta}_q$ is the noise vector.

Targets typically occupy a small portion of range-velocity coordinates in radar detection. Thus, the range-

velocity domain echo signal $y''\left(\hat{t},t_n\right)$ is sparse. Because there aren't many targets in the observation scene, it is possible to figure out the unknown vector $\boldsymbol{\theta}_q$ from the signal that was seen by solving a ℓ_1-paradigm optimization problem [31].

$$\left\langle \hat{\boldsymbol{\theta}}_q \right\rangle = \arg\min\left(\left\|\boldsymbol{\theta}_q\right\|_1\right), \text{subject to } \left\|\mathbf{y}'_q - \boldsymbol{\Psi}\boldsymbol{\theta}_q\right\|_2 \le \varepsilon \quad (15)$$

where $\hat{\boldsymbol{\theta}}_q$ is an estimate of the target amplitude in the q-th range cell. ε is the noise amplitude. The noise $\varepsilon = \left\|\boldsymbol{\delta}_q\right\|_2$ can be estimated from neighboring range or velocity cells. The $\boldsymbol{\theta}_q$ is obtained by solving (15), and the target's parameters are based on $\boldsymbol{\theta}_q$'s peak position. We use the OMP algorithm to reduce computational complexity [32].

3.4 Optimization Constraint Design Module Based on T-ACC

OMP-based target detection uses the observation scene and target signal to determine the observed signal, but the radar designer designs the dictionary matrix $\boldsymbol{\Psi}$. Equation (14) states that the dictionary matrix is deterministic and must be built beforehand. However, CS theory states that dictionary matrix column orthogonality directly affects signal stability and reconstruction accuracy. Thus, we investigated the dictionary matrix and IPCFH-IPSFC waveform agile parameters. Next, the IPCFH-IPSFC waveform parameter optimization constraint is set to ensure strong orthogonality between dictionary matrix columns to improve target recovery accuracy. The dictionary matrix contains range value intervals, velocity value intervals, carrier frequency coding, and PRF, according to (13). In practice, range and velocity value intervals are fixed. Our designed IPCFH-IPSFC waveform fixes the PRF. The carrier frequency coding of the IPCFH-IPSFC waveform determines the dictionary matrix construction in this study. Because carrier frequency coding is dynamically adjusted, the dictionary matrix differs between CPIs. Thus, optimizing the carrier frequency coding $a(n)$ must consider the dictionary matrix's performance and sparse reconstruction accuracy. We designed the T-average coherence coefficient (T-ACC) as an evaluation criterion of the dictionary matrix correlation. The correlation coefficient $\mu_T(\boldsymbol{\Psi})$ of the dictionary matrix is defined as (16):

$$\mu_T(\boldsymbol{\Psi}) \triangleq \frac{\sum_{1\le i,\, j\le N, i\ne j}(|\, G(i,\, j)\,| \ge T)\,|\, G(i,\, j)\,|}{\sum_{1\le i,\, j\le N, i\ne j}(|\, G(i,\, j)\,| \ge T)} \quad (16)$$

where $G(i,j)$ is the element in row i and column j in the Gram matrix $\mathbf{G} = \boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Psi}$. $\mu_T(\boldsymbol{\Psi})$ is the average of all Gram matrix G non-diagonal elements with absolute values greater than a threshold T.

Equations (13) and (16) express the dictionary matrix correlation coefficient as a function of carrier frequency coding $a(n)$. A smaller dictionary matrix $\mu_T(\boldsymbol{\Psi})$ means stronger orthogonality between columns and higher accuracy in sparsely reconstructing the target signal. Thus, the dictionary matrix column correlation coefficient optimization objective function is min $\mu_T(\boldsymbol{\Psi})$:

$$\min \mu_T(\boldsymbol{\Psi}) \triangleq \min \frac{\sum_{1\le i,\, j\le N, i\ne j}(|\, G(i,\, j)\,| \ge T)\,|\, G(i,\, j)\,|}{\sum_{1\le i,\, j\le N, i\ne j}(|\, G(i,\, j)\,| \ge T)}. \quad (17)$$

We optimize the carrier frequency coding $a(n)$ to suppress active jamming. Parameter variations affect dictionary matrix construction and target estimation accuracy. Therefore, we must find the optimal dictionary matrix to meet the FAR requirements. According to FAR parameter settings, we used a genetic algorithm (GA) to solve (17) by taking all values of $a(n)$ to get the range of $\mu_T(\boldsymbol{\Psi})$ [33].The optimal and mean value curves are shown in Fig. 4. The optimal value is the best correlation coefficient up to the current iteration, and the mean value is the average of all correlation coefficients.

Figure 4(a) shows the GA target fitness parameter Fit as 0 to find the minimum value of $\mu_T(\boldsymbol{\Psi})$ by taking all values of $a(n)$. Figure 4(b) shows the GA target fitness parameter Fit as 1 to find the maximum value of $\mu_T(\boldsymbol{\Psi})$ by taking all values $a(n)$. The GA algorithm converges after 30 iterations and reaches the global optimal solution in 50 iterations. The minimum and maximum value optimization

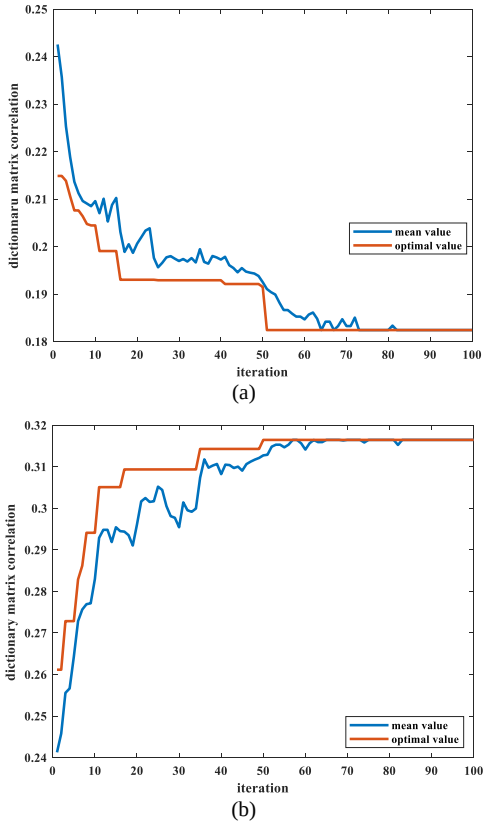

Fig. 4. $\mu_T(\boldsymbol{\Psi})$ optimization curve based on GA. (a) is the dictionary matrix correlation coefficients when the fitness parameter Fit = 0. (b) is the dictionary matrix correlation coefficients when the fitness parameter Fit = 1.

results show that the range of $\mu_T(\mathbf{\Psi})$ is (0.19, 0.32). No matter how the waveform parameters are changed, the $\mu_T(\mathbf{\Psi})$ falls within the (0.19, 0.32) range when there is no jamming signal in the observation scene. Experimental results show that different carrier frequency coding can have the same $\mu_T(\mathbf{\Psi})$. Multiple $\mu_T(\mathbf{\Psi})$ values meet parameter estimation accuracy requirements. In the observation scene, we must determine if the jamming signal affects target sparse reconstruction success. According to FAR requirements on parameter estimation accuracy and target sparse reconstruction success rate, the OMP algorithm is used to detect jammed signals. The Monte Carlo method generates large number of experimental samples to better observe experimental results [34]. We investigate how to choose the optimal threshold $\mu_T^{\text{opt}}(\mathbf{\Psi})$ in the interval (0.19, 0.32) to optimize the IPCFH-IPSFC waveform parameter. A simulation experiment will be introduced in Sec. 4.1.

3.5 Learning Algorithm Design Module Based on PPO-GPT

The FAR intelligent learning model shows that FAR transmits the IPCFH-IPSFC waveform to interact with the environment. The jammer intercepts the FAR-transmitted signal at time t and modulates the jamming signal according to the jamming strategy. The FAR creates anti-jamming actions by improving the IPCFH-IPSFC waveform parameters based on the results of the anti-jamming evaluation and the optimization constraint from the previous time t – 1. This is done after receiving the combined signal of the jamming and target. By RL theory, this procedure is analogous to the RL model. Markov decision processes (MDP) are used to model the intelligent learning method of agile waveforms against jamming. Consider the FAR as the agent and the dynamically variable jamming scenario as the environment. The MDP can be represented as a tuple $M = (S,A,P,R,\gamma)$, which includes state space S, action space A, state transition probability P, reward function R, and discount factor γ.

FAR can perceive the jamming state $s_t = [s_{t,f}, s_{t,T_j}]$ at time t, where $s_t \in S$. $s_{t,f}$ is the carrier frequency of the jamming signal. s_{t,T_j} is the jamming signal duration. The action generated by the FAR at time t is $a_t = [a_{t,a(n)}, a_{t,a_{n,k}}]$, where $a_t \in A$. $a_{t,a(n)}$ is the carrier frequency coding. $a_{t,a_{n,k}}$ is the sub-frequency coding. The $a_t = [a_{t,a(n)}, a_{t,a_{n,k}}]$ will uniquely determine the FAR-transmitted IPCFH-IPSFC waveform at time t. The waveform parameter increases the action space dimension, making traditional learning algorithms ineffective. Therefore, we divide $a_t = [a_{t,a(n)}, a_{t,a_{n,k}}]$ into $a_t^{\text{up}} = [a_{t,a(n)}]$ and $a_t^{\text{down}} = [a_{t,a_{n,k}}]$. We design the reward function based on the anti-jamming evaluation results and the optimization constraint. We designed a reward function R^{up} to optimize $a_t^{\text{up}} = [a_{t,a(n)}]$ and a reward function R^{down} to optimize $a_t^{\text{down}} = [a_{t,a_{n,k}}]$.

$$R^{\text{up}} = \begin{cases} 1, & t = N \text{ and } \mu_T(\mathbf{\Psi}) \le \mu_T^{\text{opt}}(\mathbf{\Psi}) \\ 0, & \text{others} \end{cases}, \tag{18}$$

$$R^{\text{down}} = \bar{E}_{\text{anti-jamming}} \tag{19}$$

where $\bar{E}_{\text{anti-jamming}}$ is the normalization value $E_{\text{anti-jamming}}$, which is $\bar{E}_{\text{anti-jamming}} \in [0,1]$.

The R^{up} is trained first, and then the R^{down} is trained. Finally, they are jointly optimized. The reward function R^{joint} for joint learning is defined as (20):

$$\begin{aligned} R^{\text{joint}} &= R^{\text{up}} + R^{\text{down}} \\ &= \begin{cases} 1 + \dfrac{1}{K}\sum_{k=1}^{K} \bar{E}_{\text{anti-jamming}}, & t = N \text{ and } \mu_T(\mathbf{\Psi}) \le \mu_T^{\text{opt}}(\mathbf{\Psi}), \\ \dfrac{1}{K}\sum_{k=1}^{K} \bar{E}_{\text{anti-jamming}}, & t = N \text{ and } \mu_T(\mathbf{\Psi}) > \mu_T^{\text{opt}}(\mathbf{\Psi}), \\ \bar{E}_{\text{anti-jamming}}, & t \ne N. \end{cases} \end{aligned} \tag{20}$$

In MDP, the objective of the agent is to maximize the expected future reward, i.e., the optimal action-value function $Q^*(s_t, a_t)$. The $Q^*(s_t, a_t)$ can be obtained by solving the Bellman equation [35]:

$$Q^*(s_t, a_t) = E\left[R(s_t, a_t) + \gamma Q_\pi(s_{t+1}, a_{t+1})\right]. \tag{21}$$

Order to obtain the action-value function $Q^*(s_t, a_t)$, the state transition probability P needs to be known. Due to the uncertainty of the jamming states, the state transition probabilities P cannot be obtained. By continuously updating the Q-value, the deep reinforcement learning (DRL) algorithm finds the optimal action value without knowing the state transition probability P. Due to a lack of training data, computational power, and instability, traditional DRL algorithms are only suitable for smaller state space decision-making problems. Complex environments create a larger state and action space in the FAR-jammer game. Neural networks have many parameters, so traditional DRL algorithms cannot complete the task. As an actor-critic (AC) benchmark algorithm, the PPO algorithm can solve the above problems [35]. Some neural networks have issues like low parallel computation efficiency and poor feature capture over long distances during training. To complete state space preprocessing, we introduce the GPT algorithm based on PPO to improve training effectiveness and reduce training difficulty. Therefore, we propose a dual PPO-GPT algorithm to realize the game of FAR with active jamming. The structure of the dual PPO-GPT algorithm is shown in Fig. 5. The dual PPO-GPT model mainly consists of PPO-GPT I and PPO-GPT II. They have the same structure and state space, but different action spaces and reward functions.

The action of PPO-GPT I is $a_t^{\text{up}} = [a_{t,a(n)}]$, and the reward function is R^{up}. The action of PPO-GPT II is $a_t^{\text{down}} = [a_{t,a_{n,k}}]$, and the reward function is R^{down}. PPO-GPT I

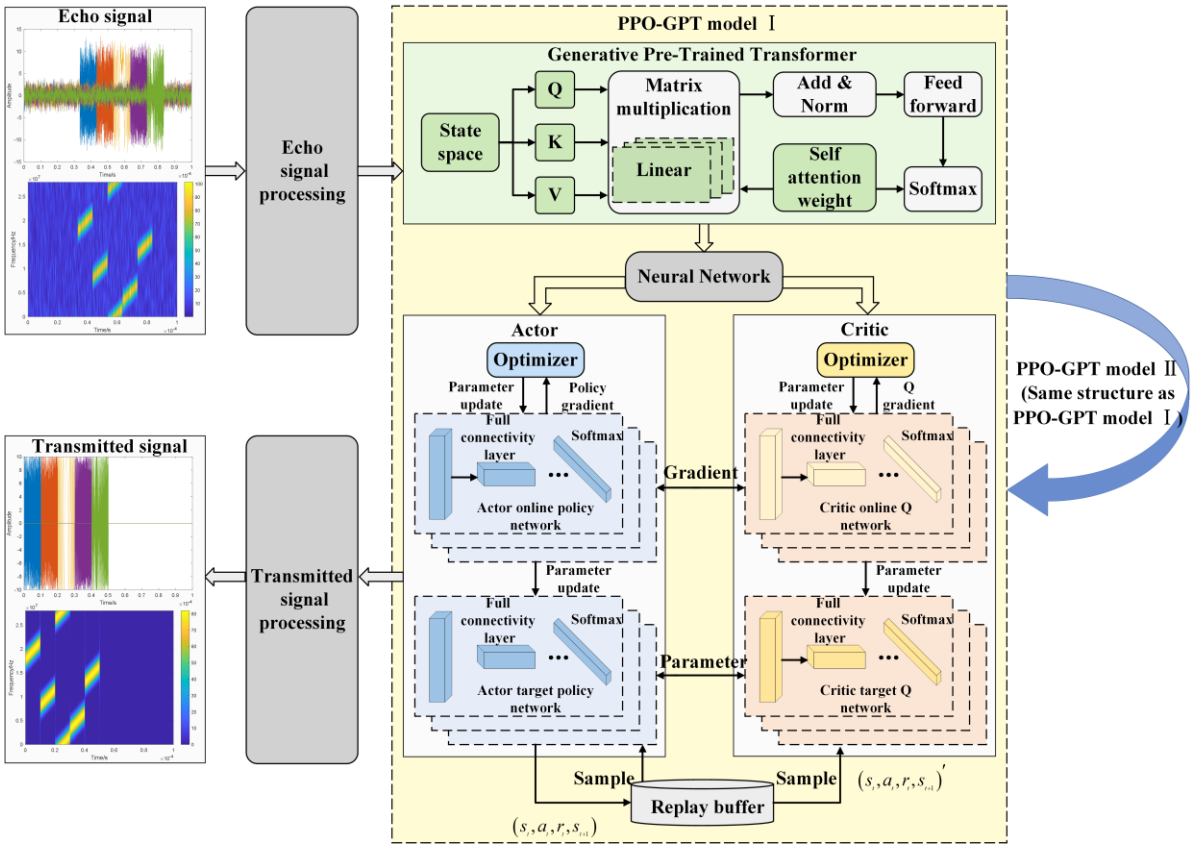

Fig. 5. Dual PPO-GPT algorithm architecture diagram.

is trained first, then PPO-GPT II is trained and finally jointly optimized. The reward function for joint learning is R^{joint}. The PPO-GPT model uses GPT to analyze state information and determine the impact on current decisions over time. State information multiplies three weight vectors: query, key, and value. It helps the neural network focus on the most important state-space regions. The PPO has actor and critic networks. The critic network calculates the state advantage function A_t^{θ}. The actor network updates policy function network parameters θ based on A_t^{θ}. The actor network stores sampled trace $\tau_i = (s_t, a_t, r_t, s_{t+1})$ from decisions. The actor network interacts with the environment to get the reward. Based on the reward and state, the critic network adjusts its weights. Algorithm 1 displays the PPO-GPT I algorithm, which is identical to the PPO-GPT II algorithm.

4. Simulation Results

We used Monte Carlo experiments to examine how jamming's JSR and sparse recovery success rate affect dictionary matrix correlation. To confine the learning procedure, the optimal threshold $\mu_T^{\text{opt}}(\boldsymbol{\Psi})$ is chosen. Therefore, Section 4.1 is to determine the optimal threshold through Monte Carlo simulation in order to provide constraints for the verification of RL-based intelligent learning methods.

We next compare the PPO-GPT algorithm to the PPO-LSTM algorithm [22], the dual DQN (DDQN) algorithm [36], the advantage actor-critic (A2C) algorithm [37], and the random method [38] using our intelligent learning model in Sec. 4.2, 4.3, 4.4, and 4.5, respectively. Performance testing of PPO-GPT-based intelligent learning algorithms includes checking for robustness and convergence, target detection, anti-jamming, and the ability to make accurate decisions in a setting with multiple jamming strategies that are used in order.

The simulation results presented in Sec. 4.2, 4.3, and 4.4 are all from the online execution of the algorithm, while the simulation results in Sec. 4.5 are from the offline execution of the algorithm. The simulated hardware environment is as follows: GPU: GeForce RTX 2060. Intel Core i7-9750H. RAM: 16G. The operating system is Windows 10. The software is Python 3.

Table 1 lists FAR simulation parameters. The specific parameters are defined in Sec. 2.1. We set the four jamming strategies designed in Sec. 2.3 as the four jamming environments for FAR, where the jammer generates jamming data according to these strategies to engage in electronic warfare against FAR. Following the PPO-GPT algorithm steps outlined in Sec. 3.5 and the four jamming environments described in Sec. 2.3, we configure the experimental parameters of the algorithm as follows: the greed factor is set to 0.95, the learning rate to 0.0003, and the discount factor γ to 0.99. The size of memory replay buffer D_{round} for the PPO-GPT algorithm is set to 100,000, with 512 training samples and 64 test samples.

Algorithm 1: PPO-GPT algorithm.

Input: State, confrontation termination time $episode_{cut}$.
Output: Action.
Initialization: Initialize network parameter θ.
For *episode* = 1 to $episode_{cut}$ **do**:
Collect the running policies $\pi(round) = \pi(\theta_{round})$ in the environment and get a collection $D_{round} = \{\tau_i\}$ of the trace.
Calculate the weights W_α of the GPT.
Get the reward R_t of the time t.
Calculate the advantage function based on the V_J obtained from the actor network.
Update the network by maximizing the objective function (using the Adam optimizer):

$$\theta_{round+1} = \arg\max_{\theta} \sum_{\tau \in D} \sum_{t=0}^{Time} \min\left(\frac{\pi_\theta(a_t, s_t)}{\pi_{\theta_{round}}(a_t, s_t)} A^{\pi_{\theta_{round}}(a_t, s_t)}, cl \right),$$

$$cl = clip\left(\frac{\pi_\theta(a_t, s_t)}{\pi_{\theta_{round}}(a_t, s_t)}, 1-\varepsilon, 1+\varepsilon \right) A^{\pi_{\theta_{round}}(a_t, s_t)}$$

Use the root mean square to update the parameters of the actor network:

$$J_{round+1} = \arg\min_{J} \frac{1}{|D_{round}| Time} \sum_{\tau \in D} \sum_{t=0}^{Time} \left(V_J(s_t) - R_t \right)^2$$

episode = *episode* + 1.
End for

Parameter	Value	Parameter	Value
N	16	f_0	14 GHz
M	24	ΔF	80 MHz
K	8	B_{sub}	5 MHz
T_{sub}	10 μs	Δf	7 MHz
T_r	100 μs	R_0	11 km
T_p	80 μs	v	20 m/s

Tab. 1. FAR simulation parameters.

During the algorithm's training process, we set *round* = 1000 adversarial rounds for each jamming environment, with a maximum training time of 1,000. The robustness of the algorithm was observed, and the results are presented in Sec. 4.2. In the testing phase, we designed *round* = 1000 rounds of adversarial experiments for each jamming environment, with each round involving 2,048 anti-jamming decisions, resulting in a total of 2,048,000 adversarial episodes. The convergence and anti-jamming performance of the algorithm were observed, and the results are detailed in Sec. 4.2 and 4.3. Each adversarial episode corresponds to the duration of one CPI, i.e., 1 episode = 1 CPI.

4.1 Monte Carlo Simulations to Determine the Optimal Threshold $\mu_T^{opt}(\Psi)$

Section 3.5 shows that $\mu_T(\Psi)$'s ideal range is (0.19, 0.32). To find the best threshold $\mu_T^{opt}(\Psi)$, we ran Monte Carlo trials. The OMP algorithm detects target information in jammed scenes. If the estimation error is within ±0.05, the target sparse reconstruction is successful. Once each 100 target detection simulation tests, the success rate is computed. The value interval is consistently split into five intervals: [0.20, 0.22), [0.22, 0.24), [0.24, 0.26), [0.26, 0.28), and [0.28, 0.30). Each interval has 40 consistent goal values that correspond to 40 IPCFH-IPSFC waveforms with varied parameters.

The jamming strategy III is utilized to intercept the FAR-transmitted waveform. To achieve Monte Carlo experiment unpredictability, we suppose the FAR creates transmitted waveforms using random parameter agility. The success rate of target sparse reconstruction is obtained when the JSR varies in [42, 50] dB. The mean success rate $\bar{H}$ was calculated using 5,000 Monte Carlo trials. The first experiment jams a quarter pulse but not three quarters pulses inside one CPI. The JSR, $\bar{H}$, and $\mu_T(\Psi)$ relationships are given in Fig. 6. The second experiment jams half of the pulses but not the other half of the pulses inside one CPI. We found the JSR, $\bar{H}$, and $\mu_T(\Psi)$ relationships in Fig. 7.

The cubic polynomial fitting curve of $\bar{H}$ and JSR for varying $\mu_T(\Psi)$ in the first experiment is shown in Figure 6(a). As JSR rises, $\bar{H}$ falls. When JSR is 50 dB and $\mu_T(\Psi) \in [0.28, 0.30)$, $\bar{H} \approx 43\%$. $\bar{H}$-$\mu_T(\Psi)$ curves for different JSRs are shown in Fig. 6(b). The higher the $\mu_T(\Psi)$, the lower the $\bar{H}$. When $\mu_T(\Psi) \in [0.28, 0.30)$ and JSR = 42 dB, $\bar{H} \approx 94\%$. Figures 6(c) and (d) show the link between $\mu_T(\Psi)$, JSR, and $\bar{H}$. As JSR rises, $\mu_T(\Psi)$ rises and $\bar{H}$ falls. To have a success rate above 90% for target sparse detection, the JSR must be below 48 dB, or $\mu_T(\Psi) \le 0.26$.

In Fig. 7(a), when JSR is 50 dB, $\mu_T(\Psi) \in [0.28, 0.30)$, $\bar{H} \approx 30\%$. In Fig. 7 (b), when $\mu_T(\Psi) \in [0.28, 0.30)$, JSR = 42 dB, $\bar{H} \approx 93\%$. In Figs. 7(c) and (d), to have a success rate above 90% for target sparse detection, the JSR must be below 46 dB, or $\mu_T(\Psi) \le 0.24$. In conclusion, random parameters cannot be used to generate carrier frequency coding of the transmitted waveform to assure target detection under strong JSR jamming. Thus, carrier frequency coding must be optimized so the jammer cannot

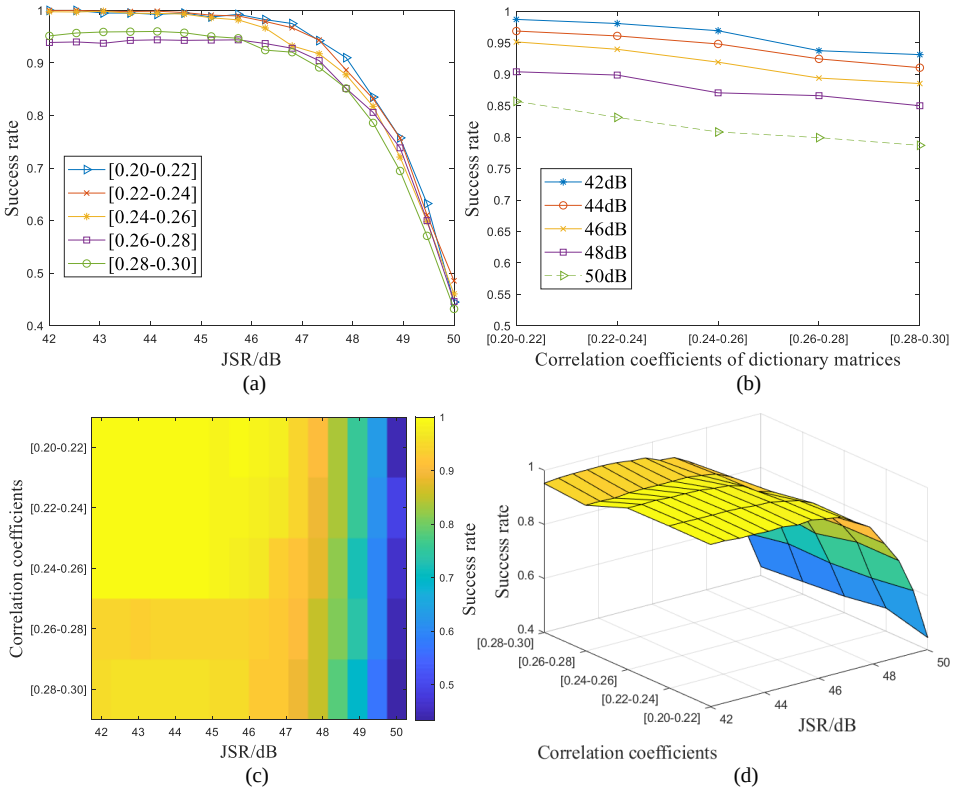

Fig. 6. Results of the relationships between the JSR, $\bar{H}$ and $\mu_T(\mathbf{\Psi})$ in the first experiment. (a) is the cubic polynomial fitting curve of JSR-$\bar{H}$. (b) is $\bar{H}$-$\mu_T(\mathbf{\Psi})$ curves. (c) is the two-dimensional connection between $\mu_T(\mathbf{\Psi})$, JSR, and $\bar{H}$. (d) is the three-dimensional connection between $\mu_T(\mathbf{\Psi})$, JSR, and $\bar{H}$.

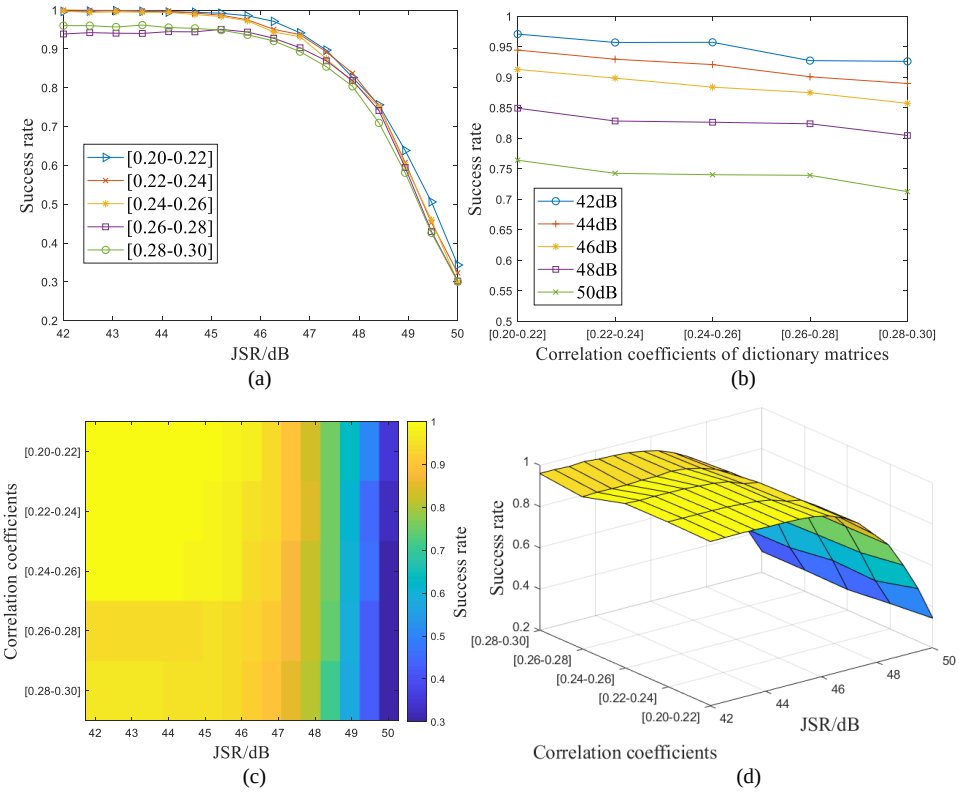

Fig. 7. Results of the relationships between the JSR, $\bar{H}$ and $\mu_T(\mathbf{\Psi})$ in the second experiment. (a) is the cubic polynomial fitting curve of JSR-$\bar{H}$. (b) is $\bar{H}$- $\mu_T(\mathbf{\Psi})$ curves. (c) is the two-dimensional connection between $\mu_T(\mathbf{\Psi})$, JSR, and $\bar{H}$. (d) is the three-dimensional connection between $\mu_T(\mathbf{\Psi})$, JSR, and $\bar{H}$.

instantaneously acquire the anti-jamming strategy. For the agile waveform to achieve successful target sparse reconstruction, $\mu_T^{opt}(\boldsymbol{\Psi}) \le 0.24$ must be made. Order to optimize the agile waveform, we determine the optimal threshold $\mu_T^{opt}(\boldsymbol{\Psi}) = 0.24$

4.2 Robustness and Convergence Verification of Intelligent Learning Method

We test the robustness and convergence of the PPO-GPT learning algorithm using the optimal threshold $\mu_T^{opt}(\boldsymbol{\Psi})$. The four jamming strategies in Sec. 2.3 were used to set the environment for four experiments. We prove our PPO-GPT algorithm's superiority over the three DRL algorithms and the random algorithm. In Experiment 1, the jammer uses strategy I and $SRJ_0 = -42$ dB. In Experiment 2, the jammer uses strategy II and $SRJ_0 = -42$ dB. In Experiment 3, the jammer uses strategy III and $SRJ_0 = -60$ dB. In Experiment 4, the jammer uses strategy IV and $SRJ_0 = -60$ dB. The parameters were uniformly set in four experiments. The unjammed echo signal power is –24 dB in the time domain. The unjammed echo signal SNR_0 is 15 dB. We assume that the noise power of the radar receiver for FAR is 0.

The DRL algorithm updates neural network learning parameters using the loss function. If the loss function gradually converges to 0, the neural network model is more stable, proving the DRL algorithm is more robust. The AC framework is not used to build DDQN, which has one neural network model. The other three algorithms use actor and critic neural networks. Since the random algorithm has no neural network model, we do not test its robustness. In Experiments 1 and 2, the jammer picks up and changes IPCFH-IPSFC sub-pulse signals, so changing the inter-pulse carrier frequency doesn't stop the jamming. The analysis of sub-frequency coding algorithm learning should be the focus. Therefore, we focus on the robustness of the PPO-GPT II model with R^{down} as the reward function. The PPO-GPT II neural network model loss curve with training times is shown in Fig. 8.

As training times increased from 0 to 1000, neural network model loss values changed. All actor neural network models have negative loss values that increase to 0. All other neural network models have positive loss values that decrease to 0. Figure 8(a) shows that after 169 training times, the actor neural network and critic neural network loss values of the PPO-GPT algorithm converge to 0. Other algorithms' neural network loss values converge to 0 at a slower rate. As shown in Fig. 8(b), the actor neural network and critical neural network loss values of the PPO-GPT algorithm reach 0 after 215 training times. The other algorithms' neural network loss value oscillates more and is harder to converge to 0.

In Experiments 3 and 4, the jammer intercepts pulse signals against the IPCFH-IPSFC waveform, so sub-frequency coding does not prevent jamming. The analysis of carrier frequency coding learning should be the focus. Therefore, we focus on the robustness of the PPO-GPT I model with R^{up} as the reward function. The loss curve of PPO-GPT I with training times is shown in Fig. 9.

All neural network models' loss values converge to 0 within 40 training times, as shown in Figs. 9(a) and (b). In the presence of jamming strategy I, our PPO-GPT algorithm is more robust. The other three DRL algorithms are weak in jamming strategy II environments. DRL learning may be harder due to the complexity of jamming states. The robustness of the four DRL algorithms remains consistent in environments with jamming strategies III and IV. Overall, jamming strategy II is the hardest to learn of the four. The proposed method stabilizes the neural network model in 215 training times, proving that the PPO-GPT algorithm is more robust than the PPO-LSTM, A2C, and DDQN algorithms. We analyze the reward function variation curve during training to verify the convergence performance of the PPO-GPT learning algorithm. The DRL algorithm's reward function shows convergence performance when finding the optimal solution. Analyzing the reward function convergence curves based on the four jamming strategies shows that the PPO-GPT algorithm is superior to the others. All neural network models' loss values converge to 0 within 40 training times, as shown in Figs. 9(a) and (b). In the presence of jamming strategy I, our PPO-GPT algorithm is more robust. The other three DRL algorithms are weak in jamming strategy II environments. DRL learning may be harder due to the complexity of jamming states. The robustness of the four DRL algorithms remains consistent in environments with jamming strategies III and IV.

We test the convergence of the dual PPO-GPT model based on the joint reward function R^{joint} in Experiments 1 and 2. The joint reward function convergence curve with episodes is shown in Fig. 10. R^{joint}'s average reward values were recorded every 1000 episodes.

The reward increases from 0 to 2 for the first 0.23 million episodes in Fig. 10(a). In the exploratory phase of increasing reward values, the DRL algorithm's neural network weight parameters are not optimal. The neural network learns the jamming strategy and generates anti-jamming actions by trial and error. Reward feedback helps find the best anti-jamming action. After 0.23 million episodes, the PPO-GPT algorithm's reward function quickly converges to optimal solution 2 and stays there. The PPO-LSTM algorithm takes 1.77 million episodes to reach an optimal solution. The DDQN algorithm is just convergent after 2 million episodes. A2C is better than random, but it converges slowly. In Fig. 10(b), the PPO-GPT algorithm's reward gradually converges to the optimal solution at 0.124 million episodes and then remains stable. It takes 0.785 million episodes for the DDQN algorithm to find the optimal solution. After 1 million episodes, the PPO-LSTM and A2C algorithms will reach an optimal solution.

We evaluate the convergence of the dual PPO-GPT model based on the joint reward function R^{joint} in Experiments 3 and 4. The R^{joint} convergence curve with episodes

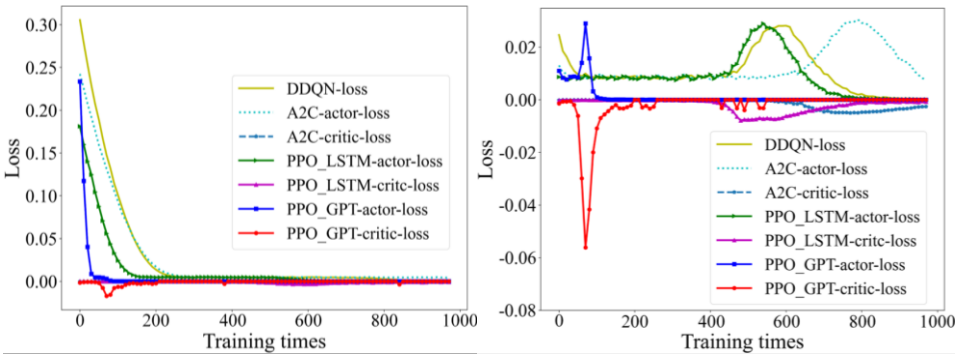

Fig. 8. Loss curve with the training times. (a) and (b) are Experiment 1 and 2 neural network model training results.

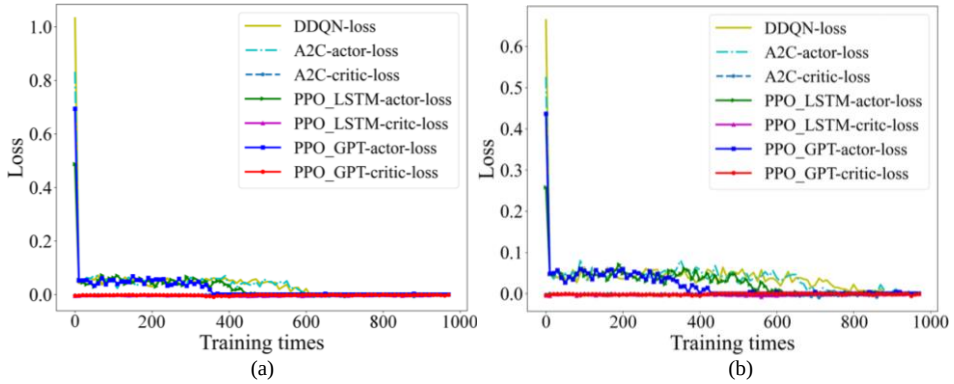

Fig. 9. Loss curve with the training times. (a) and (b) are Experiment 3 and 4 neural network model training results.

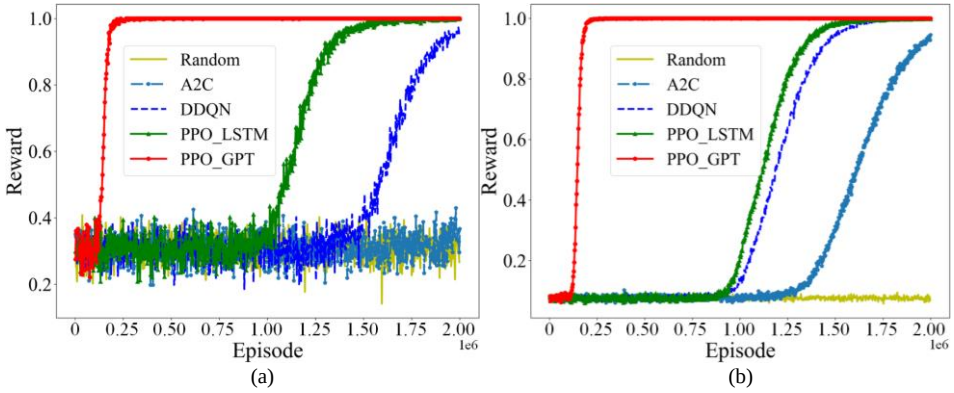

Fig. 10. Convergence curves of the reward with episodes. (a) and (b) are the reward convergence curves at Experiment 1 and 2.

is shown in Fig. 11. Figure 11(a) shows that the reward starts at 0 and increases over the first 0.015 million episodes. PPO-GPT algorithm rewards quickly converge to optimal solution 2 after 0.015 million episodes. The PPO-LSTM algorithm does not reach the optimal solution 2 until 0.035 million episodes. After 0.065 million episodes, the A2C and DDQN algorithms reach optimal solution 2. Figure 11(b) shows that the PPO-GPT algorithm reward gradually converges to optimal solution 2 from 0.023 million episodes and then remains stable. The PPO-LSTM algorithm approaches optimal solution 2 after 0.061 million episodes. DDQN reaches optimal solution 2 at 0.1 million episodes. The A2C algorithm finds optimal solution 2 at 0.104 million episodes. To conclude, Experiments 3 and 4 improve convergence by an order of magnitude over Experiments 1 and 2. It appears that these four DRL algorithms are easier to learn than jamming strategies III and IV. When optimizing sub-frequency coding, Equations (18) and (19) show that carrier frequency coding between pulses must be optimized. Optimizing carrier frequency coding between pulses does not require sub-frequency coding learning. Therefore, the DRL algorithm has a harder time learning jamming strategies I and II. Furthermore, the PPO-GPT algorithm outperforms the other four algorithms in average convergence speed by 72.5%.

4.3 Anti-jamming Performance Verification of Intelligent Learning Method

We look at how $\mu_T(\boldsymbol{\Psi})$ changes during the FAR-jammer confrontation to determine if the dictionary matrix made by IPCFH-IPSFC waveforms meets the learning

requirement. Equation (18) yields 1 if the constraint is met. The judgment result is 0 if $\mu_I(\boldsymbol{\Psi})$ does not satisfy the optimization constraint. Thus, we observe the episode-based judgment curve. To evaluate FAR's anti-jamming effect, we observe the echo signal SINR change. Equation (19) defines SINR as SJR_0 if the sub-pulse jams. SINR is SJR_0 if the sub-pulse is jam-free. Thus, we observe SINR's episode curve.

Figure 12 shows Experiment 1 and 2 observations. The judgment results curves with episodes in Figs. 12(a) and (c) have the same trend. Because jamming strategies I and II intercept sub-pulses, they do not affect carrier frequency learning. Thus, carrier frequency learning is independent of the environment and only depends on the optimization constraint. Until there are 0.162 million episodes, judgment results average is less than 1. Many episodes fail to satisfy the optimization constraint because the neural network is learning. The neural network has stabilized and learned the optimal solution after 0.162 million episodes, as the average judgment results are all 1. Except for the random algorithm, the other three algorithms can satisfy the optimization constraint but take longer. The convergence time of the judgment functions for both the DDQN and A2C algorithms is similar, approximately 0.34 million episodes. The judgment function of the PPO-LSTM algorithm takes longer to converge, and it does not reach the optimal solution even by the end of the adversarial process.

Experiment 1 and 2 average SINR curves with episodes within a CPI are shown in Figs. 12(b) and (d). The SINR trend matches the reward convergence curves in Figs. 10(a) and (b). Training begins with a low average SINR because the algorithm is learning. After learning the jamming strategy in the environment, the algorithm can adjust the IPCFH-IPSFC waveform parameters to prevent sub-pulses from interfering. At SINR = 15 dB, the echo signal is guaranteed to be signal and noise.

In Fig. 2(b), the proposed PPO-GPT algorithm achieves the fastest optimization of the maximum SINR and can make precise anti-jamming decisions, ensuring that after jamming suppression, only the target signal and noise remain. Although the PPO-LSTM algorithm optimizes more slowly, it can also achieve precise anti-jamming. However, while DDQN algorithm shows an optimization trend, the time it takes is too long, and it does not converge to the optimal SINR by the end of the adversarial process. The anti-jamming performance of the A2C algorithm is not significantly different from the random algorithm, indicating that this algorithm fails in the jamming strategy environment set in Experiment 1. In Fig. 12(d), the PPO-GPT, PPO-LSTM, and DDQN algorithms all find the optimal strategy to achieve successful anti-jamming. The A2C algorithm still fails to find the optimal SINR, and increasing the number of adversarial rounds might be helpful for it.

Experiment 3 and 4 observations are in Fig. 13. Figures 13(a) and (c) show judgment curves with episodes. After 0.8 million episodes, our PPO-GPT algorithm learns the optimal solution. Other algorithms take longer. Except for the random algorithm, all other four algorithms converge to the maximum value of 1. The convergence speed of the judgment function for the PPO-GPT algorithm is the fastest, with similar convergence and optimization performance in both Experiment 3 and Experiment 4 environments. The convergence speed of the judgment function for the PPO-LSTM algorithm is second fastest. In Fig. 13(a), the DDQN and A2C algorithms show similar convergence performance for the judgment function. In Fig. 13(c), the DDQN algorithm has the worst convergence performance, with the convergence curve of the decision function showing spikes, which may be caused by instability in the algorithm.

Experiment 3 and 4 average SINR curves with episodes within a CPI are shown in Figs. 13(b) and (d). Experiments 3 and 4 use the same SJR, but Experiment 3's minimum SINR is 20 dB higher. In Experiment 3, echo signals are jammed less during algorithm learning than in Experiment 4. The average SINR is higher in jamming strategy IV due to the shorter jamming duration of the jammer. In Fig. 13(b), the proposed PPO-GPT algorithm optimizes the SINR the fastest, achieving a 15 dB SINR after jamming suppression at around 0.8 million episodes, leaving only the target signal and noise. The PPO-LSTM algorithm optimizes more slowly, reaching the 15 dB SINR at approximately 1.12 million episodes. However, while the DDQN and A2C algorithms show an optimization trend, they take much longer, around 1.4 million episodes.

In Fig. 13 (d), the PPO-GPT algorithm continues to maintain a clear advantage, and in this Experiment 4 jamming environment, the performance of the five algorithms is similar to the anti-jamming results in Experiment 3. This is because both jamming strategy III and IV are based on pulse modulation, and the environment's effect on carrier frequency encoding is similar. By learning jamming strategies, the PPO-GPT algorithm can generate waveforms that satisfy the optimization constraint and anti-jamming performance. By blocking the jamming signal from entering the FAR receiver, the echo signal SINR can reach 15 dB.

4.4 Target Detection Verification of Intelligent Learning Method

Using the OMP-based target detection method, we output the optimal waveform parameters of the PPO-GPT algorithm to sparsely reconstruct targets, and validate the intelligent learning method's target detection performance. Experiment 1 employs an intelligent learning algorithm to optimize sub-frequency coding, targeting the sub-pulse for modulation. The optimal sub-frequency coding outputted by PPO-GPT is shown in Fig. 14(a). Blue is radar action, or the waveform sub-frequency. Red represents jamming, or the jamming signal sub-frequency. The radar sub-frequency targeted by the jammer is green. The jammer intercepts the sub-pulse at current T_{sub}, so it still targets the sub-frequency at next T_{sub}. The PPO-GPT algorithm learns the jamming strategy I and outputs an anti-jamming strategy that differentiates the sub-frequency of the current T_{sub}

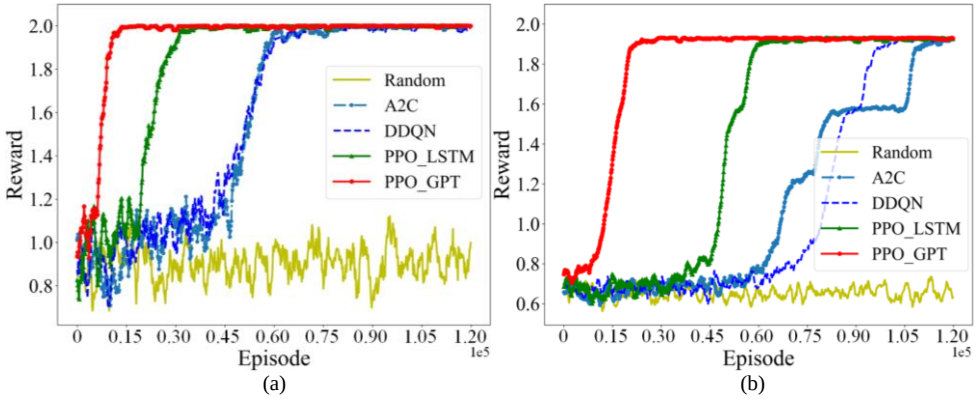

Fig. 11. Convergence curves of the reward with episodes. (a) and (b) are the reward convergence curves at Experiment 3 and 4.

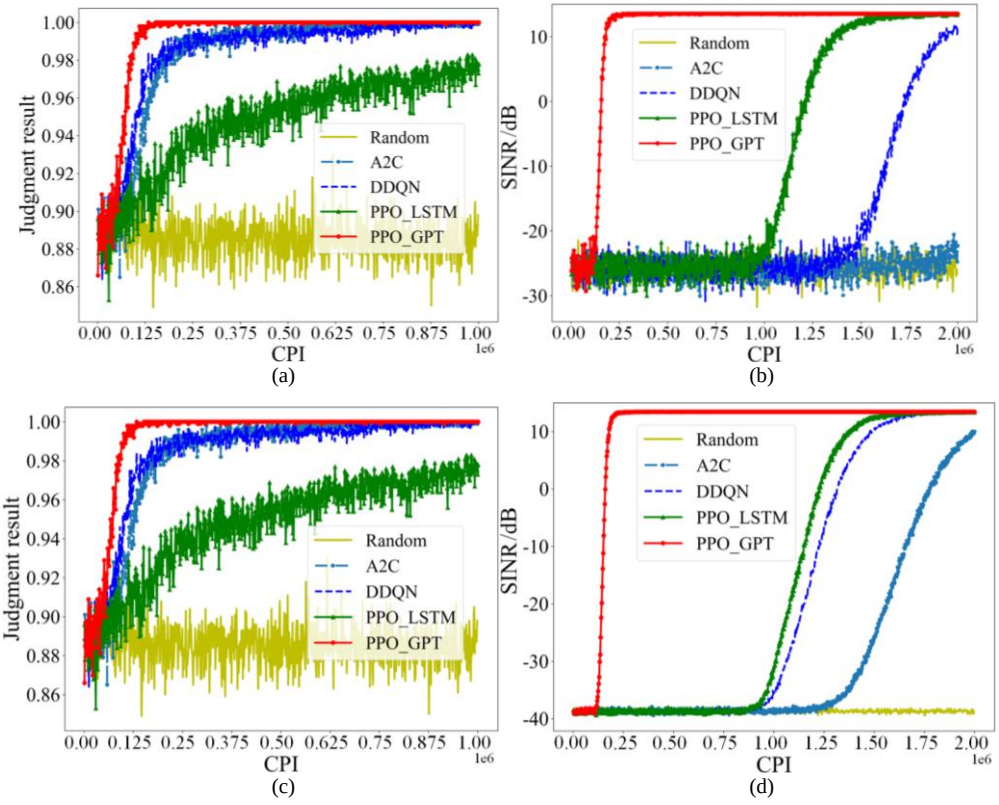

Fig. 12. The observations of Experiment 1 and Experiment 2. (a) is the curve of the judgment results with episodes in Experiment 1. (b) is the curve of the average SINR with episodes in Experiment 1. (c) is the curve of the judgment results with episodes in Experiment 2. (d) is the curve of the average SINR with episodes in Experiment 2.

from the sub-frequency of the next T_{sub}. Jamming strategy II is adjusted for sub-pulses in Experiment 2. The optimal sub-frequency coding outputted by PPO-GPT is shown in Fig. 14 (b). The width of two sub-pulses is the ISRJ repeater duration. The PPO-GPT algorithm learns the jamming strategy II and outputs an anti-jamming strategy to differentiate the sub-frequency in the current $2T_{sub}$ from the next $2T_{sub}$. We output the carrier frequency coding within a CPI and verify the target detection performance in Fig. 15. Figure 15(a) shows the ISRJ-containing sub-pulse compression result. The target has an 11 km range, but the ISRJ signal masks it, leaving only a false target at 12.5 km. The intra-pulse accumulation results of all sub-pulses with ISRJ are shown in Fig. 15(b). This occurs when the jammer accurately intercepts all sub-pulses in a PRI using the jamming strategy I. Target energy is higher than noise, but false target energy is 16 dB higher than real target energy. The fake target still fools the FAR. Figure 15(c) shows the two-dimensional sparse reconstruction of the jamming signal. The "Amplitude" axis represents the result after taking the absolute value of the signals. To ensure clarity and comparability of the results, we chose not to add units to the amplitude, as their introduction may make comparisons between different amplitudes less intuitive. Range of

the false target is 12.495 km and the velocity is 19.913 m/s. The two-dimensional sparse reconstruction of the real target signal is shown in Fig. 15(d). Real target's range is 10.9988 km, and its velocity is 19.913 m/s. They all meet FAR target detection requirements.

In Experiment 2, we tested the IPCFH-IPSFC waveform's target detection, as shown in Fig. 16. Since jammers all transmit ISRJ, Experiment 2's target detection results are similar to Experiment 1. Figures 16(a) and (b) show that when the jammer employs jamming strategy II, it can accurately intercept the radar transmission signal and use the noise signal to mask the real target, leaving only a false target at 12.5 km. As seen in Figs. 16(c) and (d), under this jamming strategy II scenario, the algorithm we designed not only achieves target range and velocity estimation but also shows that the detection noise in other range-velocity cells is nearly zero. In conclusion, the two-dimensional reconstruction results of range and velocity show that ISRJ mostly modulates target range information and not velocity information. FAR cannot detect real target information in ISRJ's environment without optimizing IPCFH-IPSFC waveform parameters.

In Experiment 3, the intelligent learning algorithm optimizes carrier frequency coding, while smart noise jamming modulates pulses. Figure 17(a) shows the optimal PPO-GPT carrier frequency coding output within a CPI. The jammer intercepts one pulse in current T_r and modulates it, so it still targets the carrier frequency of the pulse in next T_r. The PPO-GPT algorithm learns the jamming strategy III and outputs an anti-jamming strategy to differentiate the carrier frequency in the current T_r from the next T_r. Thus, it blocks the FAR receiver from receiving the jamming signal. Experiment 4 modulates jamming strategy IV for pulses. Figure 17(b) shows the optimal PPO-GPT

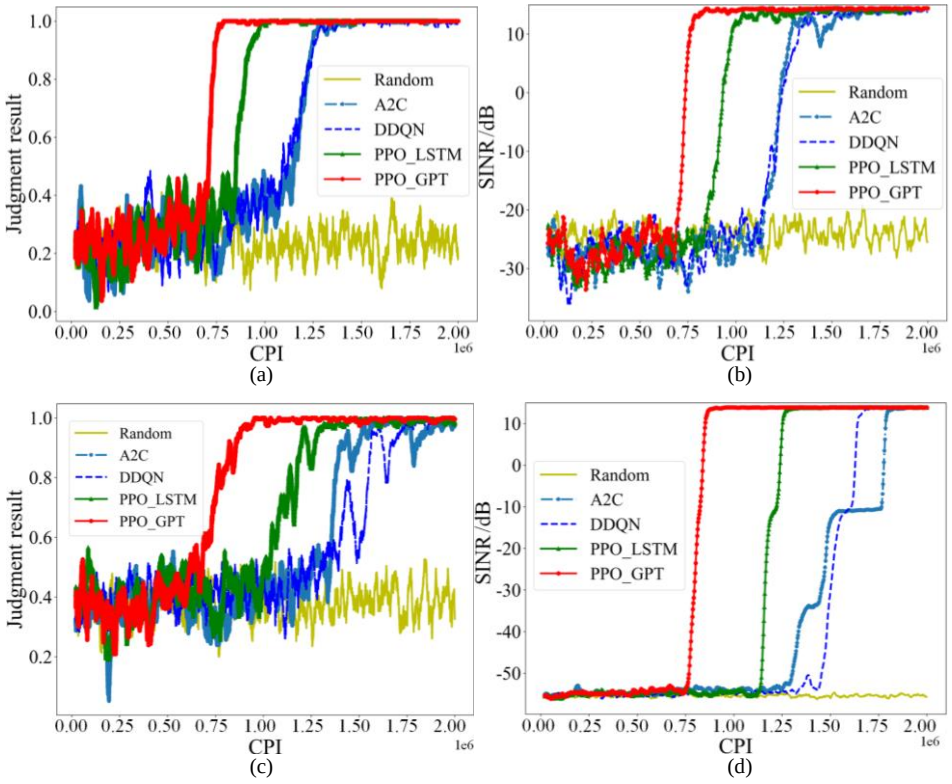

Fig. 13. The observations of Experiment 3 and Experiment 4. (a) is the curve of the judgment results with episodes in Experiment 3. (b) is the curve of the average SINR with episodes in Experiment 3. (c) is the curve of the judgment results with episodes in Experiment 4. (d) is the curve of the average SINR with episodes in Experiment 4.

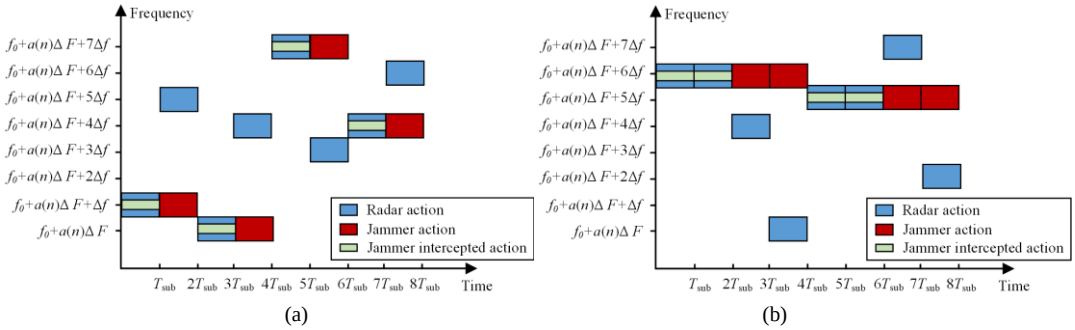

Fig. 14. Optimal sub-frequency coding. (a) and (b) are the optimal sub-frequency coding in Experiment 1 and 2.

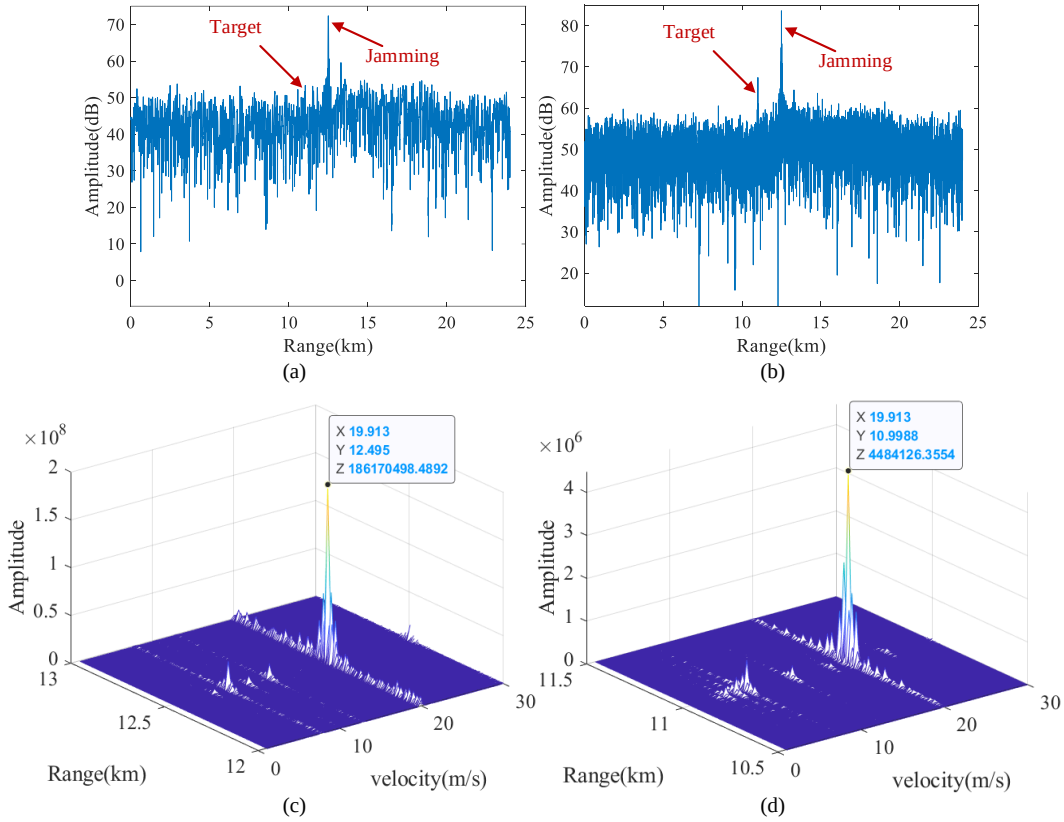

Fig. 15. Target detection result in Experiment 1. (a) is the pulse compression result for the sub-pulse signal with ISRJ. (b) is the intra-pulse accumulation result of sub-pulses with ISRJ. (c) is the two-dimensional sparse reconstruction result of the fake target signal. (d) is the two-dimensional sparse reconstruction result of the real target signal.

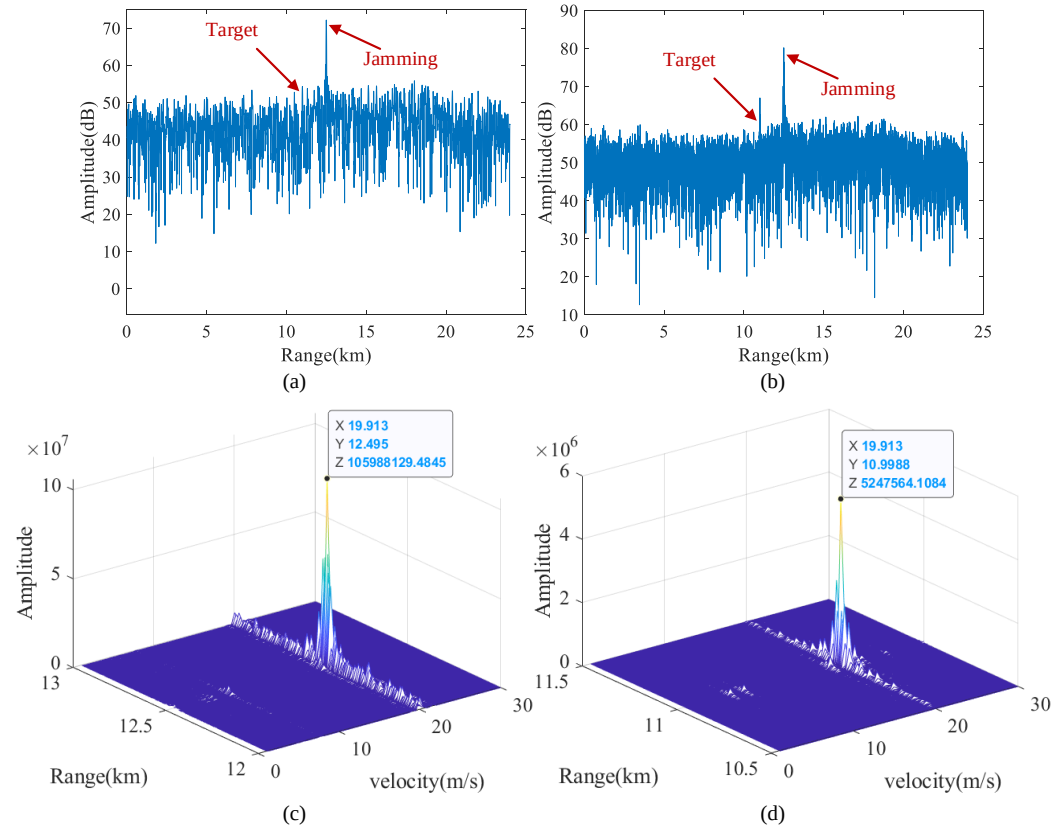

Fig. 16. Target detection result in Experiment 2. (a) is the pulse compression result for the sub-pulse signal with ISRJ. (b) is the intra-pulse accumulation result of sub-pulses with ISRJ. (c) is the two-dimensional sparse reconstruction result of the fake target signal. (d) is the two-dimensional sparse reconstruction result of the real target signal.

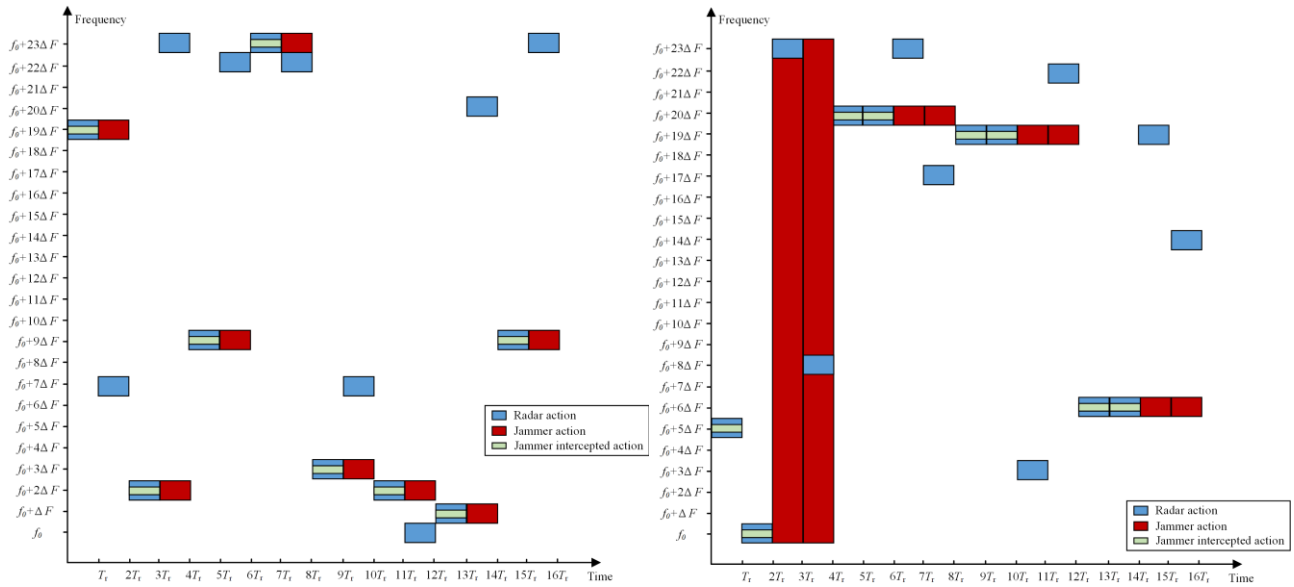

Fig. 17. Optimal carrier frequency coding. (a) and (b) are the optimal carrier frequency coding in Experiment 3 and 4.

carrier frequency coding output within a CPI. Two PRIs make up the jamming repeater's duration. Because the FAR transmits pulses with different carrier frequencies, the jammer may modulate the jamming signal to a wider band in the first two PRIs. To prevent wide-band jamming, the PPO-GPT algorithm generates transmitted signals with the same carrier frequency in two PRIs after learning the jamming strategy IV. Also, the carrier frequency in the current $2T_{\text{sub}}$ is different from the next $2T_{\text{sub}}$.

We test the IPCFH-IPSFC waveform's target detection performance using optimal carrier frequency coding in Fig. 18. Figure 18(a) shows smart noise jamming pulse compression results. Noise completely masks the target signal in range cells [7.2, 20.4] km. Figure 18(b) shows pulse accumulation within one CPI with smart noise jamming. The assumption is that the jammer accurately intercepts all CPI pulses using strategy III. The FAR cannot identify the target because the range unit only contains noise. The two-dimensional sparse reconstruction of the jamming signal is shown in Fig. 18(c) after target detection. Smart noise jamming is accumulated from many false targets, and the highest-energy false target has a range of 12.4575 km and a velocity of 21.1739 m/s. The two-dimensional sparse reconstruction of the real target signal is shown in Fig. 18(d). It can be seen that when detecting the real target in the absence of smart noise jamming, the estimation errors for its range and velocity are within acceptable limits, and no noise in other detection cells is estimated as a real target. This demonstrates the accuracy of the target detection algorithm based OMP we designed.

In Experiment 4, we tested the target detection using the IPCFH-IPSFC waveform, as shown in Fig. 19. The target detection results of Experiment 4 are similar to those of Experiment 3. Figures 19(a) and (b) show that when the jammer employs jamming strategy IV, the smart noise jamming completely masks the real target. Unlike ISRJ, which generates false target signals to mislead the radar, smart noise jamming creates numerous dense false targets accompanied by high noise power, resulting in both deceptive and suppressive jamming effects on the radar. As seen in Figs. 19(c) and (d), under jamming strategy IV, after two-dimensional sparse reconstruction, the smart noise jamming generates multiple false targets. The false target with the highest energy is located at a range of 13.1884 km and a velocity of 16.7888 m/s, while the real target remains undetectable. Since jammers transmit smart noise jamming but have different durations, the two-dimensional reconstruction results of range and velocity show that smart noise jamming can deceive and suppress target signals in range-velocity dimensions.

In conclusion, our proposed intelligent learning method can prevent the jamming signal from entering the FAR receiver by learning the jamming strategy and accurately estimating the target's range and velocity using the two-dimensional sparse reconstruction method in four jamming environments. Range and velocity estimation errors are 0.01% and 0.34%, respectively.

4.5 Decision-making Accuracy Verification of Intelligent Learning Method

We calculate the symmetric mean absolute percentage error (SMAPE) between the current and optimal decision-making results to verify the accuracy of our proposed intelligent learning method in an environment with multiple jamming strategies in chronological order. Using SMAPE, we define intelligent learning's decision-making accuracy as

$$\text{ACC} = \left|1 - \text{SMAPE}\right| \times 100\%, \tag{22}$$

$$\text{SMAPE} = \frac{100\%}{B} \sum_{b=1}^{B} \frac{\left|\hat{x}_b - x_b\right|}{\left(\left|\hat{x}_b\right| + \left|x_b\right|\right)/2} \tag{23}$$

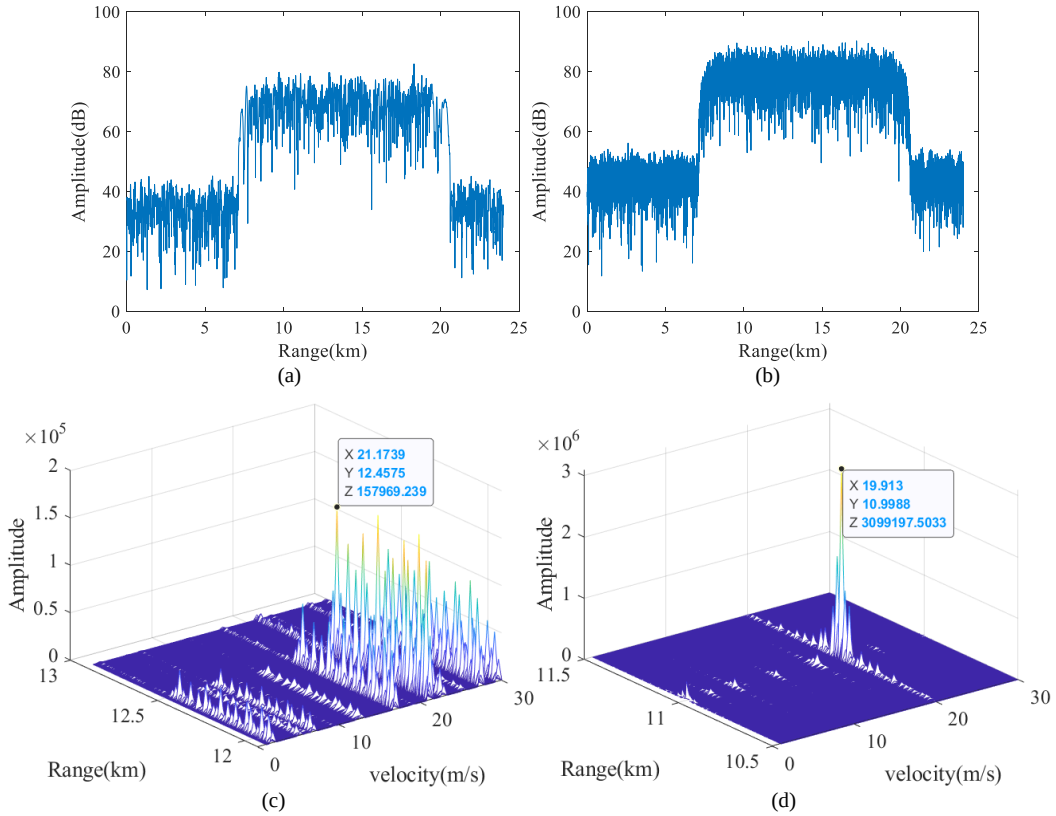

Fig. 18. Target detection result in Experiment 3. (a) is the pulse compression result for pulse signal with smart noise jamming. (b) is the pulse accumulation result of pulses with smart noise jamming. (c) is the two-dimensional sparse reconstruction result of the jamming signal. (d) is the two-dimensional sparse reconstruction result of the real target signal.

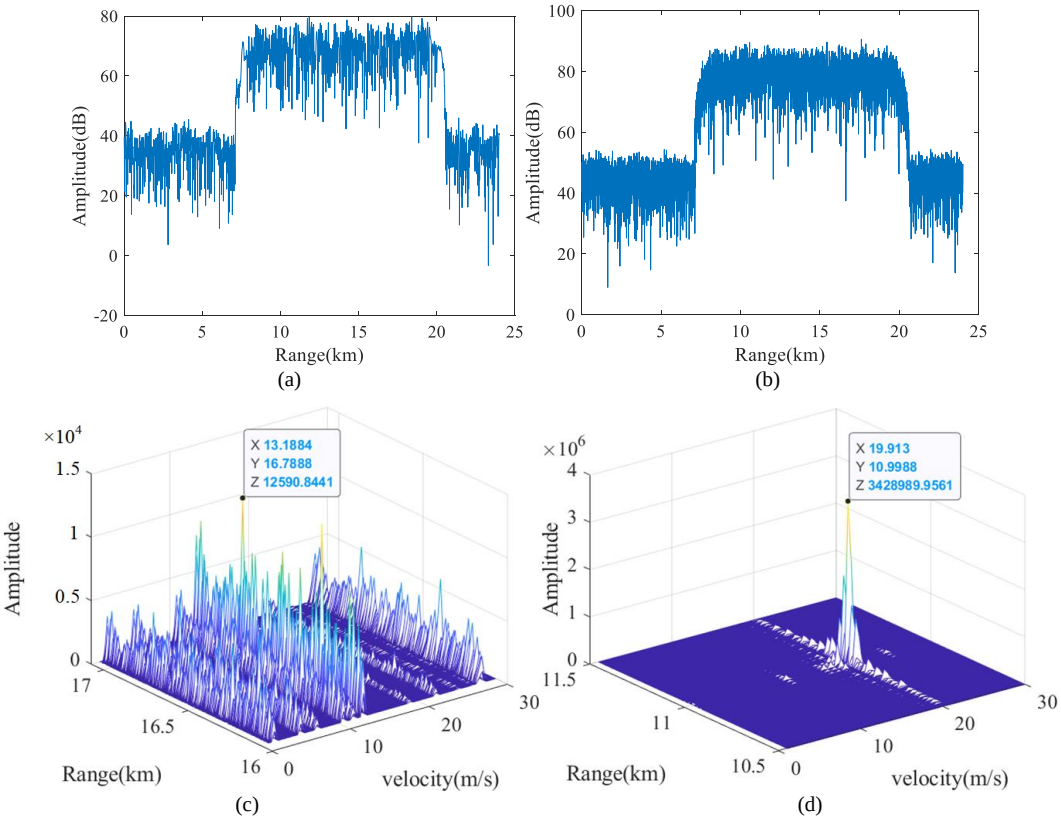

Fig. 19. Target detection result in Experiment 4. (a) is the pulse compression result for the pulse signal with smart noise jamming. (b) is the pulse accumulation result of pulses with smart noise jamming. (c) is the two-dimensional sparse reconstruction result of the jamming signal. (d) is the two-dimensional sparse reconstruction result of the real target signal.

where B is the sample size. x_b is the b-th sample optimal solution. $\hat{x}_b$ is the bth sample predicted value.

We assume a jamming environment with sequential jamming strategies. Figure 20 shows how CPI cycles transform the four jamming strategies chronologically. A game cycle has 300 CPIs: Within 0 to 60 CPIs, the jammer uses strategy III, and within 61 to 180 CPIs, strategy II. Then it uses jamming strategy IV in 181–260 CPIs. Finally, the 261–300 CPIs use jamming strategy I.

In testing the decision accuracy of the intelligent learning method, we used an approach where the algorithm is trained offline, while anti-jamming strategies are generated and executed online. Specifically, we first trained the designed FAR intelligent learning method based on the PPO-GPT algorithm using an offline approach, and the trained data was stored in an anti-jamming knowledge base. When facing different jamming strategies, the FAR system can timely select anti-jamming strategies from various knowledge bases through an online selection process.

In detail, for the four experiments mentioned above, after learning algorithm was trained and tested, we saved the last 50,000 decision results from the test in chronological order and stored them in the anti-jamming knowledge base. From a data perspective, we set up four arrays, each containing 50,000 decision results for the four experiments. The first array, Anti-jamming Strategy Library I, contains 50,000 anti-jamming strategies for jamming strategy I. The second array, Anti-jamming Strategy Library II, contains 50,000 anti-jamming strategies for jamming strategy II. The third array, Anti-jamming Strategy Library III, contains 50,000 anti-jamming strategies for jamming strategy III. The fourth array, Anti-jamming Strategy Library IV, contains 50,000 anti-jamming strategies for jamming strategy IV. When the jamming environment is set to the jamming strategy shown in Fig. 20, at each time step, after the anti-jamming assessment is complete and the jamming strategy is determined, we randomly sample strategies from the corresponding anti-jamming strategy library. We then generate the IPCFH-IPSFC waveform based on the selected anti-jamming strategy and engage in countermeasures against the jamming environment. We used Monte Carlo simulations to test the average decision accuracy of the intelligent learning methods over 1,000 adversarial cycles, as shown in Fig. 21.

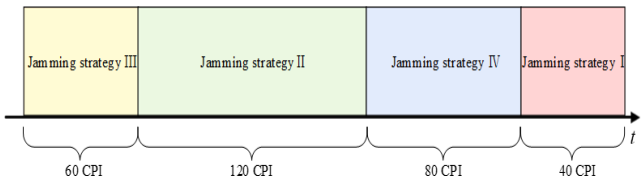

Fig. 20. Multiple jamming strategies in a game cycle.

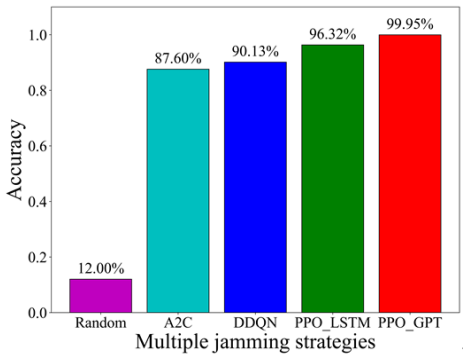

Fig. 21. The average value of the accuracy in the environment of multiple jamming strategies.

It shows that the PPO-GPT algorithm has 99.95% decision-making accuracy. The decision-making accuracy of PPO-LSTM, DDQN, and A2C algorithms is above 85%. We think the database's optimal solution percentage affects decision-making accuracy. The PPO-GPT algorithm has the highest percentage of optimal solutions in the database because it finds them quickly during training.

4.6 Applications and Discussion

The proposed method can be applied to the field of FAR for active jamming suppression through offline training and online execution. For example, in the jamming scenario involving jamming strategy I, where the jammer generates ISRJ according to jamming strategy I, FAR generally needs to follow the steps below to counter the jammer:

(a) Simulating the jamming environment: Write the code for the jamming environment according to jamming strategy I to generate the jamming signal.

(b) Generating waveform optimization constraint via simulation: According to the method in Sec. 3.4 and the experimental steps in Sec. 4.1, pre-generate the waveform optimization thresholds to provide optimization criteria for FAR to counteract jamming.

(c) Offline training of the proposed algorithm: The process of one round of confrontation between FAR and the jammer is as follows: The jammer generates a jamming signal based on jamming strategy I. FAR simulates the IPCFH-IPSFC waveform according to Sec. 2.1 and simulates the target signal received by the FAR receiver for the current round. The jamming signal and the target signal are additively combined, and then the combined echo signal is processed using the method in Sec. 3.1. The processed signal is then evaluated for anti-jamming effectiveness using the method in Sec. 3.2 and target detection is performed using the method in Sec. 3.3. Finally, the anti-jamming evaluation results and target detection results are fed back to the learning algorithm module proposed in Sec. 3.5. The PPO-GPT algorithm is used to optimize the anti-jamming strategy, generating the optimal IPCFH-IPSFC waveform parameters for the current round to engage in the next round of confrontation.

This confrontation process is repeated in a loop to train the PPO-GPT algorithm, optimizing its anti-jamming

effectiveness in the simulated jamming environment. According to the method in Sec. 4.5, the offline-trained data is stored in the anti-jamming knowledge base.

(d) Online execution of the proposed algorithm: In practical applications, the anti-jamming knowledge base is loaded onto the FAR system. According to the method in Sec. 4.5, when the real-world environment corresponds to a jamming scenario involving jamming strategy I, the FAR decision system searches for the corresponding anti-jamming strategy in the knowledge base. Based on the anti-jamming strategy, FAR generates the IPCFH-IPSFC waveform to counteract the jamming.

This process ensures that the proposed method can adapt to real-world conditions while maintaining the efficiency and effectiveness of FAR in jamming suppression.

5. Conclusions

The FAR signal model of the IPCFH-IPSFC agile waveform and active jamming signal models were created. The discontinuity of jamming signals in the time domain determines the four jamming strategies. We created a FAR intelligent learning model and designed five sub-modules. FAR achieved adaptive anti-jamming and precise target range-velocity estimates through five sub-module interactions.

We use Monte Carlo experiments to examine the relationship between sparse reconstruction success, JSR, and dictionary matrix correlation. Finally, the optimal threshold $\mu_T^{\text{opt}}(\boldsymbol{\Psi}) = 0.24$ constrains the intelligent learning method. Based on the four jamming strategies, we create four experiments to test our method's robustness, convergence, anti-jamming, and target detection. Our suggested method can quickly learn the jamming strategy in four jamming environments and stabilize the neural network model in 215 training times by assessing its loss function. Analysis of the reward function and comparison with the other four algorithms improves the average convergence speed of our proposed method by 72.5% and finds the optimal solution. Our suggested method stops the jamming signal from reaching the FAR receiver by continuously learning the jamming strategy. At the end of the learning process, the high echo signal SINR can reach up to 15 dB. Our method can accurately estimate target information, and the range and velocity estimate errors are 0.01% and 0.34%, respectively. Finally, the suggested method's decision-making accuracy is 99.95% in the environment of sequential jamming strategies.

Our future study will focus on intelligent learning methods for agile waveform anti-jamming based on chronologically ordered multiple jamming strategies. To prove our strategy works, we will conduct experiments to gather more data and expert knowledge and add them to the knowledge base. Note: Part of the source codes or additional information are available upon request from the authors.

Acknowledgments

This work was sponsored by the National Natural Science Foundation of China under Grant U20B2041.

References

[1] LI, K., JIU, B., PU, W., et al. Neural fictitious self-play for radar antijamming dynamic game with imperfect information. *IEEE Transactions on Aerospace and Electronic Systems*, 2022, vol. 58, no. 6, p. 5533–5547. DOI: 10.1109/TAES.2022.3175186

[2] LI, K., JIU, B., LIU, H., et al. Game theoretic strategies design for monostatic radar and jammer based on mutual information. *IEEE Access*, 2019, vol. 7, p. 72257–72266. DOI: 10.1109/ACCESS.2019.2920398

[3] ZHANG, Y., WEI, Y., YU, L. Interrupted sampling repeater jamming recognition and suppression based on phase-coded signal processing. *Signal Processing*, 2022, vol. 198, p. 1–7. DOI: 10.1016/j.sigpro.2022.108596

[4] ZHU, Y., ZHANG, Z., LI, B., et al. Analysis of characteristics and suppression methods for self-defense smart noise jamming. *Electronics*, 2023, vol. 12, no. 15, p. 1–14. DOI: 10.3390/electronics12153270

[5] XIE, Q., LIU, C., MO, Z., et al. A novel pulse-agile waveform design based on random FM waveforms for range sidelobe suppression and range ambiguity mitigation. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, vol. 61, p. 1–12. DOI: 10.1109/TGRS.2023.3326840

[6] XI, R., MA, D., LIU, X., et al. Intra-pulse frequency coding design for a high-resolution radar against smart noise jamming. *Remote Sensing*, 2022, vol. 14, no. 20, p. 1–20. DOI: 10.3390/rs14205149

[7] DE MARTINO, A. *Introduction to Modern EW Systems*. Beijing (China): Electronic Industry Press, 2014.

[8] LIU, S., CAO, Y., YEO, T. S., et al. Range sidelobe suppression for randomized stepped-frequency chirp radar. *IEEE Transactions on Aerospace and Electronic Systems*, 2021, vol. 57, no. 6, p. 3874–3885. DOI: 10.1109/TAES.2021.3082670

[9] QUAN, Y., WU, Y., LI, Y., et al. Range-Doppler reconstruction for frequency agile and PRF-jittering radar. *IET Radar Sonar and Navigation*, 2018, vol. 12, no. 3, p. 348–352. DOI: 10.1049/iet-rsn.2017.0421

[10] ZHOU, R., XIA, G., YUE, Z., et al. Coherent signal processing method for frequency-agile radar. In *12th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*. Qingdao (China), 2016, p. 431–434. DOI: 10.1109/ICEMI.2015.7494227

[11] KIRK, B., NARAYANAN, R., GALLAGHER, K., et al. Avoidance of time-varying radio frequency interference with software-defined cognitive radar. *IEEE Transactions on Aerospace and Electronic Systems*, 2019, vol. 55, no. 3, p. 1090–1107. DOI: 10.1109/TAES.2018.2886614

[12] WEI, S., ZHANG, L., LIU, H. Joint frequency and PRF agility waveform optimization for high-resolution ISAR imaging. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, vol. 60, p. 1–23. DOI: 10.1109/TGRS.2021.3051038

[13] LV, M., CHEN, W., MA, J., et al. Joint random stepped frequency ISAR imaging and autofocusing based on 2D alternating direction method of multipliers. *Signal Processing*, 2022, vol. 201, p. 1–11. DOI: 10.1016/j.sigpro.2022.108684

[14] WANG, X., WANG, S., LIANG, X., et al., Deep reinforcement learning: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, vol. 35, no. 4, p. 5064–5078. DOI: 10.1109/TNNLS.2022.3207346

[15] SHAO, Y., LIEW, S., WANG, T. AlphaSeq: Sequence discovery with deep reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, vol. 31, no. 9, p. 3319 to 3333. DOI: 10.1109/TNNLS.2019.2942951

[16] WEI, J., WEI, Y., YU, L., et al. Radar anti-jamming decision-making method based on DDPG-MADDPG algorithm. *Remote Sensing*, 2023, vol. 15, no. 16, p. 1–25. DOI: 10.3390/rs15164046

[17] AILIYA, WEI, Y., YE, Y. Reinforcement learning-based joint adaptive frequency hopping and pulse-width allocation for radar anti-jamming. In *IEEE Radar Conference (RadarConf20)*. Florence (Italy), 2020, p. 1–6. DOI: 10.1109/RadarConf2043947.2020.9266402

[18] LI, K., JIU, B., LIU, H., et al. Reinforcement learning based anti-jamming frequency hopping strategies design for cognitive radar. In *IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. Qingdao (China), 2018, vol. 14, no. 16, p. 1–5. DOI: 10.1109/ICSPCC.2018.8567751

[19] LI, K., JIU, B., LIU, H., et al. Deep Q-network based anti-jamming strategy design for frequency agile radar. In *2019 International Radar Conference (RADAR)*. Toulon (France), 2019, p. 1–5. DOI: 10.1109/RADAR41533.2019.171227

[20] AK, S., BRUGGENWIRTH, S. Avoiding jammers: A reinforcement learning approach. In *2020 IEEE International Radar Conference (RADAR)*. Florence (Italy), 2020, p. 321–326. DOI: 10.1109/RADAR42522.2020.9114797

[21] GENG, J., JIU, B., LI, K., et al. Radar and jammer intelligent game under jamming power dynamic allocation. *Remote Sen*sing, 2023, vol. 15, no. 3, p. 1–26. DOI: 10.3390/rs15030581

[22] LI, K., JIU, B., WANG, P., et al. Radar active antagonism through deep reinforcement learning: A way to address the challenge of mainlobe jamming. *Signal Processing*, 2021, vol. 186, p. 1–15. DOI: 10.1016/j.sigpro.2021.108130

[23] LI, K., JIU, B., LIU, H., et al. Robust antijamming strategy design for frequency-agile radar against main lobe jamming. *Remote Sensing*, 2021, vol. 13, no. 15, p. 1–24. DOI: 10.3390/rs13153043

[24] GENG, J., JIU, B., LI, K., et al. Reinforcement learning based radar anti-jamming strategy design against a non-stationary jammer. In *2022 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC)*. Xi'an (China), 2022, p. 1–5. DOI: 10.1109/ICSPCC55723.2022.9984459

[25] ZHOU, K., LI, D., HE, F., et al. A sparse imaging method for frequency agile SAR. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, p. 1–16. DOI: 10.1109/TGRS.2022.3151079

[26] LIU, H., ZHOU, S., SU, H., et al. Detection performance of spatial-frequency diversity MIMO radar. *IEEE Transactions on Aerospace and Electronic Systems*, 2014, vol. 50, no. 4, p. 3137 to 3155. DOI: 10.1109/TAES.2013.120040

[27] WEI, S., ZHANG, L., MA, H., et al. Sparse frequency waveform optimization for high-resolution ISAR imaging. *IEEE Transactions on Geoscience and Remote Sensing*, 2020, vol. 58, no. 1, p. 546–566. DOI: 10.1109/TGRS.2019.2937965

[28] QUAN, Y., LI, Y., WU, Y., et al. Moving target detection for frequency agility radar by sparse reconstruction. *Review of Scientific Instruments*, 2016, vol. 87, no. 9, p. 1–8. DOI: 10.1063/1.4962700

[29] ZHOU, K., LI, D., SU, Y., et al. Joint design of transmit waveform and mismatch filter in the presence of interrupted sampling repeater jamming. *IEEE Signal Processing Letters*, 2020, vol. 27, p. 1610–1614. DOI: 10.1109/LSP.2020.3021667

[30] LIU, Z., ZHANG, Q., LI, K. A smart noise jamming suppression method based on atomic dictionary parameter optimization decomposition. *Remote Sensing*, 2022, vol. 14, no. 8, p. 1–14. DOI: 10.3390/rs14081921

[31] LI, R., WANG, X., LI, G., et al. TEFISTA-Net: A learnable method for high-resolution range profile reconstruction with low-frequency ultra-wideband radar. *Signal Processing*, 2024, vol. 214, p. 1–14. DOI: 10.1016/j.sigpro.2023.109257

[32] POURNAGHSHBAND, R., MODARRES-HASHEMI, M. A novel block compressive sensing algorithm for SAR image formation. *Signal Processing*, 2023, vol. 210, p. 1–14. DOI: 10.1016/j.sigpro.2023.109053

[33] HUANG, W. A genetic algorithm optimized undersampling method for seismic sparse acquisition and reconstruction. *IEEE Transactions on Geoscience and Remote Sensing*, 2023, vol. 61, p. 1–10. DOI: 10.1109/TGRS.2023.3252277

[34] MÍGUEZ, J., MARIÑO, I. P., VÁZQUEZ, M. A. Analysis of a nonlinear importance sampling scheme for Bayesian parameter estimation in state-space models. *Signal Processing*, 2018, vol. 142, p. 281–291. DOI: 10.1016/j.sigpro.2017.07.030

[35] ZHENG, J., KURT, M. N., WANG, X. Stochastic integrated actor-critic for deep reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems*, 2024, vol. 35, no. 5, p. 6654–6666. DOI: 10.1109/TNNLS.2022.3212273

[36] ZHANG, W., ZHAO, T., ZHAO, Z., et al. Performance analysis of deep reinforcement learning-based intelligent cooperative jamming method confronting multi-functional networked radar. *Signal Processing*, 2023, vol. 207, p. 1–13. DOI: 10.1016/j.sigpro.2023.108965108965

[37] AKIN, E. Deep reinforcement learning-based multirestricted dynamic-request transportation framework. *IEEE Transactions on Neural Networks and Learning Systems*, 2023, p. 1–11. DOI: 10.1109/TNNLS.2023.3341471

[38] XIONG, K., ZHANG, T., CUI, G., et al. Coalition game of radar network for multitarget tracking via model-based multiagent reinforcement learning. *IEEE Transactions on Aerospace and Electronic Systems*, 2023, vol. 59, no. 3, p. 2123–2140. DOI: 10.1109/TAES.2022.3208865

About the Authors …

Jingjing WEI was born in Hebei Province, China, in 1995. She received the B.S. degree in Electrical Engineering and Automation in 2017, from Changchun University of Technology, Changchun, China. She received the M.S. degree in Electrical Engineering in 2019, from the Harbin Institute of Technology, Harbin, China. She is currently working toward the Ph.D. degree with the School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin, China. Her main research interests include radar anti-jamming strategy design and the study of intelligent learning methods for radar anti-jamming.

Lei YU (corresponding author) was born in Henan Province, China, in 1981. He received the B.Eng. degree in Signal Processing from Xidian University, Xi'an, China, in 2004, the M.Sc. degree from the University of Edinburgh, Edinburgh, UK, in 2005, and the Ph.D. degree in Electronic Engineering from the University of Sheffield, Sheffield, UK, in 2010. From 2011 to 2013, he was a postdoctoral researcher with the Department of Electrical Engineering,

Linköping University, Linköping, Sweden. He joined the Department of Electronics Engineering, Harbin Institute of Technology as an Associate Professor in 2014. His research interests include radar signal processing, anti-jamming, and waveform design.

Yinsheng WEI was born in Heilongjiang Province, China, in 1974. He received the M.S. and Ph.D. degrees in Information and Communication Systems from the Harbin Institute of Technology, Harbin, China, in 1998 and 2002 respectively. Then, he joined the Department of Electronics Engineering, Harbin Institute of Technology, as a Lecturer, and became a Professor in 2011. His research interests include anti-jamming radar waveform design, radar signal processing, and radar system analysis and simulation.

Rongqing XU was born in 1958. He received the B.S. and M. S. degrees in Electronic Engineering and the Ph.D. degree in Information and Communication Engineering from the Harbin Institute of Technology, Harbin, China, in 1982, 1984, and 1990, respectively. He is currently a Professor with the Institute of Electronic Engineering Technology, Harbin Institute of Technology. His main research interests include the fields of radar signal processing.