Radar Jamming Waveform Design Based on Hierarchical Reinforcement Learning and Cooperative Perception

1st Lijun Gao
School of Computing
Shenyang Aerospace University
Shenyang, China
gaolijun@sau.edu.cn

2nd Mengdi Zhao
School of Computing
Shenyang Aerospace University
Shenyang, China
zhaomdi2024@outlook.com

3rd Yanjie Wang
School of Computing
Shenyang Aerospace University
Shenyang, China
iris13836374130@gmail.com

4th Rui Liu
School of Computing
Shenyang Aerospace University
Shenyang, China
liurui200114@gmail.com

5th Xiyan Cao
School of Computing
Shenyang Aerospace University
Shenyang, China
cxy1748370525@163.com

6th Yanhui Liu
School of Computing
Shenyang Aerospace University
Shenyang, China
lyhhandsome2001@163.com

***Abstract*—Reinforcement Learning (RL) has attracted wide attention in many fields due to its dynamic decision-making and continuous optimization capabilities, but its practical application in radar jamming waveform generation remains limited. Traditional RL-based jamming algorithms face two core challenges: first, low sample efficiency makes it difficult to converge to the optimal strategy adapting to radar dynamic changes; second, poor adaptability of jamming waveforms fails to match radar signal characteristics in real time. Additionally, designing a reward function that accurately describes the radar-jammer interaction mechanism is highly challenging. This paper proposes an online jamming waveform generation framework: three time-frequency analysis methods (Smooth Pseudo Wigner-Ville Distribution, Fourier Synchrosqueezing Transform, and Variational Mode Decomposition-based Hilbert-Huang Transform) are used to convert complex multi-class radar signals under low signal-to-noise ratio (SNR) into three-channel time-frequency feature maps. Combined with Pulse Descriptor Words (PDW) to form a dual-stream input, this framework solves the problem of low radar signal type recognition accuracy under low SNR and serves as interpretable prior knowledge for RL. Under the hierarchical RL architecture, radar Power Spectral Density (PSD) matching degree is used as the core constraint to guide the agent for efficient directed exploration, improve sample efficiency, and enable learning more adaptive dynamic jamming strategies in complex agile environments. To further evaluate jamming effectiveness in non-cooperative scenarios, a composite reward function integrating frequency-domain and time-domain metrics is designed, combined with closed-loop optimization of the radar signal processing chain.**

***Index Terms*—Radar jamming, waveform design, electronic countermeasures, reinforcement learning, dynamic optimization**

I. Introduction

In modern electronic warfare systems, the collaborative closed loop of "radar signal recognition-jamming waveform generation" is the core to seize the initiative in electromagnetic countermeasures. Its performance directly affects the perception accuracy of battlefield electromagnetic situation and determines whether jamming strategies can take effect precisely. Traditional technologies face bottlenecks in three key stages: "unknown signal recognition", "jamming waveform generation", and "jamming effectiveness evaluation". Radar signal recognition suffers from low accuracy when identifying complex multi-class radar modulated signals under low SNR environments; jamming waveform generation mostly adopts fixed templates with low efficiency, failing to achieve precise disruption based on radar signal characteristics. In non-cooperative scenarios, effective jamming effectiveness evaluation methods have not yet been formed.

Radar signal modulation recognition is a core technology of radar reconnaissance, which is crucial for improving the performance of reconnaissance systems [1]. Early inter-pulse feature methods can hardly adapt to complex technologies, and intra-pulse features have become a research hotspot due to higher accuracy [2]. Driven by artificial intelligence, neural network-based recognition methods have become mainstream, and time-frequency feature input mode is widely used [3]–[5]. Reference [6] achieves recognition rate over 85% at -10 dB, but drops sharply by 15% at -12 dB. Reference [7] only relies on single time-frequency clustering, with purity below 75% at -10 dB. Reference [8] points out that time-frequency images are prone to energy dispersion under low SNR. Reference [9] achieves over 90% recognition rate for known signals at -10 dB, but ignores PDW time-domain parameters, resulting in a misjudgment rate of more than 30% for composite modulations such as LFM-BPSK. This paper proposes a novel recognition method based on time-frequency fusion and PDW dual-stream network.

Against the background of increasingly intelligent radar countermeasures, jamming waveform design methods are shifting from prior rule-dependent to data-driven adaptive generation paradigms. Early methods mostly adopted fixed templates or empirical formulas, introducing waveform similarity constraints to achieve energy concentration. For example, Reference [10] improves energy focusing through this constraint, and Reference [11] proposes MIMO radar waveform design based on KL divergence. Jamming generation technology has gradually evolved from "fixed template" to "adaptive optimization". Reference [12] uses DQN to optimize jamming patterns, but couples strategies and parameters into a high-dimensional action space, requiring more than 10^5 iterations for training with poor real-time performance. Reference [13] proposes a QMIX-based MARL cooperative jamming method, but fails to integrate radar prior information, leading to more than 40% invalid trial-and-error [14]. In summary, the main contributions of this paper are as follows:

1. A novel unknown radar signal recognition method is proposed: enemy I/Q signals are converted into fused time-frequency images via SPWVD, HHT, and FSST. Combined with PDW sequences to mine time-domain statistical features, a dual-stream neural network extracts features and an integration module completes adaptive weighting.

2. A jamming waveform generation method based on "two-level hierarchical reinforcement learning" is proposed: the upper-level policy network makes rapid decisions on jamming type and core parameter ranges based on signal recognition prior; the lower-level parameter optimization network adjusts parameters such as PRI under policy constraints to achieve local optimum.

3. A composite reward function integrating frequency-domain (PSD, JSR) and time-domain (autocorrelation peak) metrics is designed, constructing an RL-based cognitive jamming decision system for unknown radars.

The rest of this paper is organized as follows: Section II presents the research methodology. Section III details the experimental results and discussions. Finally, conclusions are drawn in Section IV.

II. Methodology

A. Unknown Signal Recognition

1) Overall Framework: To solve the defects of intra-pulse recognition methods such as difficult feature extraction and weak anti-jamming capability, dual-stream input is introduced for radar pulse recognition, as shown in Fig.1.

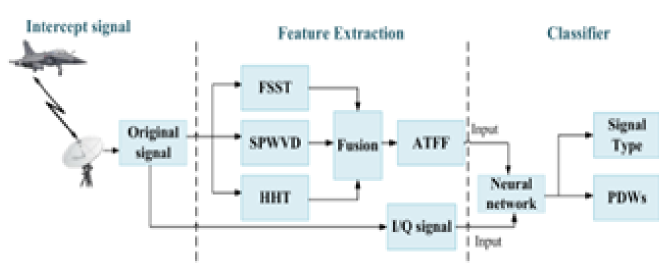

Fig. 1: Radar signal modulation recognition system

2) ATFF Features and PDW: Aiming at the insufficient robustness of single time-frequency representation under low SNR, this paper fuses three complementary time-frequency analysis technologies (SPWVD, FSST, VMD-HHT) to construct multi-channel time-frequency features (ATFF), making up for their limitations in resolution, anti-noise performance and multi-component analysis capability.

In the specific process: first, the same radar signal is converted into three single-channel time-frequency images via the above three methods respectively. After grayscale normalization, adaptive binarization, morphological dilation and resampling to 256×256, they are concatenated into a 256×256×3 tensor along the channel dimension. Meanwhile, Pulse Descriptor Words (PDW) contain physically interpretable parameters such as carrier frequency and pulse width, which can reflect the macro working characteristics of radar. This paper constructs ATFF and PDW dual-stream input: ATFF captures micro time-frequency structures, and PDW provides macro parameter prior. They are encoded separately in the subsequent network and make collaborative decisions, providing a reliable input basis for modulation recognition and intelligent jamming.

B. Hierarchical Reinforcement Learning

1) Overall Framework: To solve the core problem that traditional static jamming decision strategies are difficult to adapt to dynamic radars, this paper introduces a novel framework based on hierarchical reinforcement learning, as shown in Fig.2. It decomposes complex decision tasks into two collaborative subtasks: 1) Global jamming strategy formulation: determine jamming type and frequency range based on radar modulation type. 2) Local waveform parameter optimization: fine-tune parameters under upper-level constraints to maximize jamming effectiveness.

This chapter defines the state set, action set, environmental feedback and core elements of Markov Decision Process (MDP) for the model. All definitions are consistent with the actual process of "modulation type recognition→jamming strategy decision→waveform parameter optimization" to ensure engineering feasibility.

2) Global Jamming Strategy Formulation: The upper-level policy focuses on macro jamming decisions, selecting the optimal jamming type and frequency range based on radar modulation type to provide constraints for lower-level parameter optimization. This decision process is modeled as a Markov decision process, and the policy is learned through the Actor-Critic network.

State: Defined as $s_{\text{upper}} = \{T_{\text{mod}}, T_{\text{jam}}, F_{\text{range}}, P_{\text{PSD}}\}$, where T_{mod} is the radar signal modulation type; T_{jam} is the jamming type (deception jamming/suppression jamming) of global decision, matched from T_{mod} according to TableI; F_{range} is the jamming frequency range; P_{PSD} is the Power Spectral Density (PSD) matching degree between the current jamming waveform and radar signal.

Action: A combined decision of "jamming type + frequency range", generated based on the matching rules of T_{mod}, defined

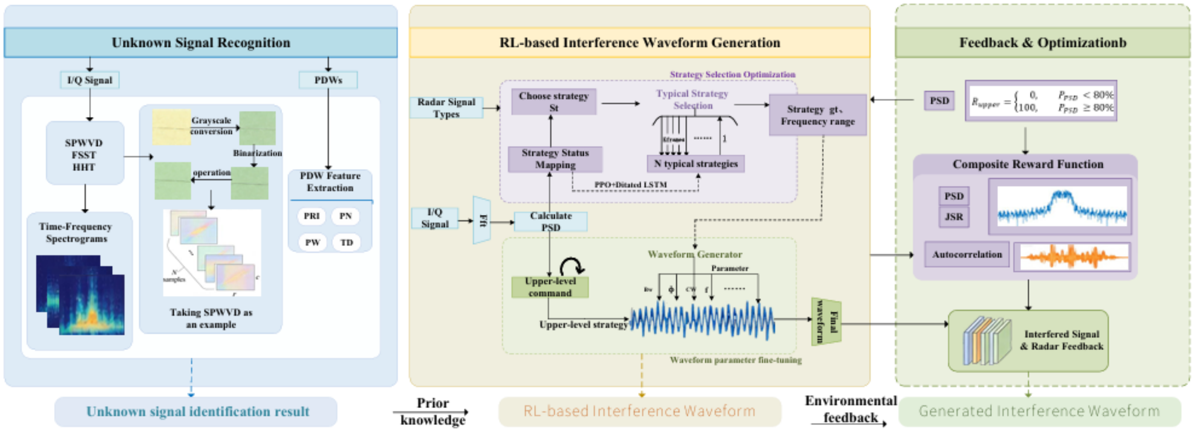

Fig. 2: Reinforcement learning modeling of Radar-Jammer system

TABLE I: T_{mod} Matching Table

Radar Type	Radar Function Type	Upper-level Jamming Type
Coherent Unmodulated	Short-range Search Radar	Suppression Jamming
LFM	Medium/Long-range Search Radar	Suppression Jamming
Barker	Tracking Radar	Deception Jamming
Polyphase Barker	Fire Control Radar	Deception Jamming (Priority: Range/Angle)
Frank	Medium-range Tracking Radar	Deception Jamming

as $A_{\text{upper}} = (T_{\text{jam}}, F_{\text{range}})$, which only serves as a constraint condition for lower-level actions.

Reward: A binary reward to guide the upper level to lock effective jamming strategies, defined as:

$$R_{\text{upper}} = \begin{cases} 0, & P_{\text{PSD}} < 80\% \\ 100, & P_{\text{PSD}} \geq 80\% \end{cases} \tag{1}$$

Transition Function: The upper-level MDP is defined as a quadruple $M_{\text{upper}} = (S_{\text{upper}}, A_{\text{upper}}, R_{\text{upper}}, P_{\text{trans, upper}})$. The upper-level state update only depends on P_{PSD}, i.e., $S_{\text{upper}}^{t+1} = \{T_{\text{mod}}, T_{\text{jam}}, F_{\text{range}}, P_{\text{PSD}}^{t+1}\}$, where $P_{\text{PSD}}^{t+1} = \text{Corr}(W_{\text{jam}}^{t+1}, W_{\text{radar}})$, and W_{jam}^{t+1} is generated by lower-level action iteration. The transition probability $P_{\text{trans, upper}} = 1$, only P_{PSD} is updated in real time with lower-level results, and other states remain fixed. The decision goal is to maximize $R_{\text{upper}} = 100$.

Policy Network: Dilated LSTM [15] with dilation rate $dr = 2$ is adopted as the core structure. The network inputs the upper-level state vector s_{upper}, processes it through the embedding layer and encoding layer, and generates the probability distribution of jamming strategy A_{upper} via the Softmax function.

3) Local Waveform Parameter Optimization: Local waveform parameter optimization focuses on micro-details, fine-tuning jamming waveform parameters under upper-level policy constraints. This process is also modeled as an MDP, realizing refined parameter optimization through reinforcement learning in continuous action space.

State: Defined as $s_{\text{lower}} = \{P_{\text{par}}, P_{\text{PSD}}, R_{\text{JSR}}, C_{\text{auto}}, C_{\text{upper}}\}$, where P_{par} is the set of adjustable jamming waveform parameters $\{P_{\text{jam}}, \text{PRI}, \text{PW}, f_c, B_{\text{FM}}, R, \theta\}$; R_{JSR} is the Jamming-to-Signal Ratio (ranging from $[0, 5]$), reflecting the intensity advantage of jamming signals; C_{auto} is the time-domain autocorrelation of radar signals (ranging from $[0, 1]$), reflecting the degree of interference to radar signal correlation; C_{upper} is the upper-level constraint (i.e., $(T_{\text{jam}}, F_{\text{range}})$), which limits the boundary of lower-level parameter optimization.

Reward: A multi-index cumulative reward to guide gradual parameter optimization, defined as:

$$\begin{aligned} R_{\text{lower}} &= R_{\text{PSD}} + R_{\text{JSR}} + R_{\text{auto}} \\ &= \frac{P_{\text{jam}} \cdot P_{\text{radar}}}{\|P_{\text{jam}}\| \cdot \|P_{\text{radar}}\|} \\ &\quad + \tanh\left(\frac{\text{Actual JSR} - \text{Target JSR}}{\text{Target JSR}}\right) \\ &\quad + \frac{1}{N} \sum_{k=1}^{N} \frac{|r_{\text{jam}}(k) - r_{\text{radar}}(k)|}{r_{\text{radar}}(k)} \end{aligned} \tag{2}$$

$$R_{\text{JSR}} = \begin{cases} 20, & R_{\text{JSR}} \geq 3 \\ 10, & 2 \leq R_{\text{JSR}} < 3 \\ -5, & R_{\text{JSR}} < 2 \end{cases} \tag{3}$$

$$R_{\text{auto}} = \begin{cases} 20, & C_{\text{auto}} \leq 0.5 \\ 5, & 0.5 < C_{\text{auto}} \leq 1 \\ -5, & C_{\text{auto}} > 1 \end{cases} \tag{4}$$

Transition Function: Defined as a quadruple $M_{\text{lower}} = (S_{\text{lower}}, A_{\text{lower}}, R_{\text{lower}}, P_{\text{trans, lower}})$. The state update satisfies $S_{\text{lower}}^{t+1} = \{P_{\text{par}}^t + A_{\text{lower}}^t, P_{\text{PSD}}^{t+1}, R_{\text{JSR}}^{t+1}, C_{\text{auto}}^{t+1}, C_{\text{upper}}\}$. The transition function $P(S'_{\text{lower}} | S_{\text{lower}}, A_{\text{lower}}) = 1$. After executing the lower-level action, the jamming parameters satisfy $P'_{\text{par}} = P_{\text{par}} + A_{\text{lower}}$. P'_{PSD}, R'_{JSR} and C'_{auto} are calculated through the new parameters to finally complete state iteration. The decision goal is to maximize R_{lower}. The lower-level policy network has the same structure as the upper-level.

4) Hierarchical Collaborative Training: This section proposes a reinforcement learning algorithm framework based on hierarchical collaborative training, realizing adaptive optimization of radar jamming waveforms through alternate updates of upper and lower-level policies. The framework adopts a hierarchical Dilated LSTM network structure, corresponding to the upper-level (jamming strategy planning layer) and lower-level (waveform parameter adjustment layer) respectively, and realizes collaborative training combined with the Clipped Proximal Policy Optimization (Clipped PPO) algorithm.

The specific execution process is as follows: First, the upper level generates frequency-domain jamming targets based on the radar feature library and current jamming effectiveness. The lower level outputs parameter adjustment actions through the Actor network based on the target and current waveform parameters, and calculates three indicators (power spectral density matching degree, jamming-to-signal ratio and autocorrelation feature) as immediate rewards. The lower-level Critic network estimates the state value function and calculates the advantage function, then updates the policy network through Clipped PPO to complete one lower-level iteration. Subsequently, the upper level calculates the global reward according to the jamming effectiveness after lower-level execution (e.g., whether the power spectral density matching degree reaches the success threshold), and updates the upper-level policy through a similar Clipped PPO mechanism after an episode ends. The two levels form a closed loop through target transmission and reward feedback: the upper level is responsible for macro jamming planning, and the lower level is responsible for refined parameter adjustment, ultimately realizing collaborative optimization of jamming strategies and waveform parameters. The specific process is shown in Algorithm1.

Algorithm 1 FeUdal PPO-Dilated LSTM for Radar Jamming Optimization

Require: Upper/Lower Dilated LSTM (PPO) initialized; $\alpha = 0.001$, $\gamma = 0.9$, $\varepsilon_{\text{clip}} = 0.2$, $d_r = 2$, suc_thres $= 80\%$, sta_thres $= 5\%$

for ep $= 1$ **to** Total_Episodes **do**
  $T \leftarrow$ Radchar Dataset; $S_U^0 = \{T, P_{\text{PSD}} = 0\%, \emptyset, \emptyset\}$
  goal $\leftarrow$ Upper_PPO(S_U^0, d_r)
  $P^0 = \{\text{Power} = 20\,\text{W}, \text{PRI} = 10\,\text{ms}, \text{PW} = 1\,\mu\text{s}\}$
  $S_L^0 = \{P^0, P_{\text{PSD}} = 0\%, R_{\text{JSR}} = 0, C_{\text{auto}} = 1, \text{goal}\}$
  sta_cnt $= 0$, $R_L^{\text{prev}} = 0$, $R_U^{\text{total}} = 0$
  for $t = 1$ **to** Max_Steps **do**
    $a_L \leftarrow$ Lower_Actor$(S_L^t, \text{goal}, d_r)$; $P^{t+1} = P^t + a_L$
    $W_j^{t+1} = f(P^{t+1}, \text{goal.Freq})$
    $P_{\text{PSD}}^{t+1} = \text{Corr}(W_j^{t+1}, W_{\text{radar}})$; $R_{\text{JSR}}^{t+1} = P_j^{t+1}/P_{\text{radar}}$
    $C_{\text{auto}}^{t+1} = \text{AutoCorr}(W_{\text{radar}} \oplus W_j^{t+1})$
    $R_L^t = 0.5 \times \Delta P_{\text{PSD}} + R_{\text{JSR\_reward}} + C_{\text{auto\_reward}}$
    $R_U^t = 100$
    if then
      $\text{P}_{\text{PSD}}^{t+1} \geq$ suc_thres
  else
    $R_U^{\text{total}} \mathrel{+}= \gamma^t \times R_U^t$
    Adv $= (R_L^t + \gamma \cdot \text{Lower\_Critic}(S_L^{t+1}, d_r)) -$ Lower_Critic(S_L^t, d_r)
    ratio $= \pi_{\text{new}}^L / \pi_{\text{old}}^L$
    $\text{loss}_L = -\min(\text{ratio} \cdot \text{Adv}, \text{clip}(\text{ratio}, 1-\varepsilon, 1+\varepsilon) \cdot \text{Adv}) + 0.5 \cdot \text{TD\_err}^2$
    Update Lower Dilated LSTM (minimize loss_L)
    $S_L^{t+1} = \{P^{t+1}, P_{\text{PSD}}^{t+1}, R_{\text{JSR}}^{t+1}, C_{\text{auto}}^{t+1}, \text{goal}\}$
    sta_cnt $=$ sta_cnt $+ 1$
    if then
      —$\text{R}_L^t - R_L^{\text{prev}}| \leq \text{sta\_thres} \cdot R_L^{\text{prev}}$
  else
    $S_U^{t+1} = \{T, P_{\text{PSD}}^{t+1}, \text{goal}, R_U^t\}$; $R_L^{\text{prev}} = R_L^t$
    $S_L^t = S_L^{t+1}$; $S_U^t = S_U^{t+1}$
    if $R_U^t = 100$ **and** sta_cnt ≥ 5 **then**
      break
    end if
  end for
  $\text{ratio}_U = \pi_{\text{new}}^U / \pi_{\text{old}}^U$
  $\text{loss}_U = -\min(\text{ratio}_U \cdot R_U^{\text{total}}, \text{clip}(\text{ratio}_U, 1-\varepsilon, 1+\varepsilon) \cdot R_U^{\text{total}})$
  Update Upper Dilated LSTM (minimize loss_U); Opt_Params $= P^{t+1}$
end for=0

III. Experimental Results and Discussion

A. Experimental Setup

The dataset adopts RadChar, covering various known modulation types such as BSK and LFM. The sampling frequency of simulated signals is 200 MHz, and the pulse width is 10 μs. The experiment is divided into two parts: "unknown radar signal recognition experiment" and "dynamic jamming waveform generation experiment".

B. Baseline

1) Radar Unknown Signal Recognition: To comprehensively evaluate the effectiveness of ATFF features and PDW dual-stream input, six types of time-frequency image features (WVD, CWD, FSST, SPWVD, HHT, ATFF) are selected for comparison. Subsequently, the recognition accuracy of various radar signal types under different SNR is compared, as well as the radar recognition accuracy without PDW stream input.

2) Jamming Waveform Generation Framework: To fully verify the effectiveness of the radar jamming waveform generation framework, three RL-based radar jamming methods are selected for comparison:

DQN [16]: Learns the optimal waveform selection strategy by maximizing the cumulative Q-value.

PPO [17]: Directly learns the probability distribution of waveform selection, optimizing multi-dimensional parameters

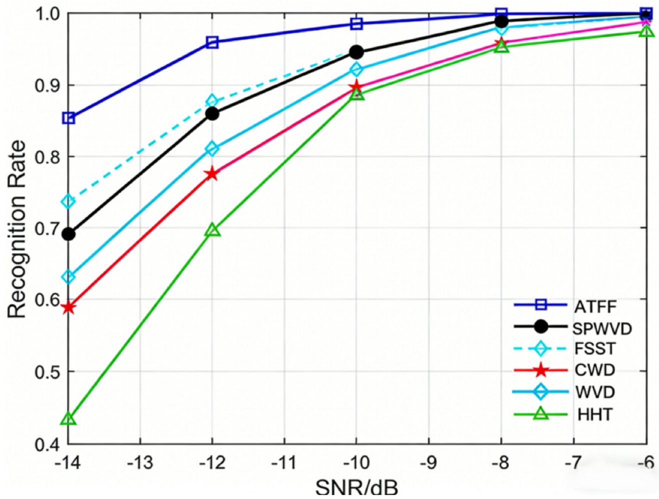

Fig. 3: Recognition rate comparison of different time-frequency analysis methods under various SNR. The horizontal axis is SNR (-14-6 dB), and the vertical axis is recognition rate. Data generated by reinforcement learning, features extracted and calculated for plotting.

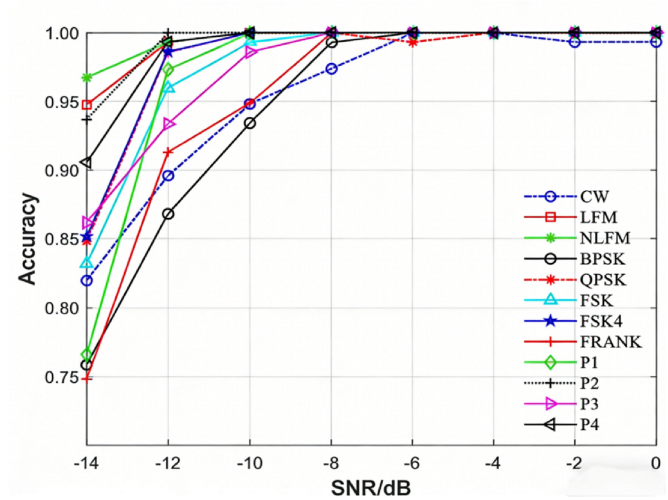

Fig. 4: Recognition rate of radar signals under different SNR. The horizontal axis is SNR (-14-6 dB), and the vertical axis is recognition rate. It compares the recognition rate change trends of 11 radar signals (LFM, BPSK, NPFSK, QPSK, FSK, FSK4, FRANK, P1, P2, P3, P4) with SNR improvement .

such as waveform frequency, bandwidth and phase.

ICM-DQN [18]: Introduces the Intrinsic Curiosity Module (ICM) on the basis of DQN. The external reward comes from anti-jamming effectiveness (SINR), and the internal reward comes from the "exploration value" of waveform selection to the jamming environment.

Subsequently, the proposed method is compared without radar signal recognition prior knowledge, and hierarchical reinforcement learning is compared with one layer removed.

C. Main Experimental Results

1) Radar Unknown Signal Recognition: As shown in Fig.3, the recognition accuracy of all features exceeds 90% at -6 dB SNR; ATFF performs best in low SNR environments, which is 16.38% and 38.94% higher than SPWVD and WVD respectively at -14 dB. The core reasons for the poor performance of other methods are weak anti-noise performance, feature redundancy, high computational cost or sensitivity to noise, and reliance only on single-domain information.

Fig.4 shows the performance of the proposed method in single radar signal classification tasks. NLFM signals achieve the best recognition accuracy, followed by LFM signals. The accuracy of P4 and P2 signals both exceeds 90% at -14 dB SNR; FRRANK, P1 and BPSK signals have relatively weak recognition effects. With SNR increased to -6 dB, the recognition accuracy of the method for all signal types approaches 100%.

As can be seen from Fig.5, the recognition accuracy of ATFF single feature in radar recognition is lower than that of ATFF+PDW dual-stream feature. Their fusion realizes dual coverage of time-domain and parameter-domain features, enabling the model to construct a more complete feature

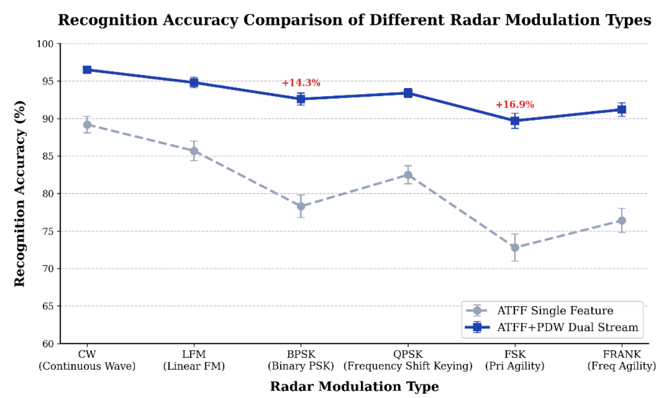

Fig. 5: Recognition accuracy of different radar modulation types: comparison between ATTF single feature and ATTF+PIWD dual feature, showing accuracy improvement of the proposed scheme.

representation that is more in line with the physical nature of radar signals.

2) Jamming Waveform Generation Framework: The reward convergence curve in Fig.6 shows that this result verifies the effectiveness of the proposed radar jamming waveform generation framework. The comparison methods have obvious shortcomings: DQN only relies on external rewards and is prone to local optimum; PPO policy updates are prone to oscillation; ICM-DQN fails to effectively combine the physical characteristics of jamming scenarios.

The results in TableII show that the jamming performance of the proposed architecture is significantly better than the hierarchical architecture without prior knowledge and the single-layer architecture. The core reason is that the proposed method integrates radar-jamming strategy prior knowl-

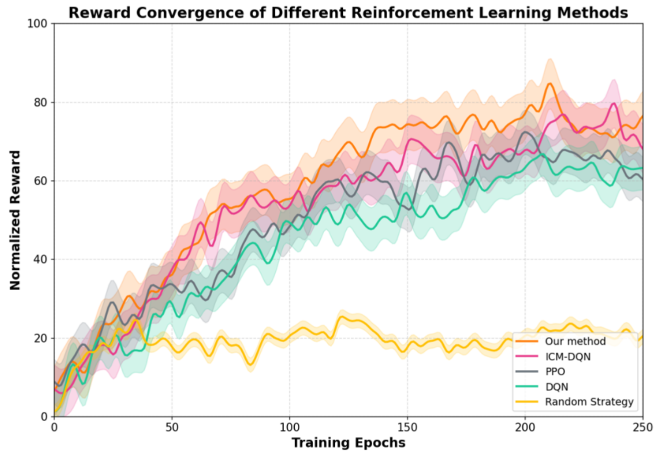

Fig. 6: Reward convergence curves of RL methods: comparison of five strategies (Our Method, ICM-DQN, PPO, DQN, Random Strategy) in stability and learning efficiency.

TABLE II: Jamming performance comparison of reinforcement learning algorithms with different architectures (mean ± standard deviation)

Algorithm Architecture	R_{PSD}	R_{JSR}	R_{auto}
Proposed Method	89.7±2.3	3.8±0.4	0.32±0.08
Hierarchical Architecture without Prior	82.1±3.5	3.1±0.5	0.45±0.11
Single-layer Architecture	73.5±4.2	2.5±0.6	0.61±0.13

edge, and the hierarchical architecture decouples "decision-parameter optimization". The dual-layer architecture without prior knowledge leads to blind strategy decisions due to lack of prior guidance. The single-layer architecture needs to take into account both types of tasks simultaneously, resulting in decision redundancy and target confusion.

IV. Conclusion

Aiming at the problems of poor adaptability and low waveform generation efficiency of traditional radar jamming strategies, this paper proposes an online jamming waveform generation framework based on hierarchical reinforcement learning and cooperative perception. The framework first constructs a time-frequency image and PDW dual-stream network, fuses multi-source time-frequency features to realize accurate recognition of unknown radar signals; then designs a hierarchical reinforcement learning architecture, where the upper level completes jamming strategy decision and the lower level optimizes waveform parameters, forming closed-loop optimization combined with a multi-index composite reward function. Simulation experiments show that compared with traditional methods, the framework has significant advantages in signal recognition accuracy under low SNR and can generate high-matching jamming waveforms, providing an effective technical solution for intelligent electronic countermeasures in complex electromagnetic environments.

References

[1] RAO G N, SASTRY C V S, DIVAKAR N. Trends in electronic warfare[J]. IETE Technical Review, 2003, 20(2):139-150.

[2] WU Y X, GUOCE H, LI W. Waveform design for radar and extended target in the environment of electronic warfare[J]. Journal of Systems Electronics, 2018, 29(1):48-57.

[3] WU Z L, HUANG X X, DU M, et al. Intra-pulse recognition of radar signals via bicubic interpolation WVD[J]. IEEE Transactions on Aerospace and Electronic Systems, 2023, 59(6):8668-8680.

[4] LIAO Y P, JIANG F, WANG J L. Intra-pulse modulation recognition of radar signals based on multi-feature random matching fusion network[J]. The Journal of Supercomputing, 2023, 79(6): 6422-6451.

[5] WU L Y, GUO P C, LIU C, et al. Radar signal modulation type recognition based on attention mechanism enhanced residual networks[J]. Acta Armamentarii, 2023, 44(8):2310-2318.

[6] QU Z Y, LI G, DENG Z A. Radar signal recognition method based on knowledge distillation and attention map[J]. Journal of Electronics & Information Technology, 2022, 44(9):3170-3177.

[7] CAO P Y, YANG C Z, SHI L M, et al. Unknown radar signal processing based on PSO-DBSCAN and SCGAN[J]. Systems Engineering and Electronics, 2022, 44(4):1158-1165.

[8] WU X, DENG Z A. Research on radar signal modulation recognition method based on deep learning[D]. Harbin: Harbin Engineering University, 2023.

[9] XIAO Y H, WANG L, GUO Y X. Radar signal modulation type recognition based on denoising convolutional neural network[J]. Journal of Electronics & Information Technology, 2021, 43(8):2300-2307.

[10] Li J, Guerci J R, Xu L. Signal waveform's optimal under restriction design for active sensing[C]. Fourth IEEE Workshop on Sensor Array and Multichannel Processing, 2006:382–386.

[11] Grossi E, Lops M. Space-time code design for MIMO detection based on Kullback-Leibler divergence[J]. IEEE Transactions on Information Theory, 2012, 58(6):3989–4004.

[12] XIE F, LIU H Y, HU X K, et al. A radar anti-jamming method under multi-jamming scenarios based on deep reinforcement learning in complex domains[J]. Journal of Radars, 2023, 12(6): 1290-1304.

[13] CAI B C, LI H R, ZHANG N M, et al. A cooperative jamming decision-making method based on multi-agent reinforcement learning[J]. Autonomous Intelligent Systems, 2025, 5(3): 1-15.

[14] ZHANG W X, MA D, ZHAO Z K, et al. Design of cognitive jamming decision-making system against MFR based on reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2023, 72(8): 10048-10062.

[15] Vezhnevets A, Osindero S, Schaul T, et al. FeUdal Networks for Hierarchical Reinforcement Learning[J]. arXiv preprint arXiv:1703.01161v2, 2017.

[16] Zhang W, Ma D, Zhao Z, et al. Design of cognitive jamming decision-making system against MFR based on reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2023, 72(8):10048-10062.

[17] Zhang W, Lei X, Zhao Z, et al. A dual-decision-maker frequency domain cooperative jamming method against multi-function radar based on PPO[J]. Digital Signal Processing, 2026, 135:105709.

[18] Zhang Y, et al. Airborne radar anti-jamming waveform design based on deep reinforcement learning[J]. Sensors, 2025, 25(12):4589.