Research on Clutter Suppression Based on Complex-Valued Residual Network and Dynamic Reward Mechanism

Yi CHENG [1,2], Jiaxin LIU [1,2], Junjie SU [1]

[1] School of Control Science and Engineering, Tiangong University, Bin Shui West Road 399, Tianjin, 300387, China
[2] School of Control Science and Engineering, Tianjing Key Laboratory of Intelligent Control for Electrical Equipment, Tiangong University, Bin Shui West Road 399, Tianjin, 300387,China

chengyi@tiangong.edu.cn, ljx1793561421@163.com, 2231251122@tiangong.edu.cn

Submitted August 21, 2025 / Accepted November 18, 2025 / Online first December 22, 2025

Abstract. *As deep reinforcement learning becomes increasingly applied to clutter suppression, existing methods have shown a certain level of adaptability. However, their capabilities in feature representation and generalization remain limited. To address the shortcomings associated with the static reward mechanism-namely, its limited adaptability and slow learning speed-a Complex-Valued Residual Deep Q-Network based on a Dynamic Reward Function (CV-ResDQN-DRF) is proposed in this study. In this method, complex-valued residual units are introduced into the complex-valued neural network framework. Through these units, a complex-valued residual network is constructed to enhance the representational capacity of both amplitude and phase features of signals. Simultaneously, a dynamic reward mechanism is designed, wherein the feedback is adaptively adjusted in real time according to the environmental states and the agent's behavior, thereby accelerating the learning process. Experimental results show that the proposed CV-ResDQN-DRF model achieves an average signal-to-clutter-plus-noise ratio (SCNR) improvement of approximately 2.3 dB on simulated data and 1.8 dB on real measured data, and exhibits a significantly faster convergence speed. These results demonstrate a significant enhancement in clutter suppression performance under complex and non-stationary environments.*

Keywords

Clutter suppression, complex-valued residual network, dynamic reward function, deep reinforcement learning

1. Introduction

In practical operating environments, radar systems receive not only target echoes but are also inevitably affected by clutter interference. This interference originates from natural environments such as the ground, sea surface, and atmospheric conditions [1], [2]. These clutter signals typically exhibit high energy levels, which can easily obscure target echoes, thereby hindering the radar's ability to effectively distinguish targets from clutter. As a result, clutter suppression has become an indispensable technique in radar target detection and remains a critical research focus in the field of radar signal processing.

In the early stages of radar development, Moving Target Indication (MTI) was one of the earliest clutter suppression techniques. It exploits the Doppler frequency difference between stationary clutter and moving targets using delay-line cancellers [3], [4]. However, its performance degrades significantly when detecting slow-moving targets or under Doppler ambiguity, where clutter and target frequencies overlap [5]. To enhance adaptability, Adaptive MTI (AMTI) introduces adaptive filtering based on clutter covariance estimation [6–8]. This method improves clutter rejection but increases computational complexity. Moving Target Detection (MTD) further extends this idea by employing Doppler filter banks and CFAR detection to handle targets with different radial velocities [9], [10]. However, its adaptability remains limited in nonstationary clutter.

With the advancement of matrix decomposition, SVD-based methods have been used to separate clutter and target components through singular value analysis [11–13]. These methods are effective but computationally expensive and perform poorly at low signal-to-clutter ratios. The SVD-FRFT approach helps handle nonstationary clutter but is highly sensitive to transform parameters [14], [15].

In summary, traditional methods such as MTI, AMTI, MTD, and SVD-based approaches form the foundation of radar signal processing. However, their reliance on prior knowledge and limited adaptability reduce their effectiveness in complex and dynamic environments [16], [17]. These limitations motivate the development of deep learning-based clutter suppression techniques.

DOI: 10.13164/re.2026.0001

Against this backdrop, deep reinforcement learning (DRL) offers a novel approach to addressing the aforementioned challenges. In [18], an adaptive clutter suppression method based on DRL was proposed. This approach integrates the environmental perception capabilities of deep learning with the decision-making optimization abilities of reinforcement learning. This enables the system to continuously optimize filter parameters through ongoing interaction with the environment, even in the absence of accurate prior knowledge, thereby achieving dynamic and adaptive clutter suppression. Although existing DRL-based methods have demonstrated strong adaptability in complex environments, there remains room for improvement in terms of feature representation and generalization performance. Furthermore, the use of static reward mechanisms may result in slow learning processes when confronted with rapidly changing clutter characteristics, thereby limiting the agent's adaptation efficiency [19], [20]. To this end, this paper proposes a method named Complex-Valued Residual Deep Q-Network with Dynamic Reward Function (CV-ResDQN-DRF). It builds upon the original DQN architecture and incorporates two key enhancements to improve algorithmic performance. First, a complex-valued residual structure is introduced [21–23], where three ResUnits are stacked to form a Complex-Valued Residual Network (CV-ResNet) that serves as the Q-network in the deep reinforcement learning framework. This network is capable of more effectively extracting amplitude and phase features from complex-valued signals, thereby improving the precision and flexibility of filter parameter tuning. Second, a Dynamic Reward Function (DRF) is designed to adaptively adjust the reward based on the current environmental state and the agent's actions. This mechanism more accurately reflects the actual impact of filter adjustments and accelerates the convergence of the learning process. These two enhancements significantly improve the convergence speed and adaptability of deep reinforcement learning in complex clutter environments, providing an effective solution for achieving efficient clutter suppression.

2. Signal Model

This study adopts the Linear Frequency Modulation (LFM) radar signal model proposed in [18] to ensure the comparability of subsequent algorithms under consistent signal conditions. It is assumed that the radar system transmits an LFM signal with a pulse width of T_p, a pulse repetition interval (PRI) of T_r, and a modulation bandwidth of B. This configuration results in a chirp rate $\mu = B/T_\mathrm{p}$, and a carrier frequency of f_c. Under these assumptions, the transmitted radar signal in the n-th pulse period can be expressed as:

$$S_n(t) = \mathrm{rect}\left[\frac{t - nT_\mathrm{r}}{T_\mathrm{p}}\right] \exp\left[\mathrm{j}2\pi\left(f_\mathrm{c}(t - nT_\mathrm{r}) + \frac{1}{2}\mu(t - nT_\mathrm{r})^2\right)\right] \tag{1}$$

where, $\mathrm{rect}\left[\frac{t-nT_\mathrm{r}}{T_\mathrm{p}}\right] = 1$, when $|t| \le \frac{T_\mathrm{p}}{2} + nT_\mathrm{r}$, assuming $t = 0$, the initial distance between the target and the radar is R_0, and the target is moving with a constant velocity v. The echo delay of the target at this moment can be expressed as:

$$\tau = \frac{2(R_0 - vnT_\mathrm{r})}{c} \tag{2}$$

where c denotes the speed of light. Therefore, the echo signal of a single target in the n-th pulse period can be written in (3), where *c(t)* represents the clutter, and *n(t)* denotes the Gaussian white noise at the receiver. Based on (3), a matrix model of the echo signal can be established in both the fast time and slow time dimensions. After performing pulse compression along the fast-time dimension, the echo signal matrix is constructed as shown in (4). Here, $\mathbf{X}_\mathrm{c}$ denotes the echo data matrix with dimensions $N \times L$, where N represents the number of accumulated pulses and L is the number of range cells. The symbol j denotes the imaginary unit, and f_d represents the Doppler frequency of the target echo.

3. Adaptive Clutter Intelligent Suppression Method Based on CV-ResDQN-DRF

This section provides a detailed introduction to an adaptive clutter suppression method, CV-ResDQN-DRF, based on the integration of complex-valued residual networks and deep reinforcement learning. Built upon the Deep Q-Network (DQN) framework, the method incorporates complex-valued residual structures to enhance the extraction of amplitude and phase features from radar complex signals. In addition, it introduces a dynamic reward function to improve training efficiency and policy stability. By organically combining feature enhancement, policy learning, and reward guidance, the proposed method enables efficient target recognition and adaptive suppression of background interference in complex clutter environments. This section focuses on the overall architecture of the method, systematically describing the design of the DQN network, the complex-valued residual network, and their integration with the dynamic reward mechanism.

$$x_{nr}(t) = \mathrm{rect}\left(\frac{t - nT_\mathrm{r} - \tau}{T_\mathrm{p}}\right) \exp\left[\mathrm{j}2\pi\left(f_\mathrm{c}(t - nT_\mathrm{r} - \tau) + \frac{1}{2}\mu(t - nT_\mathrm{r} - \tau)^2\right)\right] + c(t) + n(t), \tag{3}$$

$$\mathbf{X}_\mathrm{c} = \begin{bmatrix} x_1(1)\exp(\mathrm{j}f_\mathrm{d}T_\mathrm{r}) & x_1(2)\exp(\mathrm{j}f_\mathrm{d}T_\mathrm{r}) & \cdots & x_1(L)\exp(\mathrm{j}f_\mathrm{d}T_\mathrm{r}) \\ x_2(1)\exp(\mathrm{j}2f_\mathrm{d}T_\mathrm{r}) & x_2(2)\exp(\mathrm{j}2f_\mathrm{d}T_\mathrm{r}) & \cdots & x_2(L)\exp(\mathrm{j}2f_\mathrm{d}T_\mathrm{r}) \\ \vdots & \vdots & \ddots & \vdots \\ x_N(1)\exp(\mathrm{j}Nf_\mathrm{d}T_\mathrm{r}) & x_N(2)\exp(\mathrm{j}Nf_\mathrm{d}T_\mathrm{r}) & \cdots & x_N(L)\exp(\mathrm{j}Nf_\mathrm{d}T_\mathrm{r}) \end{bmatrix} \tag{4}$$

3.1 Adaptive Clutter Suppression Mechanism Based on Deep Q-Network

1) DQN Network: The Deep Q-Network (DQN) is a representative value-based algorithm in deep reinforcement learning, which approximates the optimal state-action value function by leveraging the deep neural networks. This approach addresses the limitation of traditional Q-learning, which becomes infeasible in the high-dimensional state spaces due to the inability to store the Q-table. The core idea of DQN lies in integrating policy learning from reinforcement learning with the feature extraction capabilities of deep learning, thereby enabling the agent to make adaptive decisions in complex environments.

As shown in Fig. 1, the overall architecture of the DQN network proposed in [18] for radar clutter suppression consists of four main components: the DQN module, the agent, the trainable filter module, and the environment interaction mechanism. The detailed process is as follows:

At the beginning of training, the filter initialization module configures the filter based on the prior knowledge (such as the spectral center and the spectral width), endowing it with a basic capability for clutter suppression. Subsequently, throughout the training process, the training model and the Q-network jointly form a learnable decision-making unit. The Q-network extracts features from the current state and predicts Q values via a deep neural network, while the filter parameters are continuously updated during training. This enables the system to achieve adaptive suppression of clutter signals. During each interaction, the estimated Q-network predicts the Q values for all possible actions a_t based on the current input state s_t. These Q values represent the expected long-term cumulative rewards associated with taking each possible action in the current state. The agent subsequently follows the ϵ-greedy strategy. In this process, it explores the action space by selecting a random action with a probability of ϵ (exploration rate), and exploits the learned policy by choosing the action with the highest predicted Q-value with a probability of $1-\epsilon$. Ultimately, the action associated with the maximum Q-value is selected, and the corresponding Q-value $Q(s_t, a_t; \theta)$ is produced as the output. The next state s_{t+1} is fed into the target value network, where the Q-values $\hat{Q}\left(s_{t+1}, a'_{t+1}; \theta^-\right)$ corresponding to all possible actions a'_{t+1} are estimated. Among these, the maximum Q-value is selected, which is considered to represent the maximum expected cumulative reward attainable by the agent from state s_{t+1}. This maximum value is then utilized as the target in the learning update process.

During each training iteration, the Q-network updates the parameters of the estimation network by computing the difference between the predicted Q-value and the target Q-value. This difference is denoted as the loss $L = \left[r_t + \gamma \max_{a'} \hat{Q}\left(s_{t+1}, a'_{t+1}; \theta^-\right) - Q\left(s_t, a_t; \theta\right)\right]$, so as to make the predicted Q-value closer to the true Q-value. The parameters of the target network are identical to those of the Q-network but are updated at a lower frequency. By periodically updating the parameters of the target network, instability in the parameters during training is reduced, thereby improving training stability and convergence speed.

2) State Space Design: In this study, the state space is defined as the set of feature information received by the agent from the environment. Specifically, the agent's observed state consists of the echo signals at its current range cell. Assuming that the agent is located at the l-th range cell and that N pulses have been accumulated along the fast-time axis at this cell, the state can be represented as

$$S = \left[x_{1,l}, x_{2,l}, ..., x_{N,l}\right] \tag{5}$$

where, N denotes the total number of received pulses, and l represents the index of the agent's range cell, satisfying $1 \le l \le L$. At the beginning of each training episode, the environment state is reset, enabling the agent to adapt to varying scenarios and conditions.

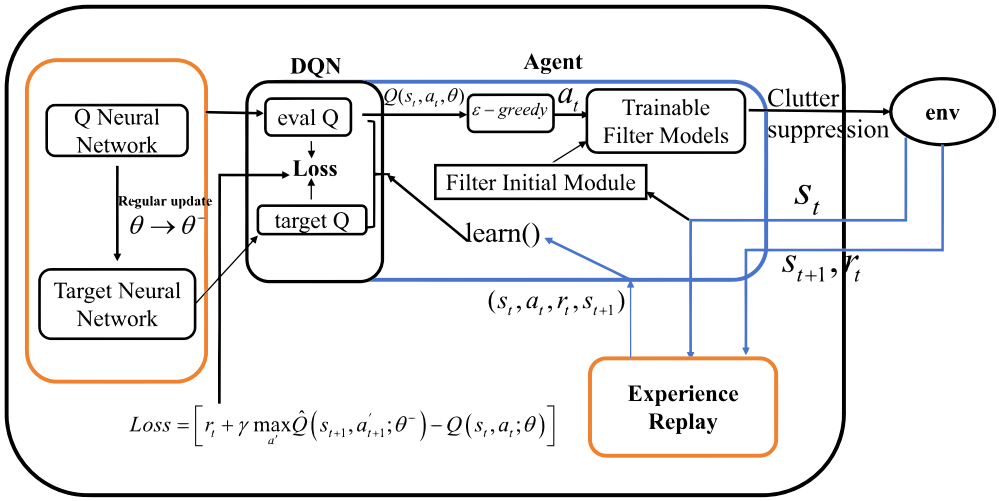


Fig. 1. The network parameter update process of DQN.

3) Action Space Design: In this study, the action space is designed by fully incorporating the impact of filter parameters on clutter suppression performance. Based on this, an action set A is constructed as follows:

$$A = \left[f_0 + 0.1, f_0 - 0.1, \delta_f + 0.1, \delta_f - 0.1, M + 1, M - 1\right] \quad (6)$$

where f_0 denotes the center frequency of the filter, δ_f represents the spectral width, and M is the filter order. This action set enables the agent to dynamically control the frequency-domain characteristics by fine-tuning key filtering parameters, thereby achieving improved suppression performance under complex clutter conditions. Adjustments to f_0 and δ_f allow the filter to better match the spectral distribution of the target and clutter, while variations in M affect the frequency resolution and smoothing capability of the filter. To ensure filter stability and practical feasibility, constraints are imposed on parameter ranges, with $\delta_f \in [0.1, 120]$ and $M \in [3, 16]$. This parameter-level action space design provides the agent with a compact yet flexible set of options to optimize filter configurations, enhancing its adaptability to diverse radar clutter scenarios.

During the training process, each agent autonomously selects and updates its dedicated filter parameters based on the features of its corresponding range cell. This approach ensures strong adaptability of the filter to varying environmental characteristics. In the initial phase, filter parameters are initialized through a predefined module, after which they are continuously adjusted through iterative updates within the deep reinforcement learning framework. This strategy not only enables dynamic adaptation to the optimal solution under different observation states, but also maintains the convergence and stability of the parameters. As a result, targeted clutter suppression is achieved while effectively preserving the target signals.

3.2 Complex-Valued Residual Network

The complex-valued residual neural network is composed of a series of stacked complex-valued residual units, each of which is represented as follows:

$$y_1 = h(x_l) + F(x_l, W_1) \quad (7)$$

where $h(x_l)$ denotes the identity mapping of the input x_l, which serves as both the input and output of the l-th residual unit. The term $F(x_l, W_l)$ represents the residual feature extracted through a series of nonlinear transformations, including complex-valued convolution, complex batch normalization (CV-BN), and complex activation functions.

Within each complex residual unit, the input x_l is first processed by a complex batch normalization layer followed by a complex activation function (such as complex-valued ReLU). The output is then passed through a complex convolutional layer to extract feature representations. This process is repeated to obtain a second set of transformed features. Finally, the residual feature $F(x_l, W_l)$ is added to the original input x_l via a skip connection, thereby enhancing the network's feature representation capability. The architecture of the residual module is illustrated in Fig. 2, where each complex residual block is composed of multiple combinations of complex batch normalization (CV-BN), complex ReLU activation, and complex convolution layers.

Building upon the Deep Q-Network (DQN) architecture proposed in [18], three complex-valued residual units are further stacked in this study to enhance the network's capacity for deep feature extraction and robustness. Each complex residual unit consists of two consecutive sets of operations: complex batch normalization, complex activation function, and complex convolution. Through residual connections, the input features are directly added to the output features, thereby effectively improving the network's ability to model complex signals. Figure 3 illustrates the architecture of the Complex-Valued Residual Deep Q-Network (CV-ResDQN), in which complex-valued residual units are integrated to enhance the network's representational capability.

3.3 Dynamic Reward Function

In [19], the Signal-to-Clutter-plus-Noise Ratio (SCNR) of the echo data matrix and the clutter attenuation at the agent's current range cell are selected as evaluation metrics, as defined in (8) and (9). The SCNR serves to quantify the prominence of the target signal under complex clutter and noise conditions, thereby effectively reflecting the overall suppression performance and signal quality. Meanwhile, the clutter attenuation at the agent's range cell captures the actual variation in local clutter energy, indicating the algorithm's real-time adaptability and its ability to suppress interference in critical range cells. The combination of these two metrics provides a comprehensive assessment from both global and local perspectives. It also enables a more accurate analysis of the reinforcement learning strategy's responsiveness to clutter characteristics and its robustness in decision-making. The reward function is defined in (10).

$$SCNR = 10\log_{10}(P_s) - 10\log_{10}(P_c), \quad (8)$$

$$CA_l^{(\text{agent\_id})} = \frac{C_i^{(\text{agent\_id})}}{C_o^{(\text{agent\_id})}}. \quad (9)$$

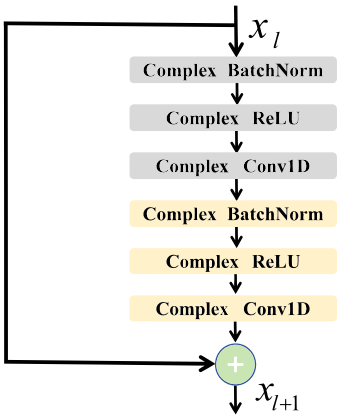


Fig. 2. Complex-valued residual module.

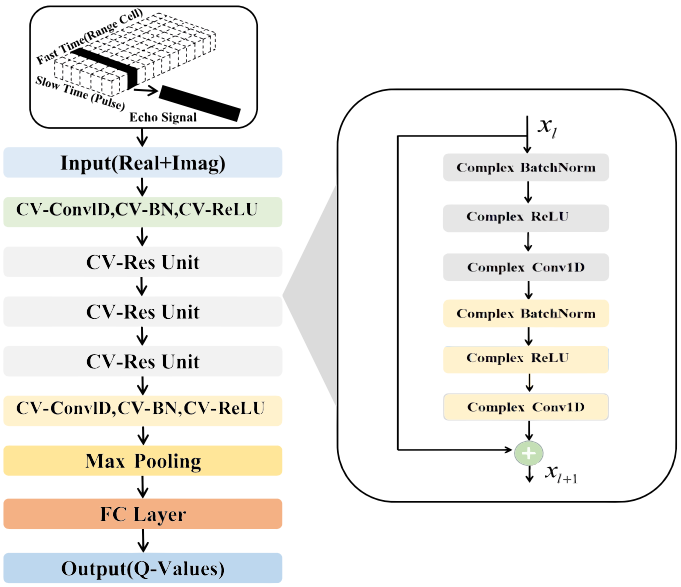


Fig. 3. Schematic of the complex-valued residual deep Q-network.

$$R = \begin{cases} 100, & \text{if } SCNR_t > SCNR_{\text{peak}} \text{ and } CA_t^{(\text{agent\_id})} > CA_{\text{peak}}^{(\text{agent\_id})} \\ 30, & \text{if } SCNR_t \leq SCNR_{\text{peak}} \text{ and } CA_t^{(\text{agent\_id})} > CA_{\text{peak}}^{(\text{agent\_id})} \\ -1, & \text{if } SCNR_t > SCNR_{\text{peak}} \text{ and } CA_t^{(\text{agent\_id})} \leq CA_{\text{peak}}^{(\text{agent\_id})} \\ -10, & \text{if } SCNR_t \leq SCNR_{\text{peak}} \text{ and } CA_t^{(\text{agent\_id})} \leq CA_{\text{peak}}^{(\text{agent\_id})} \end{cases} \tag{10}$$

In (8), P_{s} denotes the average power of the target signal, P_{c} represents the average power of the clutter. In (9), $CA_t^{(\text{agent\_id})}$ refers to the clutter attenuation at the specified range cell where the agent is located at time step t. Specifically, $C_{\text{i}}^{(\text{agent\_id})}$ denotes the clutter power before suppression, and $C_{\text{o}}^{(\text{agent\_id})}$ denotes the clutter power after suppression at the corresponding range cell.

The core idea lies in discretizing the reward or penalty based on the absolute relationship between the current performance and the historically optimal value. Although this method can initially guide the agent in optimizing filter parameters, it presents the following limitations:

- Sparse reward trap: Rewards are allocated based solely on a binary judgment of whether the current value reaches the historical maximum. This leads to sparse reward signals during the early stages of training and results in inefficient convergence.
- Conflict between global and local objectives: When optimization objectives between SCNR (a global metric) and CA (a local metric) conflict, the fixed-weight mechanism fails to dynamically balance the two, limiting the adaptability of the strategy.

To this end, a Dynamic Reward Function (DRF) is proposed, with its mathematical formulation defined as follows:

$$R_t = \alpha \cdot \Delta\text{CA}_t + \beta \cdot \Delta ISCNR_t \tag{11}$$

where $\Delta CA_t = CA_t^{(\text{agent\_id})} - CA_{\text{origin}}^{(\text{agent\_id})}$ denotes the clutter attenuation achieved by the agent at the specified range cell at time step t, and $\Delta ISCNR_t = SCNR_t - SCNR_{t-1}$ captures the incremental improvement in global clutter suppression between consecutive time steps, thereby enhancing the temporal continuity of the reward signal. The dynamic weights α and β are adaptively adjusted according to the training progress:

$$\begin{cases} \alpha = 0.8 \cdot \left(1 - \frac{t}{T}\right) + 0.2 \\ \beta = 0.2 \cdot \frac{t}{T} + 0.1 \end{cases} \tag{12}$$

where T represents the total number of training steps. The weighting coefficient α is linearly decayed from an initial value of 1.0 (at $t = 0$) to 0.2 (at $t = T$). This decay guides the agent to prioritize the optimization of local clutter attenuation (CA) during the early training phase. As training progresses, the focus gradually shifts toward improving the global signal-to-clutter-plus-noise ratio (ISCNR). This strategy alleviates the multi-objective conflict inherent in traditional fixed-weight reward mechanisms. By computing the instantaneous improvement of ΔCA_t and $\Delta ISCNR_t$ instead of relying on absolute thresholds, the function delivers a continuous reward signal. Compared to the discrete reward scheme in (10), this approach mitigates reward sparsity and significantly accelerates policy exploration efficiency.

Furthermore, the interaction among the Q-network, the environment, and the dynamic reward mechanism fundamentally improves the training dynamics of the proposed method.

During each training iteration, the agent perceives the current clutter environment through the Q-network, updates its filtering policy, and then receives real-time feedback from the dynamic reward function. This closed-loop interaction allows the network to continuously refine its policy in response to non-stationary clutter distributions. The dynamic reward function provides dense and performance-sensitive feedback, which effectively alleviates the sparse-reward issue typical in conventional DQN training. Consequently, the agent can receive meaningful learning signals even in the early training stages, promoting faster convergence and reducing instability.

Moreover, the adaptive weighting of the local (CA) and global (SCNR) reward terms enables a progressive shift of optimization focus-from local clutter attenuation in the early stage to overall signal quality enhancement in later stages. This staged optimization strategy smooths parameter updates and prevents overfitting to transient clutter variations. Through this coordinated mechanism, the CV-ResDQN-DRF framework achieves faster policy convergence, improved learning stability, and enhanced generalization to unseen clutter environments. These improvements collectively contribute to the superior training dynamics observed in the proposed method.

3.4 CV-ResDQN-DRF Algorithm Model

Based on the definitions of the state space, action space, and dynamic reward function, the parameter decision-making process of the filter is illustrated. The filter is built upon the complex-valued residual Deep Q-Network. The corresponding procedure is presented in Algorithm 1.

Algorithm 1. Filter parameter decision based on CV-ResDQN-DRF.

1: Initialize the evaluation Q-network with random parameters (adopting a complex-valued residual structure), and set the target Q-network parameters to $\theta^- = \theta$.
2: Initialize the experience replay buffer D; define the initial parameter policy of the filter, and configure key hyperparameters including the discount factor γ, learning rate α, greedy factor ϵ, target network update frequency τ, and other related parameters.
3: Initialize the dynamic reward function's weight adjustment strategy by setting the initial values $\alpha(0)$ and $\beta(0)$, and define their update rules based on the training step count.
4: Deploy a number of agents equal to the number of range cells in the radar echo matrix, each responsible for optimizing the filter parameters within its respective range cell.
5: **for** episode = $1, 2, \ldots, M$ **do**
6: Initialize the environment and obtain the initial state s_t.
7: Each agent selects action a_t according to the ϵ-greedy policy, exploring with probability ϵ, or exploiting by choosing the action with the highest current Q-value with probability $1 - \epsilon$.
8: Execute action a_t to update the corresponding filter parameters, perform clutter suppression on the echo signal, and obtain the next state s_{t+1}.
9: Calculate the reward based on the current filtering performance:

$$r_t = \alpha(t) \cdot \Delta CA_t + \beta(t) \cdot \Delta ISCNR_t$$

10: Store the sample (s_t, a_t, r_t, s_{t+1}) into the experience replay buffer D.
11: Randomly sample a batch of transitions (s, a, r, s') from D, feed them into both the evaluation and target networks, and compute the target:

$$y = r + \gamma \cdot \max_{a'} Q(s', a'; \theta^-)$$

$$L(\theta) = [y - Q(s, a; \theta)]^2$$

12: Update the evaluation Q-network parameters θ via backpropagation.
13: Every τ steps, update the target network parameters by setting $\theta^- \leftarrow \theta$.
14: **end for**

4. Experiments and Results Analysis

The overall performance of the conventional DQN-based clutter suppression method and the proposed Complex-Valued Residual Deep Q-Network with Dynamic Reward Function (CV-ResDQN-DRF) was comparatively evaluated in this study. To ensure fairness in the comparison, the clutter data and target range cell configurations from [18] were adopted in the simulation environment. The output effectiveness of both methods was assessed in terms of signal-to-clutter-plus-noise ratio (SCNR), and the simulation network parameters are summarized in Tab. 1. All experiments were conducted on a workstation equipped with an NVIDIA RTX 4090 GPU (24 GB VRAM), an Intel Core i9-13900K CPU, and 64 GB of RAM. The models were implemented in Python using the PyTorch 2.1 framework and CUDA 12.1.

Model parameters	Values
Batch size	100
Learning rate α	0.001
Discount factor γ	0.99
Experience replay buffer capacity	10000
Initial greedy factor ε	0.9
Greedy factor attenuation coefficient	0.1
Target value network update cycle τ	5

Tab. 1. Network simulation parameters.

Parameters	Values
Pulse repetition frequency [Hz]	2000
Pulses number	64
Range cell number	100
Target range cell	19, 31, 53, 74
Doppler frequency [Hz]	440
Target power [dB]	45

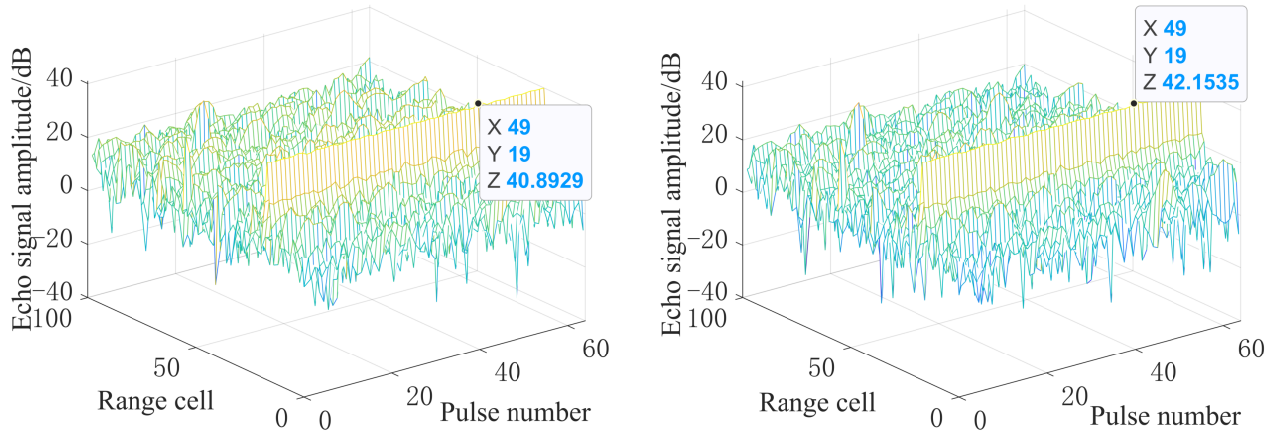
Tab. 2. Simulated radar system parameters.

Parameters	Values
Pulse repetition frequency [Hz]	2000
Pulses number	64
Clutter power [dB]	55
Average noise power [dB]	0
Amplitude type	K distribution
Range cell of simulated Clutter Region 1	1–50
Range cell of simulated Clutter Region 2	51–100
Spectral center of simulated Clutter Region 1 [Hz]	0

Tab. 3. Simulated clutter parameters.

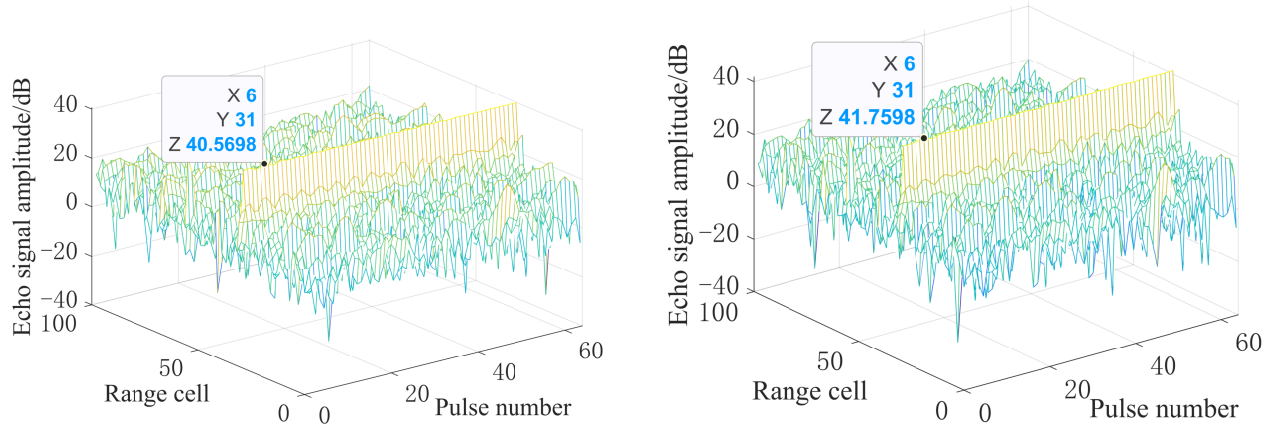
4.1 Analysis of Suppression Serformance Under Simulated Clutter Data

To verify the effectiveness of the proposed method, this section employs the same simulated clutter dataset as used in [18], with four representative range cells selected for evaluation (19, 31, 53, and 74). The simulated target parameters and clutter parameters are presented in Tabs. 2 and 3.



(a) DQN method (b) CV-ResDQN-DRF method

Fig. 4. Three-dimensional view after clutter suppression by two methods when the target is located in the 19th range cell.

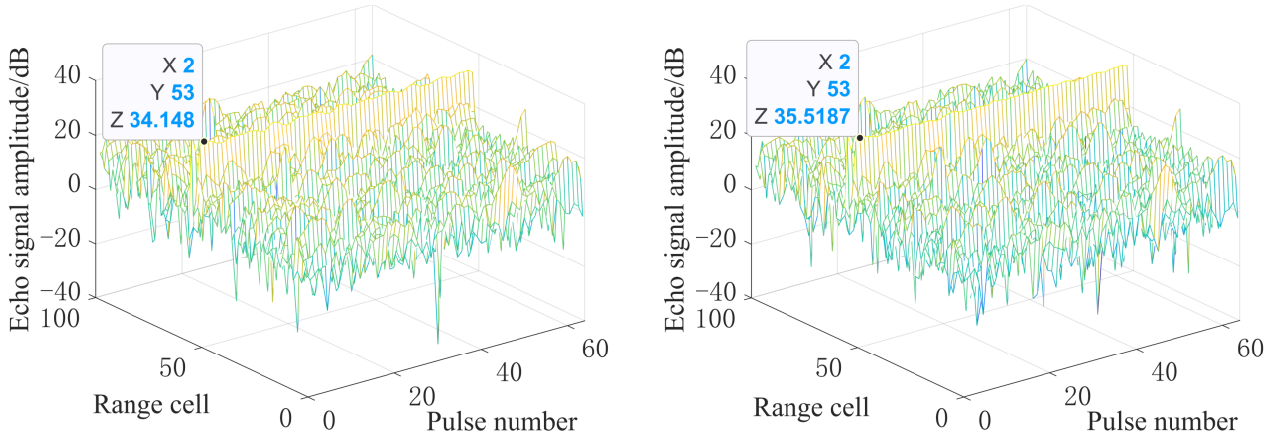


(a) DQN method (b) CV-ResDQN-DRF method

Fig. 5. Three-dimensional view after clutter suppression by two methods when the target is located in the 31th range cell.

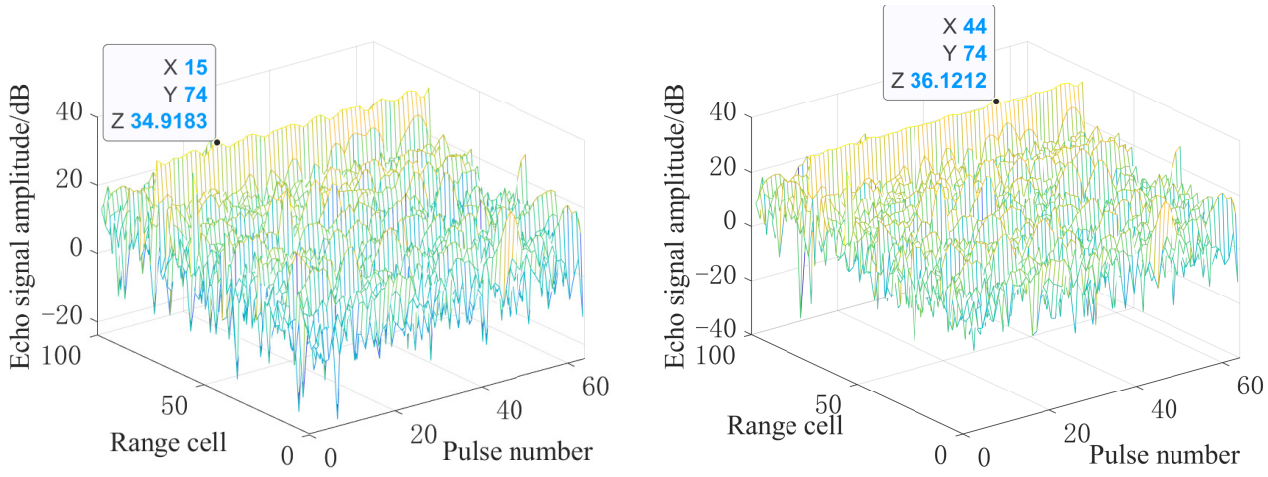
The experimental results are illustrated in Figs. 4–7. These figures present the clutter suppression outcomes at range cells 19, 31, 53, and 74 using the two respective methods. In each figure, the pulse with the highest target signal power within the corresponding target-containing range cell is highlighted. This highlights the differences in target signal preservation performance between the two algorithms.

As shown in Figs. 4 and 5, the target is located at range cells 19 and 31 respectively, both of which are within Clutter Region 1, where the clutter center frequency is 0 Hz. In this region, both algorithms effectively suppress the clutter. However, the performance is significantly improved with the introduction of the CV-ResDQN-DRF enhancement. Specifically, in these two range cells, the highest target signal power after clutter suppression by the DQN method is 40.8829 dB and 40.5698 dB, respectively, while the improved algorithm achieves 42.1535 dB and 41.7598 dB, respectively.Given that the original target power is 45 dB, this indicates that the proposed improved algorithm results in less target signal loss and better signal preservation. The target signal power loss is reduced from 4.1171 dB and 4.4302 dB to 2.8465 dB and 3.2402 dB, respectively, further validating the advantages of the improved algorithm in both clutter suppression effectiveness and target signal retention. As shown in Figs. 6 and7, the target is located at range cells 53 and 74, respectively. Both cells are within Clutter Region 2, where the clutter center frequency is 150 Hz. In this region, both algorithms effectively suppress the high-frequency clutter. However, the CV-ResDQN-DRF method still outperforms the traditional approach under these high-frequency clutter conditions. Specifically, the highest target signal power after processing by the DQN method is 34.148 dB and 34.9183 dB (with an initial target power of 45 dB), corresponding to power losses of 10.8520 dB and 10.0817 dB. In contrast, the proposed method leverages refined feature extraction via the complex-valued residual network combined with the dynamic reward function. It raises the target power to 35.5187 dB and 36.1212 dB. The power losses are reduced to 9.4813 dB and 8.8788 dB, respectively. These results demonstrate that the improved algorithm maintains superior signal preservation capability compared to the conventional method in high-frequency clutter scenarios.



(a) DQN method (b) CV-ResDQN-DRF method

Fig. 6. Three-dimensional view after clutter suppression by two methods when the target is located in the 53th range cell.



(a) DQN method (b) CV-ResDQN-DRF method

Fig. 7. Three-dimensional view after clutter suppression by two methods when the target is located in the 74th range cell.

Based on the analysis of signal power loss, Table 4 further quantifies the global performance improvement of the proposed method in terms of signal-to-clutter-plus-noise ratio (SCNR). As shown, for four representative range cells (19, 31, 53, and 74), the SCNR outputs of the conventional DQNmethod described in Sec. 3 are 26.7871 dB, 26.7136 dB, 20.4597 dB and 21.3220 dB, respectively, whereas the proposed method (CV-ResDQN-DRF) improves these values to 29.0028 dB, 28.9626 dB, 23.0925 dB and 23.6698 dB. These correspond to gains of 2.2157 dB, 2.2490 dB, 2.6329 dB and 2.3478 dB over the conventional DQN method. Notably, the improvement in Clutter Region 2 is more pronounced. This indicates that under high-frequency clutter conditions, the complex-valued residual network, through refined feature extraction, enhances local clutter suppression performance. This enables the algorithm to maintain strong signal preservation and clutter attenuation capabilities when confronting high-frequency Doppler shifts and broad spectral widths. The independent contribution of the dynamic reward function module will be further analyzed in subsequent ablation studies.

Target range cell	**19**	**31**	**53**	**74**
SCNR output by DQN clutter suppression method [dB]	26.7871	26.7136	20.4597	21.3220
SCNR output by CV-ResDQN-DRF method [dB]	29.0028	28.9626	23.0925	23.6698
SCNR gain improvement by CV-ResDQN-DRF method [dB]	2.2157	2.2490	2.6329	2.3478

Tab. 4. The results of clutter suppression when the target is located at different range cells.

To verify the statistical reliability of the proposed method, each target range cell was independently tested five times under different random seeds. The mean value and 95% confidence interval (CI) of the SCNR obtained by the CV-ResDQN-DRF method were calculated to evaluate the model's convergence stability and reproducibility. The DQN method was adopted from the literature as a fixed reference, and therefore no confidence interval was computed for it.

Target range cell	19	31	53	74
CV-ResDQN-DRF mean SCNR [dB]	29.2707	28.9641	22.9507	23.5781
CV-ResDQN-DRF 95% CI [dB]	29.27 ± 0.45	28.96 ± 0.17	22.95 ± 0.34	23.58 ± 0.22
Mean SCNR gain [dB]	2.49	2.25	2.49	2.26
SCNR gain 95% CI [dB]	2.49 ± 0.45	2.25 ± 0.17	2.49 ± 0.34	2.26 ± 0.22

Tab. 5. Mean SCNR and SCNR gain (with 95% CI) for different target cells.

Table 5 presents the SCNR outputs and corresponding gains of the DQN baseline and the proposed CV-ResDQN-DRF method across four target range cells. The results show that the CV-ResDQN-DRF consistently achieves significant SCNR improvements in all target cells. Moreover, the narrow confidence intervals indicate that the proposed method exhibits strong stability and reproducibility.

4.2 Analysis of Suppression Effects Under Measured Clutter Data

In the analysis of suppression performance using measured clutter data presented in this section, the target range cells and clutter region divisions from the measured sea clutter dataset in [18] were adopted. The measured sea clutter dataset, along with the X-band radar parameters, is summarized in Tab. 6. The suppression capabilities of the original DQN method and the proposed improved method in the actual sea clutter environment are compared. The experimental results are presented in Tab. 7. Figure 8 illustrates the range profiles after clutter suppression in different clutter regions. Specifically, these include the strong clutter region at range cell 136, the mixed clutter-noise region at range cell 1273, and the noise region at range cell 2337. This demonstrates the differing abilities of the two algorithms in preserving the target signal.

As shown in Fig. 8, three representative range cells (136, 1273, and 2337) were selected, and the target signal power of a specific pulse within each cell was compared to illustrate the differences in target signal preservation capabilities between the two methods. For range cell 136, the target signal power after processing by the DQN method is 82.0145 dB, whereas the improved method increases it to 83.7004 dB, indicating an enhancement of 1.6859 dB and better preservation of the target signal. At range cell 1273, the target signal power after clutter suppression by the DQN method is 82.7569 dB, which is improved to 84.0126 dB by the proposed method, yielding a gain of 1.2557 dB. At range cell 2337, the DQN method outputs a signal power of 81.6633 dB, while the improved method achieves 82.6502 dB, representing an increase of approximately 0.9869 dB. Overall, the differences in signal power across the range cells are relatively small. However, the improved method consistently demonstrates superior target signal preservation capabilities in all selected range cells. More pronounced gains are observed at the nearer range cell (136) and the mid-range cell (1273). This indicates that the network enhanced with CV-ResNet and DRF can more effectively reduce target signal loss during clutter suppression, thereby improving the overall clutter suppression performance and robustness of the system.

Parameters	Values
Sea clutter data	20210106155330-01-staring
Operating frequency band	X
Carrier frequency [GHz]	9.3–9.5
Polarization mode	HH
Pulse repetition frequency [Hz]	1697
Scan bandwidth [MHz]	25
Pulse width [μs]	3
Azimuth angle [°]	42.18
Sea state	3 4
Distance resolution [m]	61
Antenna working mode	Gaze scan

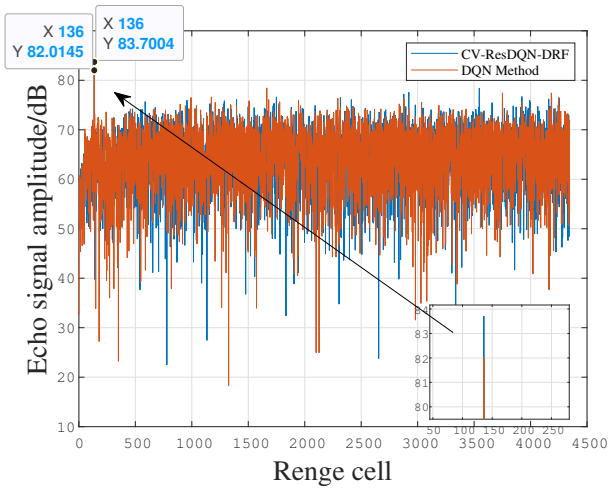
Tab. 6. X-band experimental radar parameters.

Target range cell	136	705	1273
SCNR output by DQN method [dB]	14.0699	13.1362	12.8506
SCNR output by CV-ResDQN-DRF method [dB]	16.1761	15.4295	14.5287
SCNR gain improvement by CV-ResDQN-DRF method [dB]	2.1062	2.2933	1.6781
Target range cell	**1618**	**2337**	**4153**
SCNR output by DQN method [dB]	12.7624	11.3966	11.1626
SCNR output by CV-ResDQN-DRF method [dB]	14.3490	13.0427	12.3600
SCNR gain improvement by CV-ResDQN-DRF method [dB]	1.5866	1.6461	1.1974

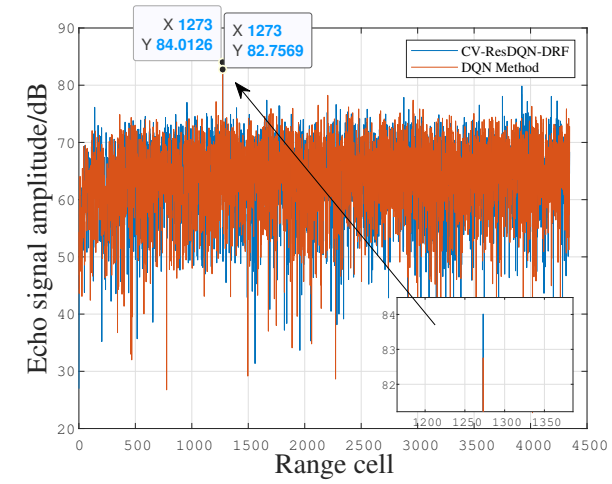
Tab. 7. The results of clutter suppression when the target is located at different range cells.

As shown in Tab. 7, SCNR improvements are achieved by the proposed method compared to the DQN method at the target range cells 136, 705, 1273, 1618, 2337 and 4153. Notably, in the strong clutter regions (range cells 136 and 705), gains of 2.1062 dB and 2.2933 dB are observed, respectively. This improvement is attributed to the enhanced capability of the proposed algorithm (based on CV-ResDQN-DRF) to extract sea clutter features more effectively under strong sea clutter conditions. It creates deeper and wider notches at the clutter frequency center, thereby reducing target signal loss during filtering and significantly increasing the output signal-to-clutter-plus-noise ratio. In the mixed clutter-noise and noise regions, the distinction between clutter and target signals is reduced, limiting the enhancement achievable even with the refined filtering approach. Consequently, the gains decrease to 1.68 dB, 1.59 dB, 1.65 dB, and 1.20 dB, respectively. Although the proposed method still outperforms the DQN method, the SCNR improvement is relatively modest. This phenomenon indicates that while the

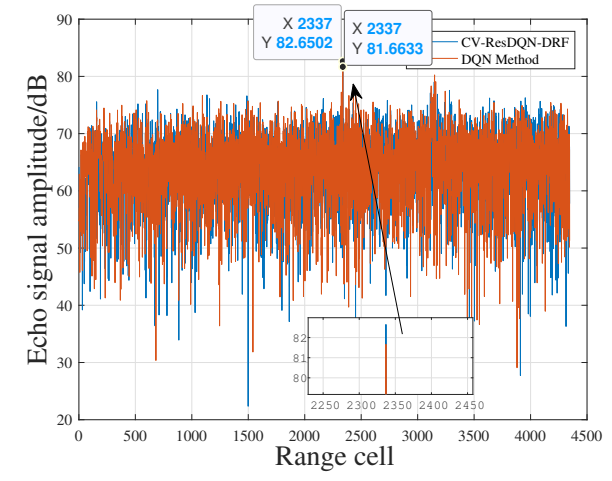
CV-ResDQN-DRF method demonstrates clear advantages in preserving target signal integrity and effectively suppressing clutter, its performance gains remain constrained by factors such as environmental complexity.



(a) Target located at range cell 136



(b) Target located at range cell 1273



(c) Target located at range cell 2337

Fig. 8. Range profiles after clutter suppression with the target located at different range cells.

Similarly, five independent experiments were conducted with different initialization seeds. For each target range cell, the SCNR and its corresponding 95% confidence interval (CI) were calculated to quantify the reliability of the results. The statistical outcomes are presented in Tab. 9. It can be observed that the proposed method maintains consistent SCNR improvements over the DQN baseline across all target range cells. Moreover, the narrow confidence intervals indicate minimal variation among different runs, demonstrating the model's robustness, reproducibility, and stable convergence in real clutter environments.

4.3 Performance Impact Analysis of Each Module Under Simulated Clutter Data

This section experimentally verifies the individual performance contributions of each module (CV-ResNet and DRF). Building upon the baseline DQN method, the DRF and the CV-ResNet are introduced separately and then combined. This allows analysis of their respective impacts on SCNR improvement. This quantifies the gain contributions of each module and confirms their effectiveness. The simulated clutter dataset employed is the same as that used in [18] and in the aforementioned proposed method, with simulation parameters detailed in Tabs. 2 and 3. Four representative range cells (19, 31, 53, and 74), consistent with those in Sec. 4.1, are selected for evaluation. Table 8 presents the comparison of output SCNR across different range cells under various module configurations. It illustrates the clutter suppression performance of the DQN, DQN-DRF, CV-ResDQN, and CV-ResDQN-DRF methods at each range cell.

Compared to the baseline DQN method, all variants-whether incorporating the dynamic reward function, the complex-valued residual structure, or both-achieve improvements in SCNR. This validates the effectiveness of both the dynamic reward mechanism and the complex-valued residual architecture. Specifically, the SCNR achieved by CV-ResDQN at each range cell is consistently higher than that of the DRF-enhanced DQN, indicating that the complex-valued residual module is more effective in preserving target signal energy while suppressing clutter. Furthermore, the CV-ResDQN-DRF method achieves a further improvement in SCNR, reaching the best overall performance. However, its gain is not a linear superposition of the individual gains contributed by the two modules, suggesting a possible complementary relationship between them.

Target range cell	**19**	**31**	**53**	**74**
SCNR output by DQN method [dB]	26.7871	26.7136	20.4597	21.3220
SCNR output by DQN-DRF method [dB]	27.9183	27.7668	21.4373	22.3537
SCNR output by CV-ResDQN method [dB]	28.0106	28.1434	22.1824	22.8294
SCNR output by CV-ResDQN-DRF method [dB]	29.0028	28.9626	23.0925	23.6698

Tab. 8. Comparison of the output SCNR of each range cell under different module configurations.

Target range cell	136	705	1273	1618	2337	4153
CV-ResDQN-DRF mean SCNR [dB]	16.2404	15.3520	14.5509	14.3808	13.0452	12.3168
CV-ResDQN-DRF 95% CI [dB]	16.24 ± 0.16	15.35 ± 0.25	14.55 ± 0.16	14.38 ± 0.22	13.05 ± 0.27	12.32 ± 0.12
Mean SCNR gain [dB]	2.17	2.22	1.70	1.62	1.65	1.15
SCNR gain 95% CI [dB]	2.17 ± 0.16	2.22 ± 0.25	1.70 ± 0.16	1.62 ± 0.22	1.65 ± 0.27	1.15 ± 0.12

Tab. 9. Mean SCNR and SCNR gain (with 95% CI) for different target cells.

Additionally, it is observed that the SCNR gains at range cells 53 and 74 are lower than those at range cells 19 and 31. This is because the simulated clutter is centered at 150 Hz, while the target signal is located at 440 Hz. As a result, part of the target signal is inevitably attenuated during the filtering process, leading to a smaller SCNR improvement in distant range cells compared to those closer to the radar. Nevertheless, the proposed method still demonstrates the best performance in terms of signal preservation, maintaining the highest SCNR even at distant range cells. This highlights its robustness and superiority in complex clutter environments.

4.4 Performance Impact Analysis of Each Module Under Measured Clutter Data

This experiment further analyzes the results presented in Sec. 4.2. The aim is to evaluate the individual performance contributions of each module (complex-valued residual structure and dynamic reward function) under real-world conditions. Based on the baseline DQN clutter suppression method, the dynamic reward function (DRF) and complex-valued residual network (CV-ResDQN) are introduced separately, and then combined, to assess their respective impacts on the improvement of signal-to-clutter-plus-noise ratio (SCNR). This approach enables the quantification of individual gain contributions and verifies the effectiveness of each module in measured clutter environments. The same measured clutter dataset as previously described was employed, with specific simulation parameters referenced in Tab. 6. Table 10 presents a comparison of SCNR performance across different range cells under various module configurations. It evaluates the clutter suppression effects of the DQN, DQN-DRF, CV-ResDQN, and CV-ResDQN-DRF methods on the measured clutter data. Six representative range cells (136, 705, 1273, 1618, 2337 and 4153) were selected to validate the applicability of each module in real-world scenarios.

However, regardless of the environment, the CV-ResDQN-DRF method consistently achieves the highest SCNR. The gains at range cells 136, 705, 1273, 1618, 2337, and 4153 are 2.1062 dB, 2.2933 dB, 1.6781 dB, 1.5866 dB, 1.6461 dB, and 1.1974 dB, respectively, further validating the effectiveness of their combination. It is worth noting that the magnitude of these gains is not simply a linear superposition; for example, the gains at range cells 2337 and 4153 are significantly lower than those in the strong sea clutter regions. This may be related to the spectral distribution characteristics of the target signal and clutter, which affect the signal preservation capability of the filter at these range cells. In summary, the experiments demonstrate that both the CV-ResDQN and DQN-DRF methods effectively enhance SCNR, each exhibiting strengths under different types of clutter environments. Their combination, CV-ResDQN-DRF, achieves the best overall performance across all tested range cells, demonstrating strong adaptability and generalization capability.

Target range cell	136	705	1273
SCNR output by DQN method [dB]	14.0699	13.1362	12.8506
SCNR output by DQN-DRF method [dB]	14.9688	13.9831	13.6610
SCNR output by CV-ResDQN method [dB]	15.6392	14.4126	14.0831
SCNR output by CV-ResDQN-DRF method [dB]	16.1761	15.4295	14.5287
Target range cell	**1618**	**2337**	**4153**
SCNR output by DQN method [dB]	12.7624	11.3966	11.1626
SCNR output by DQN-DRF method [dB]	13.7177	12.0357	11.6790
SCNR output by CV-ResDQN method [dB]	13.4232	11.4369	11.4695
SCNR output by CV-ResDQN-DRF method [dB]	14.3490	13.0427	12.3600

Tab. 10. Comparison of the output SCNR of each range cell under different module configurations.

4.5 Analysis of the Policy Optimization Process

During the training process of reinforcement learning algorithms, the efficiency of policy optimization and convergence characteristics are critical to the final model performance. In this section, the experimental parameter settings and model configurations are kept consistent with those used in the "Performance Impact Analysis of Each Module under Measured Clutter Data" to ensure comparability of results. Figure 9 illustrates the Q-value variations during training for different methods in this experiment. By comparing the Q-value ascent rates and final stable values at various training stages, the contribution of each module to policy optimization can be intuitively evaluated. This comparison reveals the respective roles of the dynamic reward mechanism and the complex-valued residual unit in accelerating learning and enhancing long-term convergence.

As shown in Fig. 9, the Q-value convergence processes of the four algorithms DQN, DQN-DRF, CV-ResDQN, and CV-ResDQN-DRF exhibit distinct characteristics. Specifically, the traditional DQN clutter suppression method shows a relatively slow increase in Q-values during the initial training phase and only stabilizes after a prolonged training period. This indicates limitations in parameter ad-

justment and feature extraction capabilities when dealing with complex-valued clutter environments, thereby affecting the overall efficiency of policy optimization. The DQN-DRF method, which incorporates the dynamic reward mechanism, initially has a slower Q-value increase compared to the CV-ResDQN method. However, it surpasses CV-ResDQN at around the 87th training step. This suggests that the dynamic reward mechanism, through progressive weight adjustment, significantly enhances policy optimization efficiency in the early stages. It does so by modulating the reward distribution, thereby accelerating policy convergence. However, as training progresses, particularly around the 301st step, the Q-values of the CV-ResDQN method once again surpass those of the DQN-DRF method. This indicates that the complex-valued residual unit demonstrates stronger feature extraction and optimization capabilities during the later stages of learning, enabling more effective separation of clutter interference from target signals and thereby improving decision quality. Overall, the CV-ResDQN-DRF method, which combines both the dynamic reward function (DRF) and complex-valued residual network (CV-ResDQN) mechanisms, consistently demonstrates superior performance throughout the entire training process. In the early training stages, it effectively leverages the dynamic reward mechanism to accelerate the increase of Q-values, while in the later stages, it relies on the complex-valued residual network to ensure policy stability and convergence to the optimal solution. This behavior is clearly reflected in the quantitative comparison of Q-value progression. It further validates the adaptability and robustness of the CV-ResDQN-DRF method in complex environments, and highlights its significant advantages in accelerating learning and enhancing final performance.

4.6 Analysis of the Policy Optimization Process

This section employs the same parameter settings and model configurations as those used in the "Performance Impact Analysis of Each Module under Measured Clutter Data" to further compare the error convergence characteristics of different methods during the training process. As shown in Fig. 10, the traditional DQN exhibits the highest overall training error and a relatively slow convergence rate, with a final error value that remains comparatively large. This indicates that in complex clutter environments, the DQN's ability to recognize clutter characteristics is limited, making it difficult to adjust parameters promptly to reduce error. Consequently, its training process shows significant fluctuations and insufficient final stability. In contrast, the DQN-DRF method, which incorporates a dynamic reward mechanism, demonstrates improved performance. By including a reward component in the error function, the model is guided to some extent by the reward signals during training, thereby effectively reducing the error to a certain degree. However, although DQN-DRF reduces both the peak error and the final convergence error compared to the traditional DQN, its improvement is primarily reflected in lowering the overall convergence value. Its capability to better identify clutter characteristics remains limited, as evidenced by persistent fluctuations during the training error decline, showing only minor oscillations after convergence. Building upon this, the method employing the complex-valued residual network (CV-ResDQN) demonstrates significantly superior performance. The CV-ResDQN method is able to more rapidly identify and suppress clutter during training, resulting in a faster decline in training error and significantly lower peak error values compared to the DQN and DQN-DRF methods. More importantly, CV-ResDQN maintains smaller oscillations after convergence, indicating greater stability and robustness throughout prolonged training. Ultimately, the CV-ResDQN-DRF method, which integrates both the dynamic reward and complex-valued residual mechanisms, demonstrates the best training error convergence characteristics. It not only achieves a rapid error reduction in the early stages but also maintains the lowest error values and minimal post-convergence fluctuations throughout the entire training process.

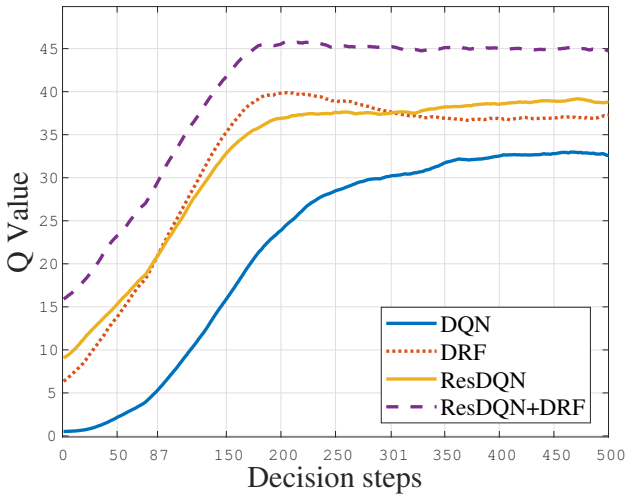


Fig. 9. Q-value variations during training for different methods.

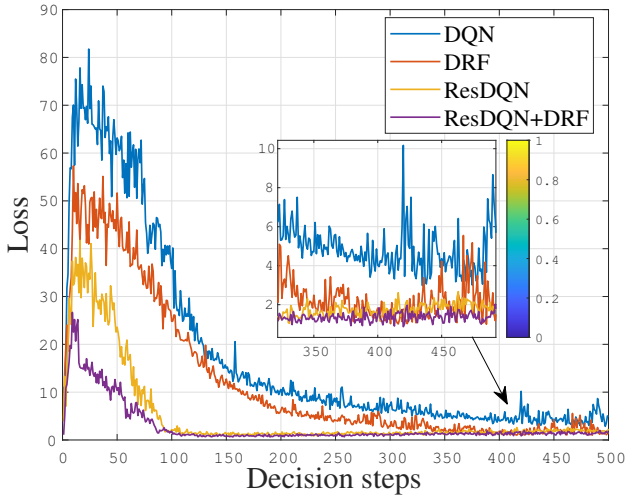


Fig. 10. Loss error variation during training process for different methods.

5. Conclusions

This paper addresses the challenges of low signal-to-noise ratio and poor adaptability of traditional filtering methods in radar target detection under complex clutter environments. To overcome these issues, a deep reinforcement learning model is proposed, which integrates a complex-valued residual network with a dynamic reward mechanism (CV-ResDQN-DRF). The approach leverages complex-valued residual units to enhance feature extraction of radar complex signals' amplitude and phase, while the dynamic reward function mitigates training sparsity issues. Experiments on both simulated and measured datasets demonstrate that the proposed method significantly outperforms the conventional DQN approach in terms of SCNR. Ablation studies further confirm the critical roles of the complex-valued residual structure and dynamic reward mechanism. Despite these promising results, there remains room for improvement, including network architecture optimization, multi-dimensional reward design, and online learning capability. Future work may incorporate attention mechanisms, multi-objective reward frameworks, and real-time deployment strategies to further enhance the practical value of the algorithm.

References

[1] TORRES, S. M., WARDE, D. A. Ground clutter mitigation for weather radars using the autocorrelation spectral density. *Journal of Atmospheric and Oceanic Technology*, 2014, vol. 31, no. 10, p. 2049–2066. DOI: 10.1175/JTECH-D-13-00117.1

[2] CHEN, X., WANG, G., DONG, Y., et al. Sea clutter suppression and micromotion marine target detection via radon-linear canonical ambiguity function. *IET Radar, Sonar & Navigation*, 2015, vol. 9, no. 6, p. 622–631. DOI: 10.1049/iet-rsn.2014.0318

[3] BARRETT, C. R. MTI and pulsed doppler radar. Chapter in: *Eaves, J. L., Reedy, E. K. Principles of Modern Radar*, 1987, p. 422–464. ISBN: 9781461319719

[4] AOYAGI, J. A study on the MTI weather radar system for rejecting ground clutter. *Papers in Meteorology and Geophysics*, 1983, vol. 33, no. 4, p. 187–243. DOI: 10.2467/mripapers.33.187

[5] RANNEY, K., MARTONE, A., SOUMEKH, M. Indication of slowly moving targets via change detection. *Radar Sensor Technology XI*, 2007, vol. 6547, p. 196–207. DOI: 10.1117/12.720743

[6] ORESHKIN, B. N. Adaptive filters for the moving target indicator system [in Russian]. *arXiv preprint*, 2020, p 1–149. DOI: 10.48550/arXiv.2012.15440

[7] SHORT, R. D. An adaptive MTI for weather clutter suppression. *IEEE Transactions on Aerospace and Electronic Systems*, 2007, no. 5, p. 552–562. DOI: 10.1109/TAES.1982.309268

[8] CHEN, J., HUANG, P., XIA, X. G., et al. Multichannel signal modeling and AMTI performance analysis for distributed space-based radar systems. *IEEE Transactions on Geoscience and Remote Sensing*, 2022, vol. 60, p. 1–24. DOI: 10.1109/TGRS.2022.3202567

[9] EHARA, N., SASASE, I., MORI, S. Moving target detection by quadrature mirror filter. *Electronics and Communications in Japan (Part I: Communications)*, 1996, vol. 79, no. 4, p. 55–62. DOI: 10.1002/ecja.4410790406

[10] PRABHU, K. M. M., GIRIDHAR, J. Detection performance of an adaptive MTD with WVD as a Doppler filter bank. *Signal Processing*, 2001, vol. 81, no. 4, p. 693–698. DOI: 10.1016/S0165-1684(00)00241-3

[11] LI, G., ZHANG, H., GAO, Y., et al. Sea clutter suppression using smoothed pseudo–Wigner–Ville distribution–singular value decomposition during sea spikes. *Remote Sensing*, 2023, vol. 15, no. 22, p. 1–20. DOI: 10.3390/rs15225360

[12] HE, W., LUO, Y., SHANG, X. Motion clutter suppression for non-cooperative target identification based on frequency correlation dual-SVD reconstruction. *Sensors*, 2024, vol. 24, no. 16, p. 1–17. DOI: 10.3390/s24165298

[13] POON, M. W. Y., KHAN, R. H., LE-NGOC, S. A singular value decomposition (SVD) based method for suppressing ocean clutter in high frequency radar. *IEEE Transactions on Signal Processing*, 1993, vol. 41, no. 3, p. 1421–1425. DOI: 10.1109/78.205747

[14] CHEN, Z., HE, C., ZHAO, C., et al. Using SVD-FRFT filtering to suppress first-order sea clutter in HFSWR. *IEEE Geoscience and Remote Sensing Letters*, 2017, vol. 14, no. 7, p. 1076–1080. DOI: 10.1109/LGRS.2017.2697458

[15] CHENG, Q., WU, X., ZHANG, X., et al. A novel sea clutter suppression method based on SVD-FRFT at low signal-to-clutter ratio. *Electronics Letters*, 2023, vol. 59, no. 14, p. 1–3. DOI: 10.1049/ell2.12874

[16] TORRES, S. M., WARDE, D. A. Ground clutter mitigation for weather radars using the autocorrelation spectral density. *Journal of Atmospheric and Oceanic Technology*, 2014, vol. 31, no. 10, p. 2049–2066. DOI: 10.1175/JTECH-D-13-00117.1

[17] BOWYER, D. E., RAJASEKARAN, P. K., GEBHART, W. W. Adaptive clutter filtering using autogressive spectral estimation. *IEEE Transactions on Aerospace and Electronic Systems*, 1979, vol. AES-15, no. 4, p. 538–546. DOI: 10.1109/taes.1979.308738

[18] CHENG, Y., SU, J., XIU, C., et al. Adaptive clutter intelligent suppression method based on deep reinforcement learning. *Applied Sciences*, 2024, vol. 14, no. 17, p. 1–18. DOI: 10.3390/app14177843

[19] DENG, J., SU, C., ZHANG, Z.-M., et al. Evolutionary game analysis of chemical enterprises' emergency management investment decision under dynamic reward and punishment mechanism. *Journal of Loss Prevention in the Process Industries*, 2024, vol. 87, p. 1–14. DOI: 10.1016/j.jlp.2023.105230

[20] JARRAY, R., ZAGHBANI, I., BOUALLÈGUE, S. Dynamic reward-based deep reinforcement learning algorithm for UAV path planning in large-scale environments. *Procedia Computer Science*, 2025, vol. 270, p. 692–702. DOI: 10.1016/j.procs.2025.09.189

[21] WANG, S., CHENG, H., KE, Z., et al. Complex-valued residual network learning for parallel MR imaging. In *Joint Annual Meeting ISMRM-ESMRMB*. Paris (France), 2018.

[22] YANG, L., WU, Z., JIANG, L., et al. Underwater source localization via active deep complex residual network in a shallow-water waveguide. *The Journal of the Acoustical Society of America*, 2025, vol. 158, no. 4, p. 3017–3035. DOI: 10.1121/10.0039579

[23] VIGER, R., MIROTZNIK, M., LAMBRAKOS, S. G. Synthetic aperture radar image enhancement and phase characterization using complex-valued neural networks. *Journal of Applied Remote Sensing*, 2025, vol. 19, no. 2, p. 1–29. DOI: 10.1117/1.JRS.19.026504

About the Authors ...

Yi CHENG was born in Heilongjiang Province, China, in December 1979. She received the B.S. degree in Electrical Engineering and Automation from Harbin Institute of Technology, Heilongjiang, China, in 2001, and the M.S. and Ph.D. degrees in Navigation, Guidance, and Control from Harbin Engineering University, Heilongjiang, China, in 2004 and 2008, respectively. She is currently an associate professor at Tiangong University, engaged in radar signal processing research.

Jiaxin LIU (corresponding author) was born in Liaoning Province, China. He received the B.S. degree in Electrical Engineering and Automation from Shenyang Jianzhu University, Liaoning, China. He is currently a master's student majoring in Control Science and Engineering at Tiangong University; his research focuses on radar clutter suppression.

Junjie SU received the B.S. degree in Mechanical Engineering and Automation from Tianjin University of Technology. He received his M.S. degree in Control Engineering from Tiangong University; his research focuses on radar clutter suppression.