ITC 2/54

Information Technology
and Control
Vol. 54 / No. 2/ 2025
pp. 576-592
DOI 10.5755/j01.itc.54.2.40167

Single-Pulse Detection Method of Radar Weak Target Based on a Two-Stage Deep Neural Network

Received 2024/01/14 | Accepted after revision 2025/04/02

HOW TO CITE: Qiu, M., Wang, J., Wu, G. (2025). Single-Pulse Detection Method of Radar Weak Target Based on a Two-Stage Deep Neural Network. *Information Technology and Control*, 54(2), 576-592. https://doi.org/10.5755/j01.itc.54.2.40167

Single-Pulse Detection Method of Radar Weak Target Based on a Two-Stage Deep Neural Network

Mingjie Qiu

China Academy of Electronics and Information Technology, No. 11 Shuangyuan Rd Shijingshan District, Beijing, China

Nanjing Research Institute of Electronics Technology, No. 8 Guorui Road Yuhuatai District, Nanjing, China

Jianming Wang, Guangxin Wu

Nanjing Research Institute of Electronics Technology, No. 8 Guorui Road Yuhuatai District, Nanjing, China

Corresponding author: qiumingjie_mj@outlook.com

With the increasing prevalence of drones in low-altitude airspace, the radar detection of weak targets with a low signal-to-noise ratio (SNR) still poses a crucial challenge. Traditional constant false alarm rate (CFAR) methods encounter issues of high false alarms and low accuracy when the SNR is below-15dB. This paper puts forward a two-stage deep neural network to improve weak target detection by emulating human visual perception. In the first stage (coarse detection), potential targets are rapidly localized through grid-based regression. In the second stage (fine detection), depth-wise separable convolution (DSC) and residual connections are utilized for accurate classification. Experimental results show that, at an SNR of -20dB, the detection rate of the proposed method is 20% higher than that of CFAR methods, and the inference speed is 3.66 times faster than that of single-stage networks. Ablation studies confirm the efficiency improvements brought by the coarse detection network. This approach offers a robust solution for real-time drone surveillance in complex and cluttered environments.

KEYWORDS: Radar; Single-Pulse Detection; Deep Learning; Weak Target Detection

1. Introduction

Radar, an acronym for "radio detection and ranging," is currently the most extensively applied detection method for air target surveillance [26-27]. Given its all-weather operation ability, long-range coverage, and sensitivity to small radar cross-sections (RCS), radar remains the primary means for drone detection. Distinct from optical sensors, radar functions effectively in fog, rain, and darkness, rendering it essential for continuous low-altitude surveillance. Generally, radar detects targets based on the echo amplitude of the targets. Thus, setting a reasonable detection threshold is crucial to stably detect targets while preventing excessive false alarms that could otherwise undermine detection performance [4]. Constant false alarm rate (CFAR) detection, a significant approach for target detection, has long been a focal point in radar target detection research [14, 19, 3, 35, 13].

Traditional CFAR detectors, such as cell averaging CFAR (CA-CFAR) [38], the smallest of CFAR (SO-CFAR), and the greatest of CFAR (GO-CFAR) [21], estimate noise power from reference cells to determine amplitude thresholds. Although CA-CFAR faces difficulties in heterogeneous clutter, SO/GO-CFAR enhance robustness by choosing min/max reference values, yet they demand manual parameter adjustment [22, 11, 10]. Nevertheless, CFAR algorithms suffer from several common limitations. First, they overly depend on amplitude-only features. Second, the CFAR algorithm is highly sensitive to situations where the SNR is below -10dB. In such low-SNR conditions, the signal is prone to noise corruption, severely deteriorating the detection performance and making accurate weak-target identification arduous. Finally, when handling weak targets, the CFAR algorithm fails to effectively exploit phase coherence.

For small drones with low SNR, the radar detection effect is far from satisfactory [29]. Owing to their small radar cross-section (RCS) and low flight altitude, the target signals of typical drones in radar echoes are sometimes too feeble to be effectively detected. The core challenge lies in the fact that differentiating weak drone echoes (SNR<-15dB) from clutter necessitates the simultaneous analysis of amplitude and phase features, which traditional CFAR methods cannot accomplish. Consequently, radar detection of small unmanned aerial vehicles has always been a challenging and hot topic [1, 25, 34].

With the rapid development of neural networks, detection methods based on deep learning have garnered unprecedented attention [6]. Thanks to their capacity to automatically extract deep-level features from data, these methods have been successfully applied in diverse fields, such as speech recognition [15, 12, 17] and image recognition [31], thereby offering novel research perspectives for radar applications. Capitalizing on the remarkable recognition and object detection capabilities in images, deep learning has been explored in radar target recognition and detection [40]. Currently, the application of neural networks in radar target detection primarily centers on target detection in range-Doppler (RD) spectrum images [29, 32, 30, 33, 8] and synthetic aperture radar (SAR) images [36, 14, 39, 5]. The main rationale is that neural-network-based target detection methods are typically applicable to two dimensional images, and RD spectrum images or SAR images share many similarities with them, facilitating their application. Recent advances in deep learning for radar detection have explored diverse architectural paradigms. Wang et al. [28] achieved high accuracy through a CNN-based classification approach, but their method is plagued by high latency. Transformer-based models like [37] demonstrate exceptional feature fusion capabilities through self-attention mechanisms, however their computational complexity (typically 3.8× higher than CNNs) renders them impractical for real-time detection. Lightweight architectures such as Tiny-YOLO [23] and GhostNet [7] address efficiency concerns through network pruning and feature map redundancy reduction, yet struggle with low-SNR targets due to their limited depth and insufficient phase coherence analysis. Traditional CFAR methods lack phase analysis, while deep-learning-based approaches suffer from high latency. Our work aims to bridge this gap by directly processing raw echoes and optimizing computational efficiency through a two-stage architecture.

To reduce computational complexity, on one hand, we propose using echo signals rather than RD spectrum images as inputs for deep-learning algorithms. On the other hand, we attempt to conduct fine-grained sliding window detection only in areas with a high probability of target appearance, thus reducing the algorithm's computational complexity. In this manner, a network structure is devised, where initial localization is carried out by the coarse detection network, and point by point confirmation is performed by the fine detection network. As preliminary research, we have initially validated the effectiveness of this configuration model on simulated data. However, two limitations still exist [20]. Firstly, due to the relatively simple structure of the initially constructed model, its performance on real data remains to be improved. Secondly, the computational load of the model during the fine detection phase remains relatively high, with its running time being over ten times that of CFAR detection. To address these two limitations, this paper focuses on optimization and enhancement.

The main contributions of this work are summarized as follows:

1. A two-stage architecture integrating coarse localization and fine classification, which reduces the computational load by 73.7%.
2. Integration of DSCResNet for efficient feature extraction, achieving an 85% reduction in parameters.
3. Comprehensive validation on real X - band radar data, demonstrating an AUC > 0.82 at - 20dB SNR.

The remainder of the paper is structured as follows. Section 2 introduces traditional CFAR detection methods. Section 3 then presents the proposed single pulse detection method based on deep learning step by step. In section 4, the method is tested and compared with traditional detection methods. Finally, discussions and conclusions are drawn in section 5.

2. Traditional CFAR Detection

Figure 1 depicts the flowchart of the traditional CFAR detection method used for radar target detection.

As seen in Figure 1, after the radar receives the echo signals from targets, based on the characteristic of traditional detection methods relying on the amplitudes of echo signals for target detection, the amplitudes of echo signals need to be enhanced first. Assuming that the radar transmits a linear frequency-modulated (LFM) signal with operating frequency f, pulse width τ, and frequency modulation slope K, the LFM signal can be expressed as

$$e(t) = A\,\mathrm{rect}\left(\frac{t}{\tau_p}\right)\exp\left(j\pi K t^2\right)\exp\left(j2\pi f_0 t\right), \tag{1}$$

where A represents the intensity of the reflected signal, j is the imaginary unit, and rect(·) is the rectangular signal function.

To effectively highlight the amplitude characteristics of the target echo signal, pulse compression processing needs to be performed on the received echo signal. According to the principle of matched filtering, for the transmitted signal as shown in (1), its matched filter h(t) can be expressed as

$$h(t) = \overline{s(-t)} = \mathrm{rect}\left(\frac{t}{\tau_p}\right)\exp\left(-j\pi K t^2\right) \tag{2}$$

and the signal after pulse compression can be expressed as

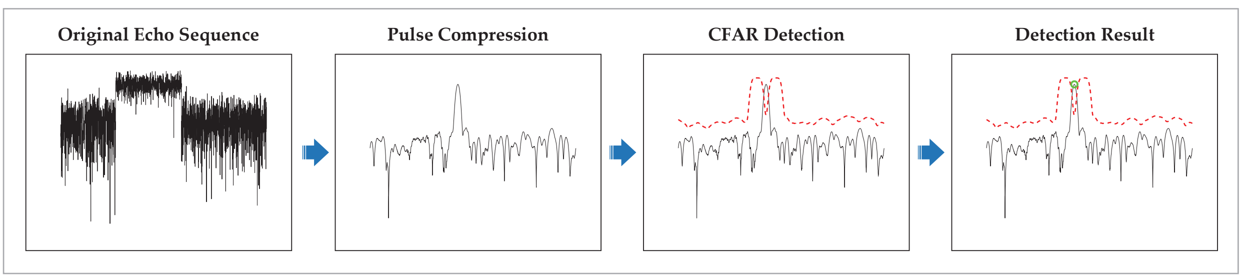
$$s_{\mathrm{pc}}(t) = r(t) * h(t), \tag{3}$$

where * represents the convolution operation.

Upon completion of the pulse compression process, the target echo data is fed into the CFAR detection algorithms, as graphically presented in Figure 2. As is evident from the figure, the CFAR detection mechanism utilizes a sliding window strategy for process-

Figure 1

Schematic workflow of the traditional CFAR detection methodology.



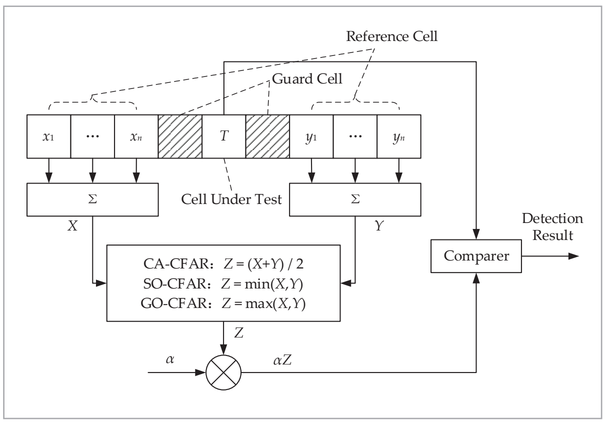
ing the input echo signal sequence. In each iteration, a specific segment of data from the echo signal sequence is extracted, and a determination is made as to whether the central cell within that segment harbors a target.

The CFAR detection approach categorizes the data collected by each sliding window into three distinct types. First is the cell under test (CUT), which serves to assess whether the current detection window encompasses a target within the specific cell. Second is the guard cell (GC), whose fundamental function is to preclude the leakage of target energy into neighboring cells. Such leakage could otherwise inflate the detection threshold and result in missed detections. Third is the reference cell (RC), which is assumed to contain solely the background noise characteristic of the target's environment. Through a meticulous statistical analysis of the signal strength within these cells, an accurate estimate of the current CUT detection threshold can be derived.

Prior to the emergence of deep learning techniques, CFAR detection methods have long held a pivotal position in the realm of radar target detection and have been extensively implemented in engineering applications. Nevertheless, CFAR detection methods have consistently encountered formidable challenges. One of the key issues lies in the selection of the number of reference cells, which significantly impacts the detection performance. Additionally, under diverse scenarios, determining the most suitable CFAR detection method among multiple options remains a complex task. The choice of the method needs to be carefully tailored to the specific characteristics of the radar environment, such as the SNR, target distribution, and interference level, further complicating the application process.

Figure 2

Operational framework of CFAR detection with sliding window mechanism and threshold adaptation process



3. Two-Stage DSCResNet Detection

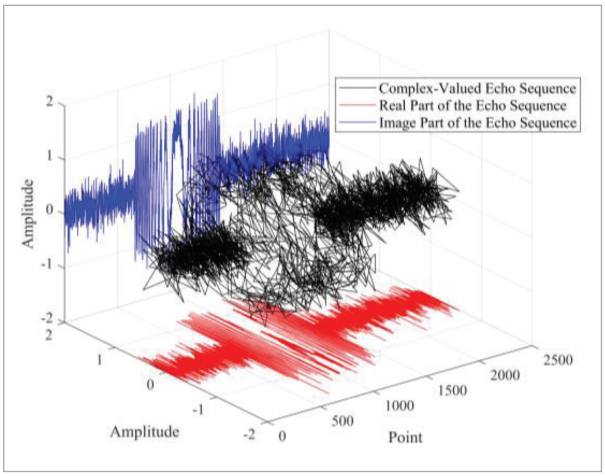
Figure 3 illustrates the real and imaginary components of a target echo sequence with a SNR of 10 dB prior to pulse compression. In the traditional CFAR detection process, it is evident that target detection typically hinges on the signal amplitude. Nevertheless, as depicted in Figure 3, even before pulse compression, the real and imaginary parts of the echo signal are capable of reflecting the phase characteristics of the chirp signal. This implies that there are potential information sources beyond the amplitude, which traditional CFAR detection methods may overlook, thus providing an opportunity for exploring more comprehensive detection strategies.

3.1 Network Structure

Figure 4 depicts the proposed two-stage deep neural network target detection process. The first stage network draws on the design concept of the YOLO detection network [2]. It partitions the input data into multiple uniform grids. Subsequently, after feature extraction using convolutional layers, a regression method is utilized to calculate the probability of a tar-

Figure 3

Time-domain representation of raw radar echo components (SNR=10dB) showing preserved phase information in real and imaginary parts.



get's presence in each grid, thereby achieving an initial approximation of the target's location. We term this as the coarse detection network.

Based on the results of the coarse detection, sequences potentially containing targets are intercepted point by point through a sliding window method. The intercepted data is then fed into a deep convolutional network for target detection, which we refer to as the fine detection network. The coarse detection network mirrors the human visual system's rapid global scanning mechanism, while the fine detection network emulates the focused attention on specific localized regions. This hierarchical detection approach effectively strikes a balance between detection speed and accuracy, offering a more efficient solution for target detection tasks in complex scenarios.

3.1.1 Coarse Detection Network

In image target detection, the YOLO algorithm in target detection is not much accurate but improves the speed of detection [9]. Considering the use of pixel-by-pixel detection in the second stage, to enhance the overall detection speed of the method, we have designed a coarse detection network with 7 convolutional layers with max pooling, inspired by YOLO's grid-based regression, as shown in Figure 5.

Figure 4

Architectural overview of the proposed two-stage detection framework.

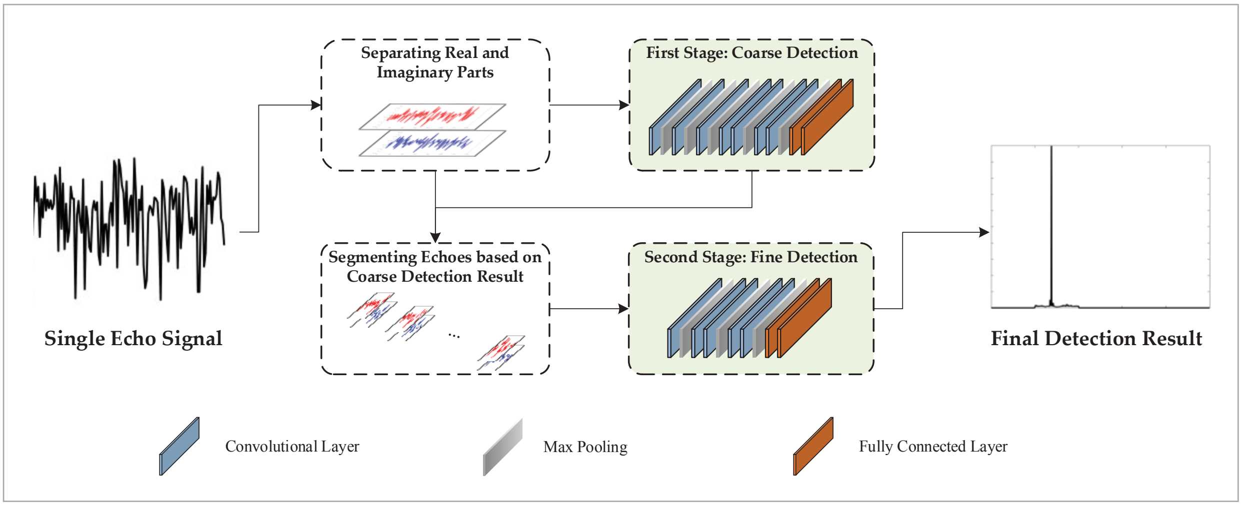
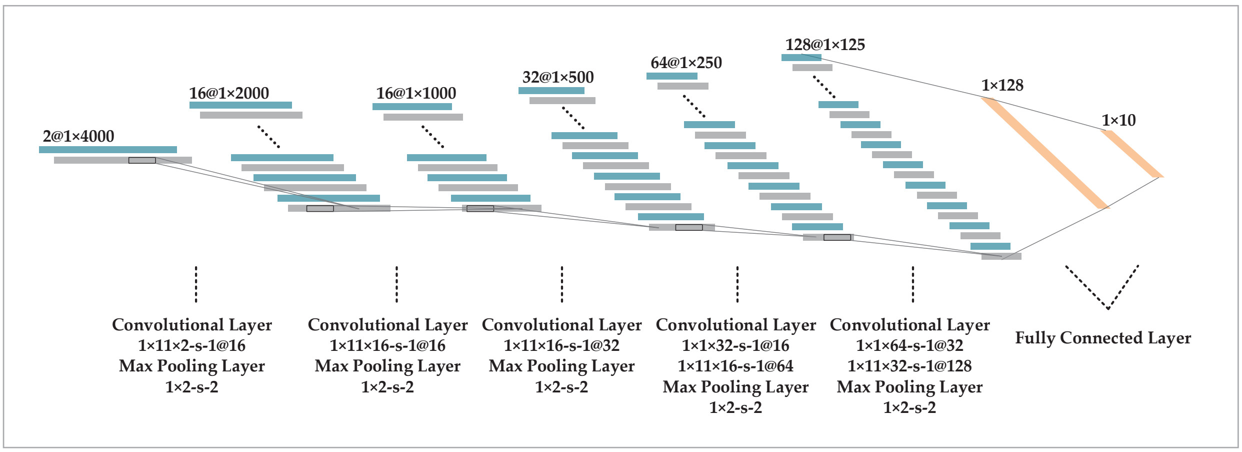


Figure 5

Detailed layer configuration of the coarse detection network featuring seven convolutional layers with max pooling operations for rapid target region proposal.



From the previous analysis, it is evident that separating the real and imaginary parts of the signal can facilitate the observation of target echo characteristics. Thus, we decompose an echo signal sequence into its real and imaginary components and use them jointly as the input to the network. When the signal length is 4000, the network input can be represented as an image with dimensions of 1×4000×2.

In the first layer, a convolution kernel of size 1×11×2 is first applied to convolve the input data. This operation enables the interaction of information from the real and imaginary parts, ensuring that subsequent convolutional operations can concurrently extract features from both components. Subsequently, the input data proceeds through the remaining layers according to the predefined model for feature extraction, yielding a feature map of size 1×125×128. The output of each convolutional layer is activated by the rectified linear unit (ReLU) function to enhance the network's non-linearity, which can be expressed as

$$\mathrm{ReLU}(x) = max(0, x). \quad (4)$$

For the network's output part, we adopt 2 fully connected layers as the output head. The sigmoid function is employed as the activation function, which is expressed as

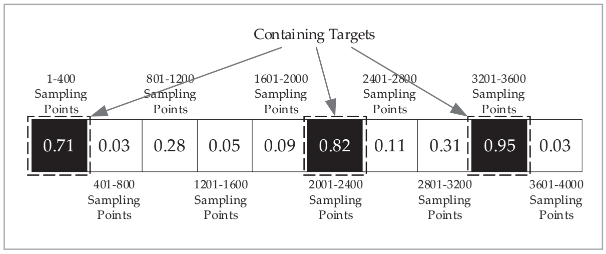
$$\mathrm{Sigmoid}(x) = \frac{1}{1+\exp(-x)}. \quad (5)$$

The first fully connected layer consists of 128 neurons, which is consistent with the depth of the final feature map. The second fully connected layer is designed to obtain the probability of a target echo signal in each of the 10 equal length segments after the input echo sequence is partitioned. Consequently, it has 10 neurons, with each neuron generating a network output value between 0 and 1.

The output of the first stage network is illustrated in Figure 6. Given that the input echo sequence has a length of 4000, each neuron output of the final output layer represents the probability of a target within 400 detection units. In this study, we generally define that if the output value of a neuron exceeds 0.5, it indicates that the probability of a target within the 400 detection units corresponding to that neuron is greater than 50%. This approach obviates the need for experimental or speculative threshold determination.

Figure 6

Grid probability mapping visualization demonstrating the coarse detection network's output.



3.1.2 Fine Detection Network

In the fine detection stage, we also employ the approach of intercepting local data point by point through sliding windows as the input for target detection. Here, the size of the input data is decreased from 1×4000×2 in the coarse detection network to 1×50×2. Thanks to the coarse detection network in the first stage, we have obtained the probability of each segment containing targets after evenly dividing the input data into 10 segments. Only the data segments with probabilities greater than 0.5 are likely to have targets. Consequently, only the data at these qualified positions need to be detected point by point, which can remarkably reduce the computational complexity of precise target detection.

Furthermore, to further enhance the running speed of the method, in the fine - detection network, we utilize the DSC to substitute conventional convolutional operations. DSC can reduce overfitting by minimizing redundant parameters, thus contributing to a more efficient fine-detection process.

DSC, initially proposed by François Chollet, has been subsequently and extensively applied in networks such as Xception [16] and MobileNet [18]. As illustrated in Figure 7, the fundamental concept of DSC lies in the decoupling of the spatial and channel dimensions of feature maps. By decomposing the convolutional operation into two independent convolutions, the network can achieve comparable feature extraction results while substantially reducing the number of network parameters and enhancing the efficiency of the method.

In the context of single pulse detection, which demands a comprehensive exploration of the detailed amplitude and phase features within the echo sequence, a challenge arises. As the number of convolutional layers increases, the detailed features extracted by the network tend to be smoothed out

Figure 7

DSC decomposition diagram showing spatial-channel separation with pointwise and depthwise convolution operations. The symbol * means convolutional operation.

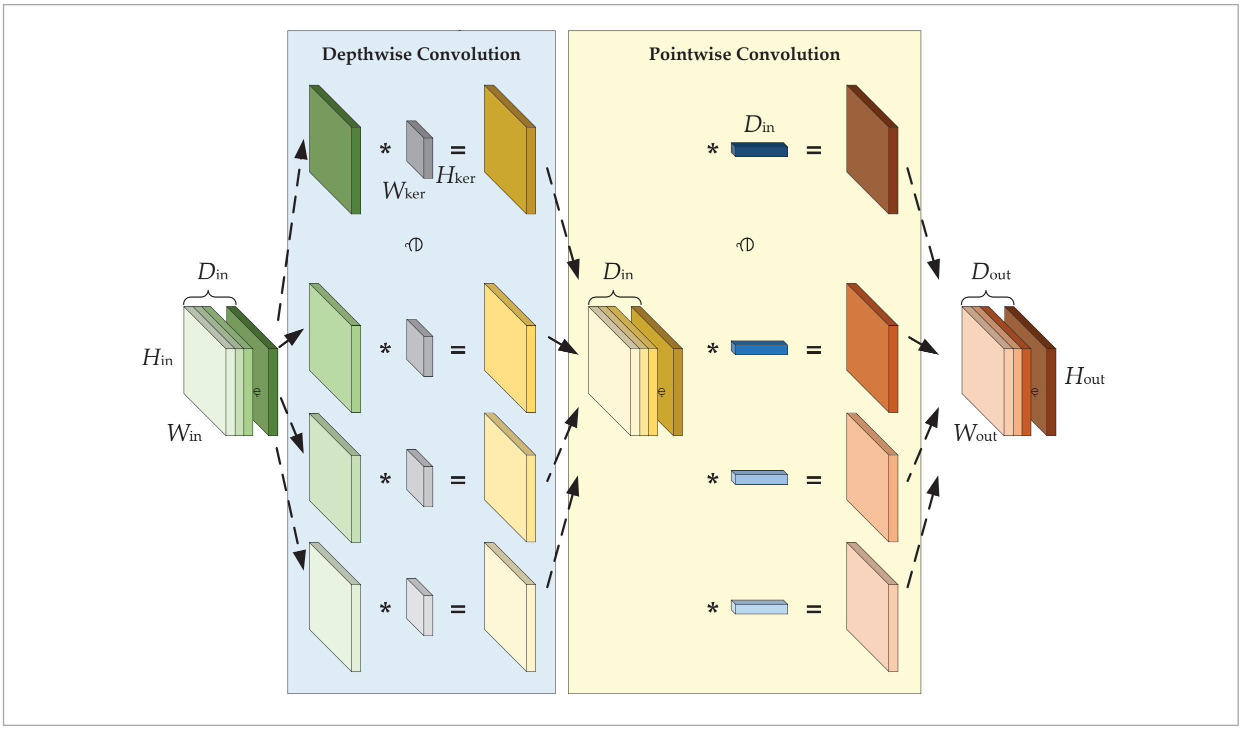
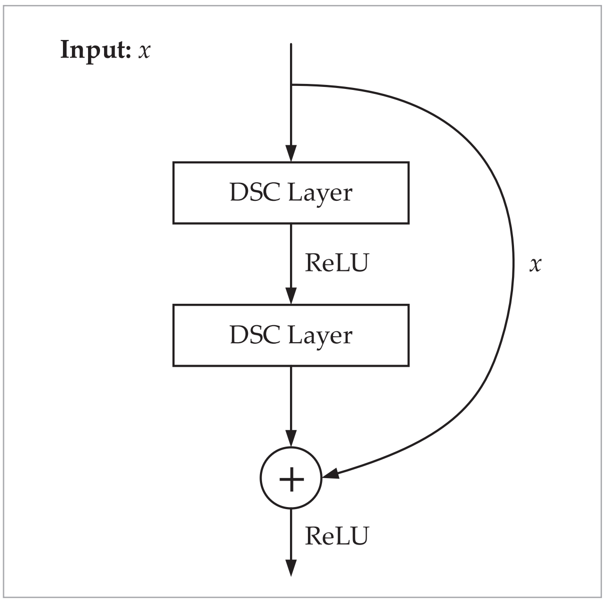


Figure 8

Structural diagram of DSCResBlock integrating depthwise separable convolutions with residual connections for feature preservation.



and lost. To fully leverage the features from different layers during the final classification process, we introduce a shortcut path, similar to that in the residual network [24], after every two DSC layers. These components collectively form a DSC residual block (DSCResBlock). The structure of a single DSCResBlock is presented in Figure 8.

The structure of the fine detection network in the second stage is illustrated in Figure 9. The fine detection network consists of a total of 5 DSCResBlocks, each of which contains depth-wise separable convolutions and residual shortcuts. Given that the fine detection network shares the same input echo sequence with the coarse detection network, yet its input is merely a partial signal cropped from the first stage. Consequently, the search range for targets in the fine detection stage has been substantially reduced. Moreover, the echo features of weak and small radar targets do not occupy a large size range within the signal; thus, there is no necessity for pooling to acquire a larger receptive field. The DSCResBlock preserves weak target features via residual connections. These connections prevent gradient vanishing and ensure stable training on noisy data, which is crucial for accurate detection of weak and small radar targets.

As shown in Figure 9, within the DSCResBlock, the input from the previous network is convolved to extract deeper features. Meanwhile, via the shortcut connection, the input information is incorporated into the output. The red lines in Figure 10 indicate the transmission path of hierarchical features from different layers within a single DSCResBlock.

Figure 9

Fine detection network architecture with five cascaded DSCResBlocks and dual fully-connected layers for precise target/non-target classification.

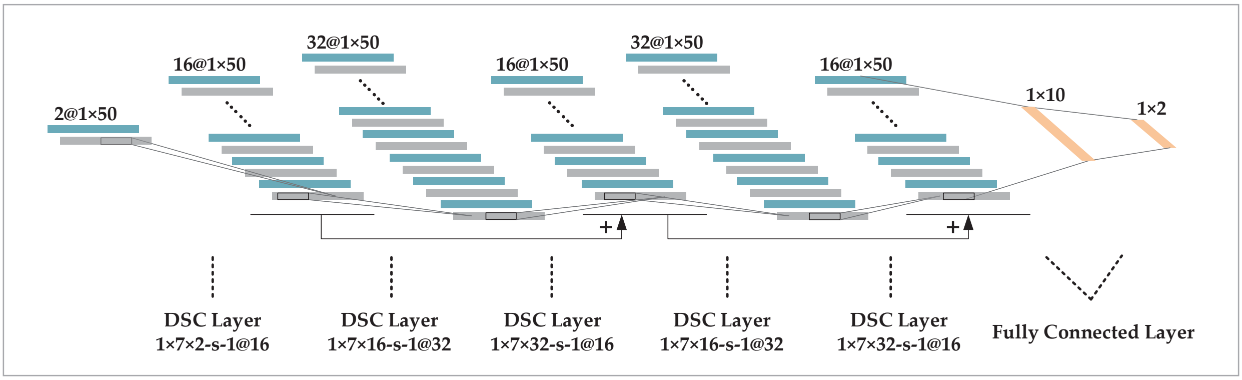


Figure 10

Feature propagation pathways within DSCResBlock highlighting multi-level feature integration through residual connections.

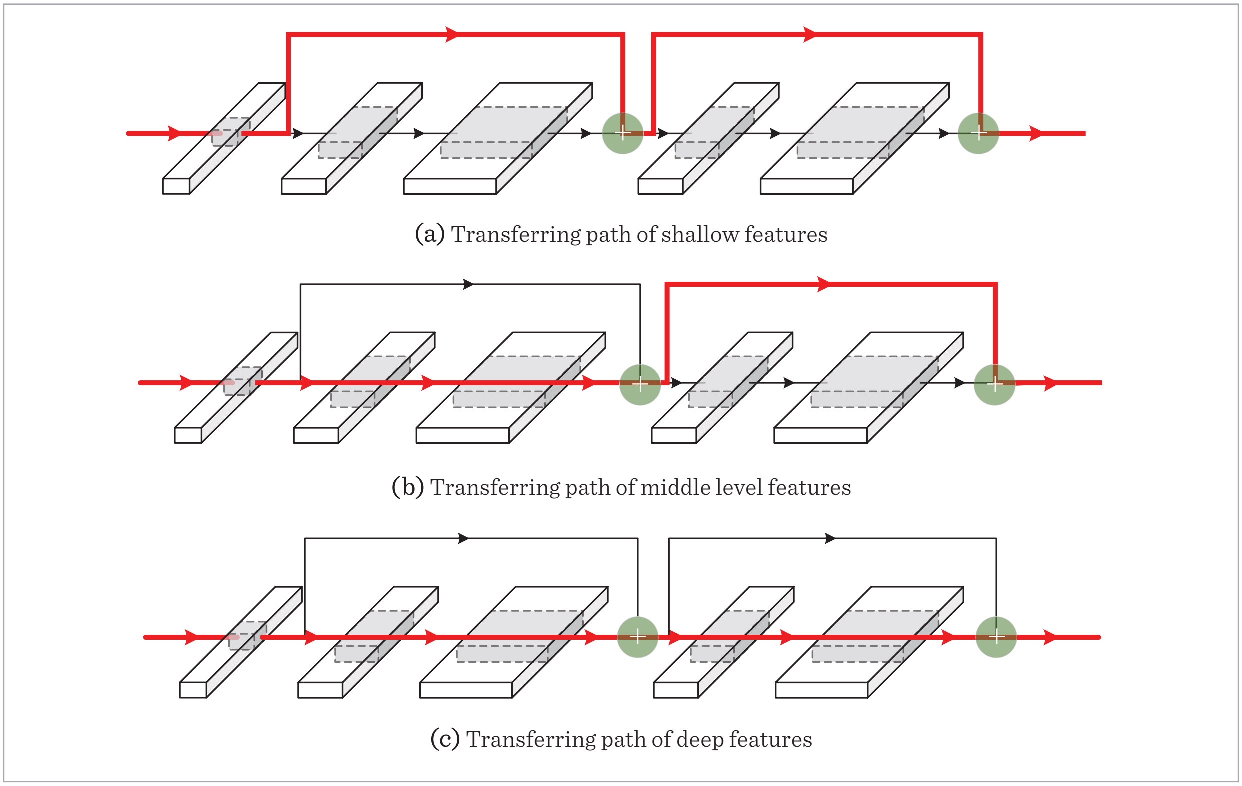


Table 1

Comprehensive architectural specifications of the two-stage DSCResNet detection network detailing layer names, kernel dimensions, activation functions, and output dimensionalities for both coarse and fine detection stages.

Stage	Layer Name	Kernel Size	Kernel Number	Active Function	Output Size
Stage 1	Convolutional Layer	1×11×2	16	ReLU	1×4000×16
	Max Pooling Layer	1×2	-	-	1×2000×16
	Convolutional Layer	1×11×16	16	ReLU	1×2000×16
	Max Pooling Layer	1×2	-	-	1×1000×16
	Convolutional Layer	1×11×16	32	ReLU	1×1000×32
	Max Pooling Layer	1×2	-	-	1×500×32
	Convolutional Layer	1×1×32	16	ReLU	1×500×16
	Convolutional Layer	1×11×16	64	ReLU	1×500×64
	Max Pooling Layer	1×2	-	-	1×250×64
	Convolutional Layer	1×1×64	32	ReLU	1×250×32
	Convolutional Layer	1×11×32	128	ReLU	1×250×128
	Max Pooling Layer	1×2	-	-	1×125×128
	Fully Connected Layer	-	128	Sigmoid	1×128
	Fully Connected Layer	-	10	Sigmoid	1×10
Stage 2	DSC Layer	1×7×2	16	ReLU	1×50×16
	DSC Layer	1×7×16	32	ReLU	1×50×32
	DSC Layer	1×7×32	16	ReLU	1×50×16
	DSC Layer	1×7×16	32	ReLU	1×50×32
	DSC Layer	1×7×32	16	ReLU	1×50×16
	Fully Connected Layer	-	10	Sigmoid	1×10
	Fully Connected Layer	-	2	Softmax	1×2

The head part of the fine detection network is identical to that of the coarse detection network, comprising 2 fully connected layers. The first layer employs the sigmoid function, while the second layer uses the softmax function as the activation function, which can be expressed as

$$\mathrm{Softmax}(x_i) = \frac{\exp(x_i)}{\sum_{n=1}^{N}\exp(x_n)}. \tag{6}$$

where x_i represents the output value of the i^{th} neuron, and N denotes the total number of neurons in that layer. In this case, $N = 2$.

In summary, the complete structure of the two-stage weak target single pulse detection neural network is presented in Table 1.

3.2 Loss Function

Since the final output layer of the network employs the softmax function for classification and consists of two neurons, the output of one neuron represents the probability that the current input contains a target. Conversely, the output of the other neuron represents the probability that the input does not contain a target. From the mathematical expression of the softmax function, it is evident that these two probabilities are complementary, and their sum is always equal to 1.

In the context of the loss function, choosing the commonly used mean squared error (MSE) as the loss metric can effectively facilitate the error propagation within the network. Notably, for the purpose of calculation, it suffices to select the output of the neuron representing the probability of target presence. This approach simplifies the computational process while maintaining the integrity of the network's learning mechanism for accurate target classification.

During the training of the network, let T_k represent the target detection result of the k^{th} input echo sequence segment of the second stage network. If the start of the input segment coincides with the start of the target echo, the ground truth is marked as 1, and in other cases, the ground truth is marked as 0. Then, when the output of the neuron representing the "presence probability of targets" in the output layer is o_{1k} for the k^{th} input segment, the MSE loss is expressed as

$$\mathrm{Loss}_{\mathrm{MSE}}(S) = \frac{1}{K}\sum_{k=1}^{K}\left(T_k - o_{1k}\right)^2, \tag{7}$$

where S is the set of input segments which containing total of K segments.

3.3 Training Parameters

In contrast to the conventional stochastic gradient descent method, the momentum-based method (SGDM) demonstrates a superior ability to prevent the network from getting trapped in local optima during the training process, thus avoiding the situation where further convergence is hindered. For this reason, SGDM is adopted for neural network training in this paper.

The training parameters of the network are presented in Table 2.

Table 2

Training hyperparameter configuration for neural network optimization including SGDM settings, learning rate scheduling, and batch specifications.

Parameter Name	Value
Learning Method	SGDM
Momentum Parameter	0.9
Initial Learning Rate	0.05
Learning Rate Drop Factor	0.5
Learning Rate Drop Period	50
Maximum Epoch	100
Batch Size	10

As presented in the table, for the momentum parameter, the conventional empirical value of 0.9 was consistently utilized. The initial learning rate was set at a relatively small value of 0.05 to ensure the convergence of the training process. The maximum number of training epochs was set to 100. After 50 epochs of training, the learning rate was reduced by half. This parameter setting allows the network to iterate and optimize rapidly at the start of training and eventually converge to the optimal solution with more refined step sizes.

4. Experiments

4.1 Experimental Data

This chapter presents experiments conducted using both simulated and real data of weak and small targets under various conditions. The real data was collected from an X-band drone detection radar. A LFM signal was employed, with a pulse width of 1 μs, a bandwidth of 20 MHz, and a pulse repetition frequency of 8 kHz. The X-band radar was installed on a 15-meter tower in an urban area. The drone utilized was a DJI Phantom 4, which flew at altitudes ranging from 10 to 50 meters, performing hover and linear flight trajectories. The echo sequence was sampled at a high data rate of 25 MHz, yielding an adjacent range cell spacing of approximately 6 meters. After manual data cleaning and sorting, a total of 3000 real echo sequences were selected for testing. For each echo sequence, only about 300 valid data points were retained starting from the detection gate.

To reliably validate the method's performance and mitigate the influence of different signal parameters, the simulated data used for network training and experimentation adopted the same pulse width, signal bandwidth, and pulse repetition frequency as the real data. The detailed information is presented in Table 3.

The simulated echo sequence targets for testing have SNRs ranging from -20dB to 0dB. For each SNR value, 20,000 data samples are generated, with target positions randomly assigned. The data set employed for network training consists of 10,000 echo sequences, all having a constant SNR of 0dB. This choice is to ensure the network learns robust features of target echoes under minimal noise interference. To mitigate domain mismatch, we augmented training data with

Table 3

Simulation parameters for radar echo generation matching real X-band system characteristics.

Parameter Name	Value
Carrier Frequence	10 GHz
Pulse Width	1 μs
Pulse Repetition Frequence	8 kHz
Bandwidth	20 MHz
Sampling Rate	25 MHz
Range Resolution	6 m
SNR for Testing Data	From -20dB to 0dB
SNR for Training Data	0dB

additive white Gaussian noise spanning -10 dB to 0 dB during fine-tuning. The two-stage networks are trained separately using data that match their respective input sizes.

4.2 Evaluation Metrics

To precisely assess the detection performance of diverse target detection methods, it is essential to first elucidate the evaluation metrics for method performance. In binary classification problems like target detection, the receiver operating characteristic (ROC) curve is extensively utilized to evaluate the performance of methods. For target – detection methods, on the ROC curve, the false positive rate (FPR) on the horizontal axis represents the proportion of falsely detected targets among all non - target detection points. Meanwhile, the true positive rate (TPR) on the vertical axis denotes the ratio of correctly detected targets to the total number of targets. This ROC-based evaluation framework provides a comprehensive and intuitive way to compare the performance of different target detection algorithms, enabling researchers to better understand the trade-off between the correct detection of targets and the occurrence of false alarms.

Suppose that there are total of N_A targets to be detected in the set $A = \{a_1, a_2, \cdots, a_{N_A}\}$, and total of N_B targets detected by the method in the set $B = \{b_1, b_2, \cdots, b_{N_B}\}$, the expressions for these two metrics are as

$$\mathrm{TPR} = \frac{\left|\{a_i \mid \exists b_j \in B, \mathrm{Dis}(a_i, b_j) \le D_T\}\right|}{N_A} \tag{8}$$

$$\mathrm{FPR} = \frac{N_B - \left|\{a_i \mid \exists b_j \in B, \mathrm{Dis}(a_i, b_j) \le D_T\}\right|}{N_{ALL} - N_A}, \tag{9}$$

where $\mathrm{Dis}(\cdot)$ is the function that computes the Euclidean distance between two elements, N_{ALL} is the total point number of echo sequences.

The range resolution of radar detection signals can be expressed as

$$\Delta R = \frac{c}{2B}, \tag{10}$$

where c is the speed of light of 3×10^8 m/s, B is the bandwidth of the signal and D_T is the distance threshold used to determine whether the detected target is correct.

According to this equation, when the bandwidth used for both real and simulated data is 20 MHz, the range resolution of the echo can be calculated as 7.5 meters. The interval between adjacent distance sampling points is 6 meters for the sampling rate of 25MHz. Therefore, the difference between the detected target position and the true target position are not supposed to exceed 1 point. Therefore, when calculating TPR and FPR, the distance threshold D_T in (8) and (9) is set to 1.

The area under the curve (AUC) of the ROC curve offers a single-valued metric that encapsulates the overall performance of a classifier across all possible thresholds. The value of AUC spans from 0 to 1. A perfect classifier exhibits an AUC of 1. This implies that for each false positive, there is a corresponding true positive. As the threshold is adjusted, the TPR escalates significantly faster than the FPR. Conversely, a random classifier has an AUC of 0.5, signifying that the TPR and FPR increase at an identical rate as the threshold varies.

The advantage of using AUC as a metric in target detection is that it is threshold-independent. AUC takes into account all possible thresholds, providing a more comprehensive and fair comparison between different object detection algorithms. It gives a better understanding of how well the algorithm can distinguish between positive and negative samples, regardless of the specific threshold used for making the classification decision.

4.3 Experiments on Simulated Data

Prior to presenting the experimental results, it is important to note that all experiments in this paper were implemented using Matlab software. They were conducted not only on simulated data in this section but also on real data in subsequent experiments, all under the same computer configuration (Intel® Core™ i7-4790CPU@3.6GHz).

Under different SNR conditions, the detection thresholds of each method were adjusted to plot the ROC curves, as depicted in Figure 11. The corresponding AUC values, obtained by integrating these curves, are presented in Table 4. Evidently, when the SNR of the target echo is -20 dB, the ROC curves of the CFAR

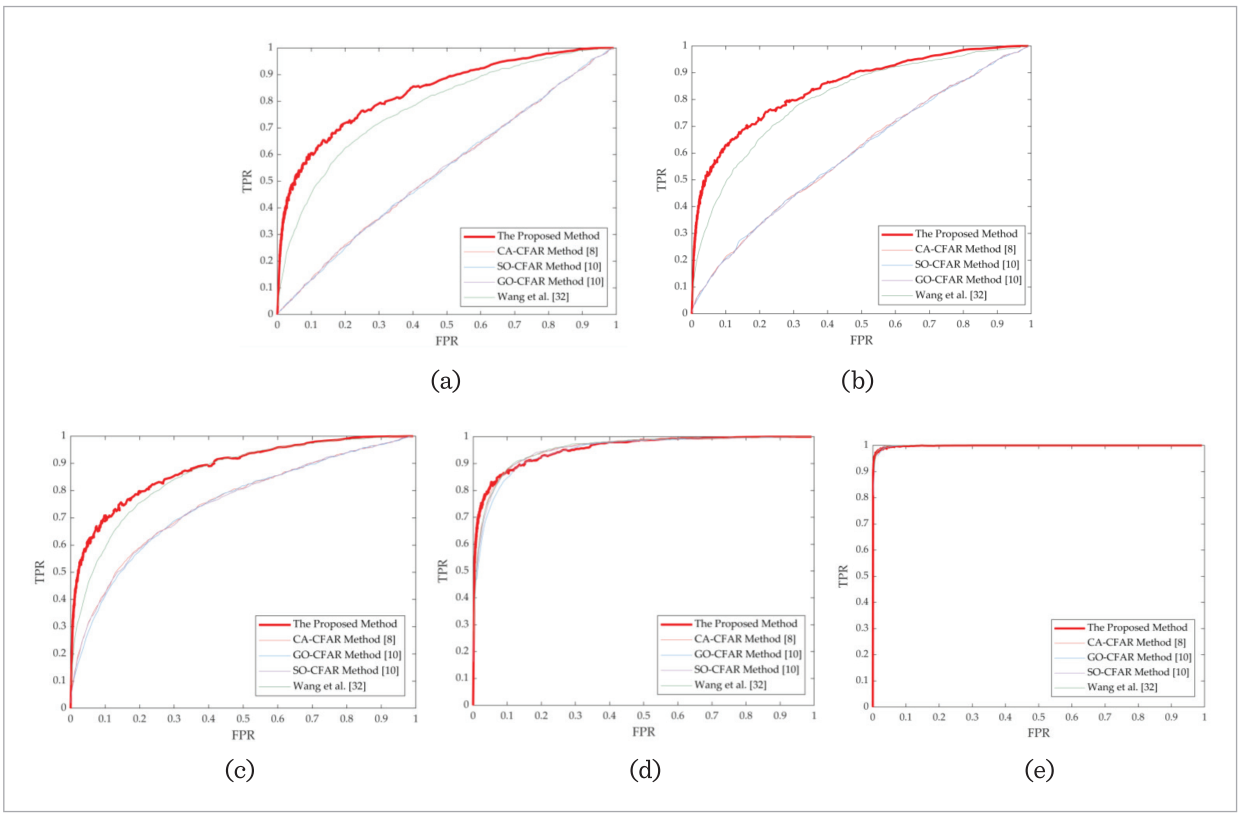
Table 4

Quantitative AUC comparison across SNR levels (-20dB to 0dB) demonstrating performance superiority of proposed method over CFAR baselines and prior neural approaches.

	AUC				
SNR	-20dB	-15dB	-10dB	-5dB	0dB
CA-CFAR [13]	0.5340	0.5916	0.7449	0.9448	0.9870
SO-CFAR [21]	0.5331	0.5898	0.7394	0.9332	0.9759
GO-CFAR [21]	0.5333	0.5906	0.7427	0.9438	0.9986
Wang et al. [28]	0.7734	0.7958	0.8491	0.9447	0.9872
The Proposed Method	0.8256	0.8366	0.8710	0.9446	0.9897

Figure 11

Comparative ROC curves of the proposed method and CFAR detection methods on the simulated data: (a) ROC curves for signals with SNR of -20dB; (b) ROC curve for signals with SNR of -15dB; (c) ROC curves for signals with SNR of -10dB; (d) ROC curves for signals with SNR of -5dB; (e) ROC curves for signals with SNR of 0dB.



methods are nearly straight lines with a 45-degree slope, and the AUC values are merely around 0.53, essentially losing their target detection capabilities. In contrast, the AUC value of the proposed method can be maintained above 0.8. As the SNR of the target echo increases, the AUC values of both the proposed method and the CFAR detection methods gradually improve. Even when the SNR reaches -10 dB, the AUC value of the proposed method is still significantly superior to that of the CFAR detection methods. When the SNR is sufficiently high (above -5dB), all detection methods can accurately detect the target, with AUC values above 0.93. We also compared the proposed method with one existing state-of-the-art neural network methods [28]. The results clearly show that the method proposed consistently outperforms others in terms of detection performance at low SNRs when using only single pulse echo signals.

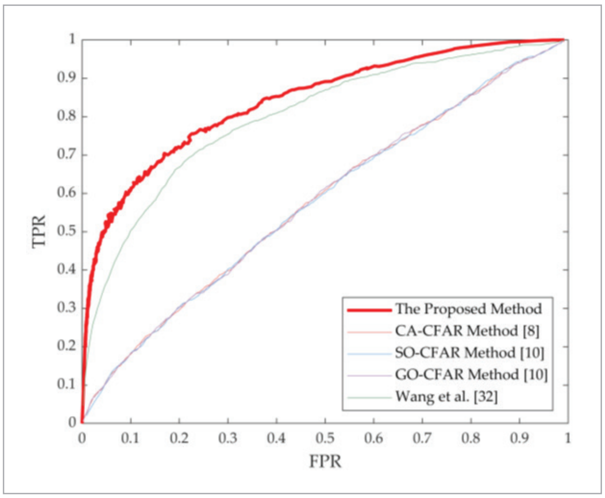
4.4 Experiments on Real Data

For real data, we compare the ROC curves in Figure 12 and calculate the AUC values in Table 5 on the entire real dataset.

As can be seen from Figure 12 and Table 5, the proposed method can achieve higher TPR under all FPR conditions, so that the AUC is around 30% lager than CFAR detection methods which reaches 0.8317.

Figure 12

ROC curves of the proposed method and CFAR variants on real data. The proposed method achieves an AUC of 0.8317, outperforming CA-CFAR (AUC=0.5704) and Wang et al. (AUC=0.7918).



4.5 Ablation Experiments

Figure 13 presents the ROC curves for simulated echo signals with an SNR of -20dB, obtained using the proposed method and the method without the first stage coarse detection network, respectively. As can be observed from the figure, the AUC value is smaller for the method without the coarse detection network compared to the proposed method. This is because, in the proposed method, the first-stage coarse detection delineates the approximate area where the target exists under relatively relaxed conditions. This allows the avoidance of detecting areas that are unlikely to contain the target in the second-stage fine detection. Consequently, under the same TPR conditions in the second-stage fine detection, when the false alarm probability is identical, the proposed method can achieve a reduction in the FPR.

Table 5

Comparison of AUCs of different detection methods on the real data.

	AUC
CA-CFAR [13]	0.5704
SO-CFAR [21]	0.5686
GO-CFAR [21]	0.5699
Wang et al. [28]	0.7918
The Proposed Method	0.8317

Figure 13

Ablation study results comparing ROC of full two-stage architecture versus single-stage implementation at -20dB SNR.

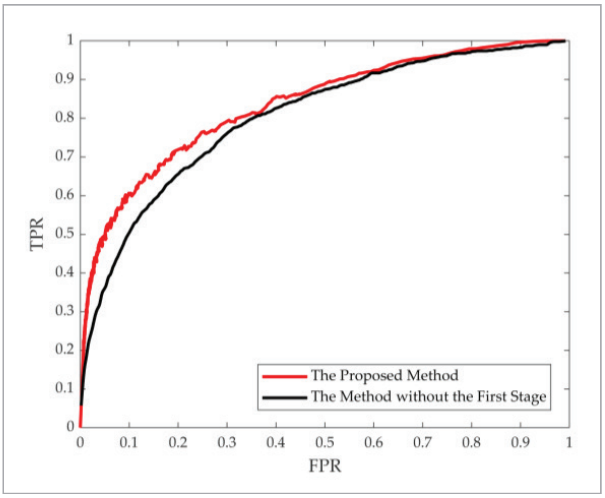


Table 6 presents the average time consumption for 1000 detections on echoes containing a single detection target, using the proposed method and the method without the first-stage. Additionally, the table lists the average single detection time of the CA-CFAR detection method for comparison. As can be seen from the results in Table 6, deep-learning-based methods consume substantially more time than the CA-CFAR method.

Notably, the proposed method, by incorporating a fast-running coarse detection network, reduces the data range required for fine detection, thereby significantly enhancing the algorithm's computational speed. The underlying principle is that the coarse detection stage shrinks the search space by 90%, enabling the fine detection network to process merely 10% of the original data volume. This hierarchical data pruning leads to a 73.7% reduction in inference time.

Table 6

Computational efficiency analysis comparing inference times between full two-stage system, single-stage variant, and traditional CFAR implementation.

	Average Time Consumption
The Proposed Method	7.86ms
Method without the First Stage	28.8ms
Wang et al. [28]	31.17ms
CA-CFAR [13]	1.23ms

4.6 Discussion

The experimental results on both simulated and real data demonstrate that the method proposed in this paper consistently showcases excellent target detection performance in low SNR echoes. It significantly outperforms traditional CA-CFAR detection algorithms. Additionally, ablation experiments fully illustrate the crucial role of the coarse detection network within the two-stage detection network of our proposed method. By employing the coarse detection network, not only is the overall efficiency of the algorithm enhanced, but the ROC curve also indicates a marginal improvement in the algorithm's target detection performance.

Nevertheless, since the algorithm in this study is neural-network-based, it has stringent requirements regarding the format and length of the input data. When the pulse width or bandwidth of the radar signal varies, which causes a change in the length of the echo sequence, the network must be retrained to achieve optimal detection performance.

5. Conclusion

The proposed two-stage DSCResNet framework represents a substantial advancement in radar weak target detection, especially for low altitude drones operating in cluttered environments. Theoretically, this research demonstrates the feasibility of integrating a coarse-to-fine hierarchical processing approach with raw radar signal analysis, thereby circumventing the limitations of traditional methods that rely on preprocessing. By harnessing DSC and residual connections, the method effectively strikes a balance between computational efficiency and feature preservation, filling a crucial void in deep-learning applications for radar systems.

In practical terms, the framework attains a 20% improvement in the detection rate at -20dB SNR compared to the CA-CFAR method. Moreover, it reduces the inference time by a factor of 3.66 relative to single-stage networks, enabling real time deployment in scenarios such as urban surveillance and border security.

The main contributions of this paper are three-fold:

1. Novel architecture: A two-stage design emulating human visual perception. The coarse detection stage, inspired by YOLO's grid regression, rapidly localizes targets, while the fine detection stage refines the classification. This design reduces the computational load by 73.7%.
2. Phase-aware feature extraction: The DSCResNet preserves both amplitude and phase coherence in weak targets. It achieves an AUC of 0.826 at -20dB SNR, representing a 54.7% improvement over CA-CFAR.
3. Real data validation: Robust performance on X-band radar data (AUC=0.8317) under diverse clutter conditions validates the operational viability of the proposed method.

In view of the limitations of the method proposed in this article, which necessitates the use of fixed-size data and targeted training for different signal waveforms, future research will focus on developing dynamic input size adjustment algorithms to adapt to variable pulse widths. Simultaneously, edge optimization is required for the algorithm to deploy the quantized DSCResNet on FPGA platforms, thereby achieving sub-millisecond latency.

In summary, this work bridges the gap between deep learning efficiency and radar signal specificity, presenting a scalable solution for next generation surveillance systems. By addressing both theoretical and practical challenges, it paves the way for future innovations in adaptive, multi-sensor target detection.

References

1. Bi, J., Zhang, G., Yang, C., Jin, L., Zhang, W. Architecture Design of Typical Target Detection and Tracking System in Battlefield Environment. 2021 3rd International Conference on Machine Learning, Big Data and Business Intelligence (MLBDBI). IEEE, 2021, 473, 477. https://doi.org/10.1109/MLBDBI54094.2021.00096
2. Chen, P., Wang, Y., Liu, H. GCN-YOLO: YOLO Based on Graph Convolutional Network for SAR Vehicle Target Detection. IEEE Geoscience and Remote Sensing Letters, 2024, 21, 1-5. https://doi.org/10.1109/LGRS.2024.3424875
3. Chen, X., Liu, K., Zhang, Z. A PointNet-Based CFAR Detection Method for Radar Target Detection in Sea Clutter. IEEE Geoscience and Remote Sensing Letters, 2024, 21, 1-5. https://doi.org/10.1109/LGRS.2024.3363041
4. Chen, X., Su, N., Huang, Y., Guan, J. False-Alarm-Controllable Radar Detection for Marine Target Based on Multi Features Fusion via CNNs. IEEE Sensors Journal, 2021, 21, 9099-9111. https://doi.org/10.1109/JSEN.2021.3054744
5. Deng, Z., Sun, H., Zhou, S., Zhao, J. Learning Deep Ship Detector in SAR Images from Scratch. IEEE Transactions on Geoscience and Remote Sensing, 2019, 57, 4021-4039. https://doi.org/10.1109/TGRS.2018.2889353
6. Gu, F., Zhang, L., Zheng, S., Chen, J., Yue, K., Zhao, Z., Yang, X. Detection of Radar Pulse Signals Based on Deep Learning. IEEE Open Journal of Signal Processing, 2024, 5, 991-1004. https://doi.org/10.1109/OJSP.2024.3435703
7. Han, K., Wang, Y., Tian, Q., Guo, J., Xu, C., Xu, C. GhostNet: More Features from Cheap Operations. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020, 1580, 1589. https://doi.org/10.1109/CVPR42600.2020.00165
8. Ji, G., Song, C., Huo, H. Detection and Identification of Low-Slow-Small Rotor Unmanned Aerial Vehicle Using Micro-Doppler Information. IEEE Access, 2021, 9, 99995-100008. https://doi.org/10.1109/ACCESS.2021.3096264
9. Jiang, P., Ergu, D., Liu, F., Cai, Y., Ma, B. A Review of YOLO Algorithm Developments. Procedia Computer Science, 2022, 199, 1066-1073. https://doi.org/10.1016/j.procs.2022.01.135
10. Jiménez, L. P. J., García, F. D. A., Alvarado, M. C. L., Fraidenraich, G., De Lima, E. R. A General CA-CFAR Performance Analysis for Weibull-Distributed Clutter Environments. IEEE Geoscience and Remote Sensing Letters, 2022, 19, 1-5. https://doi.org/10.1109/LGRS.2022.3187554
11. Kerbaa, T. H., Mezache, A., Oudira, H. Improved Decentralized SO-CFAR and GO-CFAR Detectors via Moth Flame Algorithm. 2022 International Conference of Advanced Technology in Electronic and Electrical Engineering (ICATEEE). IEEE, 2022, 1, 5. https://doi.org/10.1109/ICATEEE57445.2022.10093725
12. Kim, M., Kim, H. I., Ro, Y. M. Prompt Tuning of Deep Neural Networks for Speaker-Adaptive Visual Speech Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2025, 47, 1042-1055. https://doi.org/10.1109/TPAMI.2024.3484658
13. Kuang, C., Wang, C., Wen, B., Hou, Y., Lai, Y. An Improved CA-CFAR Method for Ship Target Detection in Strong Clutter Using UHF Radar. IEEE Signal Processing Letters, 2020, 27, 1445-1449. https://doi.org/10.1109/LSP.2020.3015682
14. Liu, Y., Lin, M., Mo, Y., Wang, Q. SAR-Optical Image Matching Using Self-Supervised Detection and a Transformer-CNN-Based Network. IEEE Geoscience and Remote Sensing Letters, 2024, 21, 1-5. https://doi.org/10.1109/LGRS.2024.3355472
15. Liu, Z., Kang, X., Ren, F. Dual-TBNet: Improving the Robustness of Speech Features via Dual-Transformer-BiLSTM for Speech Emotion Recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31, 2193-2203. https://doi.org/10.1109/TASLP.2023.3282092

16. Luo, H., Gao, P., Li, J., Zeng, D. Gaze Estimation Based on the Improved Xception Network. IEEE Sensors Journal, 2024, 24, 8450-8464. https://doi.org/10.1109/JSEN.2024.3359085
17. Lyu, B., Fan, C., Ming, Y., Zhao, P., Hu, N. En-HACN: Enhancing Hybrid Architecture with Fast Attention and Capsule Network for End-to-End Speech Recognition. IEEE/ACM Transactions on Audio, Speech, and Language Processing, 2023, 31, 1050-1062. https://doi.org/10.1109/TASLP.2023.3245407
18. Mai, C., Liao, R., Ren, J., Gong, Y., Zhang, K., Zhang, C. MobileNet-Based IoT Malware Detection with Opcode Features. Journal of Communications and Information Networks, 2023, 8, 221-230. https://doi.org/10.23919/JCIN.2023.10272350
19. Medeiros, D. S., García, F. D. A., Machado, R., Santos Filho, J. S. C., Saotome, O. CA-CFAR Performance in K-Distributed Sea Clutter with Fully Correlated Texture. IEEE Geoscience and Remote Sensing Letters, 2023, 20, 1-5. https://doi.org/10.1109/LGRS.2023.3238169
20. Qiu, M., Wang, J., Wu, G., Zhang, P. Single Pulse Target Detection Method Based on a Two-Stage Convolutional Neural Network. IET International Radar Conference (IRC 2023). 2023, 1012, 1018. https://doi.org/10.1049/icp.2024.1225
21. Sahal, M., Said, Z. A., Putra, R. Y., Kadir, R. E. A., Firmansyah, A. A. Comparison of CFAR Methods on Multiple Targets in Sea Clutter Using SPX-Radar-Simulator. 2020 International Seminar on Intelligent Technology and Its Applications (ISITIA). IEEE, 2020, 260, 265. https://doi.org/10.1109/ISITIA49792.2020.9163697
22. Sahed, M., Kenane, E., Khalfa, A., Djahli, F. Exact Closed-Form Pfa Expressions for CA- and GO-CFAR Detectors in Gamma-Distributed Radar Clutter. IEEE Transactions on Aerospace and Electronic Systems, 2022, 59, 4674-4679. https://doi.org/10.1109/TAES.2022.3232101
23. Sangaiah, A. K., Yu, F. N., Lin, Y. B., Shen, W. C., Sharma, A. UAV T-YOLO-Rice: An Enhanced Tiny YOLO Network for Rice Leaves Diseases Detection in Paddy Agronomy. IEEE Transactions on Network Science and Engineering, 2024, 11, 5201-5216. https://doi.org/10.1109/TNSE.2024.3350640
24. Shafiq, M., Gu, Z. Deep Residual Learning for Image Recognition: A Survey. Applied Sciences, 2022, 12, 8972. https://doi.org/10.3390/app12188972
25. Shao, S., Zhu, W., Li, Y. Radar Detection of Low-Slow-Small UAVs in Complex Environments. 2022 IEEE 10th Joint International Information Technology and Artificial Intelligence Conference (ITAIC). IEEE, 2022, 10, 1153-1157. https://doi.org/10.1109/ITAIC54216.2022.9836542
26. Tian, J., Wang, C., Cao, J., Wang, X. Fully Convolutional Network-Based Fast UAV Detection in Pulse Doppler Radar. IEEE Transactions on Geoscience and Remote Sensing, 2024, 62, 1-12. https://doi.org/10.1109/TGRS.2024.3358956
27. Wagner, S., Johannes, W., Qosja, D., Bruggenwirth, S. Small Target Detection in a Radar Surveillance System Using Contractive Autoencoders. IEEE Transactions on Aerospace and Electronic Systems, 2023, 60, 51-67. https://doi.org/10.1109/TAES.2023.3253469
28. Wang, C., Tian, J., Cao, J., Wang, X. Deep Learning-Based UAV Detection in Pulse-Doppler Radar. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60, 1-12. https://doi.org/10.1109/TGRS.2021.3104907
29. Wang, L., Tang, J., Liao, Q. A Study on Radar Target Detection Based on Deep Neural Networks. IEEE Sensors Letters, 2019, 3, 1-4. https://doi.org/10.1109/LSENS.2019.2896072
30. Wen, L., Ding, J., Xu, Z. Multiframe Detection of Sea-Surface Small Target Using Deep Convolutional Neural Network. IEEE Transactions on Geoscience and Remote Sensing, 2022, 60, 1-16. https://doi.org/10.1109/TGRS.2021.3122515
31. Wu, L., Xu, Y., Hou, J., Chen, C. L. P., Liu, C. L. A Two-Level Rectification Attention Network for Scene Text Recognition. IEEE Transactions on Multimedia, 2022, 25, 2404-2414. https://doi.org/10.1109/TMM.2022.3146779
32. Wu, M., Li, M., Shi, H., Cheng, X., Rao, B., Wang, W. Using Range-Doppler Spectrum-Based Deep Learning Method to Detect Radar Target in Interrupted Sampling Repeater Jamming. IEEE Sensors Journal, 2023, 23, 29084-29096. https://doi.org/10.1109/JSEN.2023.3286893
33. Xie, Y., Tang, J., Wang, L. Radar Target Detection Using Convolutional Neural Network in Clutter. 2019 IEEE International Conference on Signal, Information and Data Processing (ICSIDP). IEEE, 2019, 1, 6. https://doi.org/10.1109/ICSIDP47821.2019.9173064
34. Xiong, X., He, M., Li, T., Zheng, G., Xu, W., Fan, X., Zhang, Y. Adaptive Feature Fusion and Improved Attention Mechanism Based Small Object Detection for UAV Target Tracking. IEEE Internet of Things Journal, 2024, 11, 21239-21249. https://doi.org/10.1109/JIOT.2024.3367415

35. Yang, Z., Zhou, H., Tian, Y., Zhao, J. Improved CFAR Detection and Direction Finding on Time-Frequency Plane with High-Frequency Radar. IEEE Geoscience and Remote Sensing Letters, 2021, 19, 1-5. https://doi.org/10.1109/LGRS.2021.3066522
36. Yasir, M., Jianhua, W., Mingming, X., Hui., S., Zhe, Z., Shanwei, L., Colak, A. T. I., Hossain, M. S. Ship Detection Based on Deep Learning Using SAR Imagery: A Systematic Literature Review. Soft Computing, 2023, 27, 63-84. https://doi.org/10.1007/s00500-022-07522-w
37. Ye, T., Qin, W., Zhao, Z., Gao, X., Deng, X., Ouyang, Y. Real-Time Object Detection Network in UAV-Vision Based on CNN and Transformer. IEEE Transactions on Instrumentation and Measurement, 2023, 72, 1-13. https://doi.org/10.1109/TIM.2023.3241825
38. Zebiri, K., Mezache, A. Radar CFAR Detection for Multiple-Targets Situations for Weibull and Log-Normal Distributed Clutter. Signal, Image and Video Processing, 2021, 15, 1671-1678. https://doi.org/10.1007/s11760-021-01905-6
39. Zhang, X., Su, H., Zhang, C., Gu, X., Tan, X., Atkinson, P. M. Robust Unsupervised Small Area Change Detection from SAR Imagery Using Deep Learning. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 173, 79-94. https://doi.org/10.1016/j.isprsjprs.2021.01.004
40. Zhao, Y., Sun, T., Zhang, J., Gao, M. GAN-CNN-Based Moving Target Detector for Airborne Radar Systems. IEEE Sensors Journal, 2024, 24, 21614-21627. https://doi.org/10.1109/JSEN.2024.3397731