electronics

MDPI

Article

RaTDet: A Marine Radar Transformer Network for End-to-End Target Detection

Huaxing Kuang [1,2,3,*], Haocheng Yang [2,3] and Luxi Yang [1]

[1] School of Information Science and Engineering, Southeast University, Nanjing 210018, China; lxyang@seu.edu.cn
[2] National Key Laboratory of Electromagnetic Effect and Security on Marine Equipment, Nanjing 211100, China; yhcnanking@126.com
[3] Nanjing Marine Radar Institute, Nanjing 211153, China
* Correspondence: 230188074@seu.edu.cn

Abstract

Recent advancements in deep learning have shown considerable potential to enhance radar target detection, particularly in improving detection probability under complex environmental conditions. However, existing deep learning approaches largely operate in the real number domain, neglecting the complex-valued nature of radar data, and often inherit vision-oriented architectures that fail to address radar-specific challenges—such as sparse target echoes, the necessity for phase preservation, and constraints imposed by scanning radar systems. Meanwhile, conventional radar signal processing methods, including CA-CFAR, are limited by their dependence on idealized statistical models and often underperform in dynamic and cluttered electromagnetic environments.To overcome these issues, this paper proposes Radar Transformer for Detection (RaTDet), an end-to-end detection network that integrates complex-valued convolutional neural networks (CNNs) and Transformers. RaTDet fully leverages complex-valued data to preserve critical phase and amplitude information, enabling automated feature learning directly from raw radar signals. The model operates effectively with very few pulses, making it suitable for resource-constrained scenarios, and can serve as a pre-trained foundation model for various radar downstream tasks. Experimental results demonstrate that RaTDet achieves excellent detection performance, characterized by high detection probability (Pd) and low false alarm rate (Pfa), outperforming both traditional signal processing and conventional deep learning methods. This work bridges the gap between deep learning and radar signal processing, offering a flexible and powerful network for next-generation radar systems.

Keywords: radar target detection; complex-valued neural networks; end-to-end architecture

Check for updates

Academic Editor: Raffaele Solimene

Received: 4 March 2026
Revised: 24 April 2026
Accepted: 27 April 2026
Published: 2 May 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Maritime radar target detection serves as the cornerstone of modern ocean surveillance, maritime security, search-and-rescue operations, and naval defense, enabling the all-weather, day-and-night monitoring of vessels, small targets, and dynamic oceanic scenes under harsh meteorological and electromagnetic conditions [1,2]. Conventional detection frameworks, represented by constant false-alarm rate (CFAR) detectors and model-driven statistical methods, rely heavily on handcrafted features and predefined clutter models, which suffer from performance degradation in non-stationary, non-Gaussian sea clutter, low signal-to-clutter ratio (SCR) scenarios, and when facing small, low-radar-cross-section (RCS), or highly maneuverable targets [3,4]. These limitations have motivated

a paradigm shift toward data-driven intelligent detection, which leverages deep learning, multi-dimensional feature fusion, and self-adaptive learning to mine implicit scattering characteristics and contextual information from radar echoes, thereby breaking the bottlenecks of traditional algorithms in complex marine environments [5,6]. In recent years, the field of intelligent maritime radar target detection has witnessed rapid advancement driven by annotated datasets, and novel neural network architectures [7–9]. Previous studies [10] have shown that deep learning networks can also be effectively applied to radar target detection tasks when radar data are represented in image format, with application in radar automatic target recognition [10], radar target detection [11–13], cognitive radar [14], marine target detection [15,16], and synthetic aperture radar target detection [17]. These techniques enable high-performance detection with low false alarm probability (Pfa) and high detection probability (Pd).

It is important to note, however, that current research on neural network architectures is largely driven by applications in computer vision and natural language processing. Designing effective neural networks for radar applications may require deeper radar-specific insights. Several key differences between visual images and radar data are noteworthy. First, radar echoes—formed by waves scattered from targets—typically occupy only a small fraction of pixels in the radar image, making radar target detection inherently a small-object detection problem from a computer vision perspective. Second, vision-based algorithms generally process entire images, which is often incompatible with the sequential data acquisition of scanning radars. Third, unlike real-valued visual images, modern radar systems typically collect complex-valued data. In this case, a complex-valued neural network, whose computations adhere to the rules of complex multiplication, constrains the parameter optimization space to a more physically plausible subspace, thereby enhancing the network's efficiency in extracting features from complex-valued data.

On the other hand, conventional radar signal processing is increasingly challenged by complex electromagnetic environments. Although classic methods such as Cell-Averaging Constant False Alarm Rate (CA-CFAR) are valued for their simplicity, stability, and effectiveness, they are designed under the assumption of Rayleigh-distributed clutter. Despite the development of alternative detectors [18–26], these closed-form solutions often struggle to adapt to the dynamic and heterogeneous interference conditions faced by modern marine radars. This mismatch between theoretical assumptions and practical conditions leads to degraded detection performance. Deep neural networks, with their strong nonlinear modeling capacity, offer a promising approach to mitigate these issues.

In this paper, we propose RaTDet (Radar Transformer for Detection), a general framework for end-to-end marine radar target detection. The architecture integrates complex-valued data processing, classical radar signal processing techniques, and deep neural networks, combining the strengths of both traditional workflows and deep learning to achieve high Pd and low Pfa. The main contributions of this work are summarized as follows:

(1) We introduce a fully complex-valued network architecture that comprehensively preserves phase information from raw complex-value radar data. This allows the model to exploit more complete information from the input of complex-value radar data, leading to richer feature representations and improved detection performance.
(2) Unlike classical feature-based detectors that often require extensive pulse integration [19] the proposed model operates effectively with only a small number of successive coherent pulses (a pulse group). This makes it particularly suitable for pulse resource-constrained scenarios where rapid detection is essential.
(3) The pulse-group-based processing design offers high flexibility and compatibility across various types of pulse-Doppler radars. This adaptability facilitates deployment in diverse operational contexts without requiring major architectural modifications.

The rest of the paper is organized as follows. Section 2 involves an introduction to the motives in the design of RatDet. Section 3 addresses the typical framework of RaTDet, and the relative theories underlying its blocks. Then numerical results generated by four distinct models based on RaTDet are presented in Section 4. The performance of these models is also compared to traditional methods utilized in radar signal processing. And finally, in Section 5 we conclude our paper with a summary and a further research proposal of RaTDet.

2. Related Work

Radar maritime target detection is a core technology for maritime situational awareness, with an irreplaceable role in national maritime security, coastal defense, marine traffic management, and marine resource development. Over the past decade, radar maritime target detection has undergone a progressive evolution from traditional statistical methods to intelligent deep learning-based approaches. This section systematically reviews the related work in radar maritime target detection, focusing on three core technical categories: statistical-based detection methods, deep convolutional neural network (CNN)-based detection methods, and Transformer-based detection methods.

Constant False Alarm Rate (CFAR) detectors are the foundational paradigm for maritime radar target detection, as they maintain a stable false alarm rate by adaptively estimating sea clutter statistics. Classical CFAR variants, such as Cell-Averaging CFAR, Order-Statistic CFAR, and Greatest-Of CFAR, have been widely studied but suffer from performance degradation in heterogeneous sea clutter and nearshore regions with land interference. To address non-Gaussian sea clutter modeling, Gao and Liu compared multiple statistical models for SAR sea clutter and validated that the proposed model significantly improves CFAR detection accuracy by better fitting heavy-tailed clutter distributions, outperforming traditional Gaussian and K-distribution models in low signal-to-clutter ratio scenarios [27]. To handle nearshore land–sea mixed clutter, Wen and Lu introduced a space-time adaptive filtering preprocessing step combined with Generalized Extreme Value distribution modeling, which suppresses non-stationary clutter and enhances CFAR threshold robustness for offshore ship detection [28].

Deep CNN-based detection methods have become the mainstream paradigm of radar maritime target detection. SLA-Net is proposed for SAR ship detection guided by a hierarchical attention mechanism, which innovatively integrates a dedicated land–sea segmentation branch with a hierarchical attention mechanism [29]. The land–sea segmentation branch generates precise sea–land masks to distinguish sea regions from land clutter, while the hierarchical attention module dynamically fuses multi-scale detection features with the masks, enabling the network to focus on sea regions and suppress land interference effectively. For low-SCR small and weak target detection, Zhang et al. proposed the Laplace and LBP feature guided SAR ship detection method with an adaptive feature enhancement block, which adopts a hybrid paradigm combining traditional handcrafted features with deep CNN features [30]. This method extracts Laplacian edge and LBP texture features, then fuses them with deep CNN features through an Adaptive Feature Enhancement Block, which adaptively weights multi-scale feature maps to suppress noise and highlight target features. Liu et al. proposed GCBANet: A Global Context Boundary-Aware Network for SAR Ship Instance Segmentation, which integrates a Global Context Information Modeling Block and a Boundary-Aware Box Prediction Block to improve the accuracy of target localization and pixel-level contour extraction [31]. Additionally, lightweight CNN-based methods have made remarkable progress to achieve both high accuracy and efficiency on SAR object detection. Zhuang Zhou proposed a lightweight detector with knowledge distillation, reducing model parameters while maintaining detection performance [32].

Transformer-based detection methods have emerged as a new research hotspot in recent years, addressing the limitation of CNNs in capturing long-range feature dependencies, which is crucial for multi-scale and occluded maritime target detection. Unlike CNNs that rely on local convolution operations, Transformers use self-attention mechanisms to model global feature relationships, significantly improving the robustness of detection in complex maritime environments. To tackle multi-scale and arbitrarily oriented ship targets, Chen et al. designed a lightweight Transformer detector with deformable attention, enabling adaptive feature sampling for small and rotated ships while reducing computational overhead, suitable for real-time maritime surveillance [33]. A ship detection Transformer is proposed with integrated key scattering point feature aggregation and context refinement, which effectively suppressed false alarms from inshore clutter and improved localization accuracy for sparse scattering ship targets in SAR images, validated on public SAR datasets with state-of-the-art performance [34].

3. System Model

3.1. An Enlightening Review of Typical Methods

RaTDet is designed not only to achieve high-performance radar target detection but also to incorporate principles from traditional signal processing into a neural network framework, thereby maintaining alignment with conventional radar workflows. This section provides a concise review of three key conventional methods: CA-CFAR, feature-based detection, and moving target detection (MTD).

In CA-CFAR, a cell under test (CUT)—located at a specific range bin within a pulse—is evaluated along with a set of adjacent reference cells. The core principle of CA-CFAR is to assess whether the CUT contains a target echo by comparing its energy level against those of the reference units. Specifically, the average background energy level, denoted as σ, is estimated from the reference cells. This value is then multiplied by a threshold factor τ, which is predetermined based on the desired false alarm probability (Pfa), to form the detection threshold T as follows:

$$\begin{cases} E_{CUT} \geq T & \text{CUT belongs to a target echo.} \\ E_{CUT} < T & \text{CUT does not belong to a target echo.} \end{cases}, \tag{1}$$

where E_{CUT} indicates the energy of CUT.

CA-CFAR can be regarded as a specialized instance of feature-based detection. In general, feature-based detection methods analyze radar data sequences across different range bins. For each sequence within a specific range bin, a set of handcrafted features is extracted to construct a feature vector, representing the characteristics of that range bin in a high-dimensional feature space. Since target echoes and background clutter typically originate from distinct sources, the corresponding feature points are often separable in this space. Consequently, by leveraging these feature vectors and a suitable classifier, feature-based detection can effectively determine whether a range bin contains a target.

In the case of CA-CFAR, the energy of the cell under test (CUT) serves as the selected feature, and a manually designed threshold T functions as the classifier. In other approaches, such as those referenced in [35–41], additional features—including Hurst exponents, information entropy, and others—have been proposed, alongside classifiers such as support vector machines (SVM) and decision trees.

Moving target detection (MTD), on the other hand, enhances target echoes by applying Fourier-based filtering across pulse groups, effectively separating targets according to their Doppler shifts. CFAR detection is then independently applied within each velocity channel, and the results are aggregated to produce the final detection outcome. MTD exemplifies how signal processing techniques can amplify discriminative

features—specifically the energy discrepancy between targets and background clutter—to improve detection performance.

3.2. Summarized General Workflow of Radar Target Detection

It is evident that the core methodology of radar target detection consists of two fundamental steps: first, extracting discriminative features from radar echoes, and second, employing a classifier to partition the decision region for distinguishing targets from background clutter. In essence, the role of radar signal processing is to magnify the separation between target and background features in this representation space. Thus, from a feature-oriented perspective, the conventional radar target detection workflow can be formalized as a two-stage process as in Figure 1.

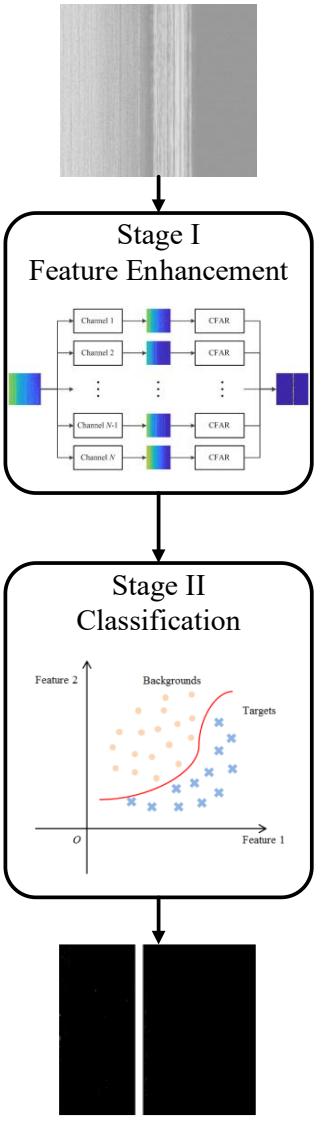


Figure 1. The two-stage procedure of traditional radar target detection.

In the feature enhancement stage, signal processing techniques—such as moving target detection (MTD) or filtering—are applied to the raw radar data to isolate and amplify characteristics that help differentiate targets from the background. Energy remains the most widely used feature, although other attributes such as Doppler shift, entropy, or statistical properties may also be employed depending on the operational context. The objective is to maximize the distinguishability between target and clutter in feature space.

In the classification stage, a classifier—such as CA-CFAR, support vector machine (SVM), or other decision rules—is used to divide the feature space into target-present and target-absent regions. The design of this classifier is typically based on statistical assumptions or expert-defined rules, often tailored to specific environmental conditions (e.g., urban clutter, sea waves, or low signal-to-noise ratio scenarios).

This well-established workflow is the product of decades of research and refinement by radar experts. Its simplicity, robustness, and effectiveness have been thoroughly validated in practice, establishing it as a highly reliable approach in the field. Nevertheless, with the increasing complexity of electromagnetic environments, human-engineered features may no longer be sufficient to maintain high detection performance, often resulting in elevated false alarm rates or the reduced probability of detection. Deep neural networks offer a promising pathway to address these emerging challenges.

4. Proposed RaTDet Network

4.1. The Radar Transformer for Detection

Given that the traditional radar target detection workflow has proven effective over many years, RaTDet is not intended to completely replace it. Rather, it is designed to enhance flexibility and learn directly from raw complex-valued data. Accordingly, and in alignment with the conventional process, RaTDet consists of two core modules: pulse embedding and sequence analysis.

The pulse embedding module processes a group of successive pulses into a multi-channel sequence of feature vectors along the range bin dimension using neural networks. This module functionally corresponds to the signal processing stage in the traditional workflow, acting as a feature extraction step. As previously noted, feature extraction in conventional systems relies heavily on human expertise. However, human-engineered feature design has notable limitations. First, it is challenging for experts to identify highly discriminative features that reliably separate target echoes from background clutter. Second, aside from a limited set of simple attributes, most handcrafted features require a large number of pulses within a pulse group to be effective, a requirement often impractical for scanning radars due to resource and time constraints.

In contrast, the pulse embedding module in RaTDet automatically extracts a rich set of features from a limited number of pulses, without human intervention. These features, which are often beyond human ability to design or interpret, enable more adaptive and efficient detection in complex electromagnetic environments.

As is shown in Figure 2, the size of the input pulse group is designated as $N_{in} \times C_{in} \times M$, where N_{in} is the number of pulses in a pulse group, C_{in} is the number of features for a certain range bin, and M is the number of range bins. For the input data, C_{in} equals 1 for the input feature is the value of signal itself. After being processed by a feature extracting neural network, the output is a matrix of the size of $N_{out} \times C_{out} \times M$, $N_{out} = 1$ is the number of vector sequences along the dimension of range bins, C_{out} is the number of features extracted, and M is the number of range bins.

The subsequent module is sequence analysis, transforming the feature vector sequence acquired by the pulse embedding module into a binary vector indicating the location of target range bins. Rather than a classification neural network, the sequence analysis is a counterpart of CA-CFAR detector, trying to figure out various relations between different range bins and directly accomplishing the mapping process. As is shown in Figure 2, the input feature vector sequence of size $1 \times C_{out} \times M$ is transformed into a $1 \times 1 \times M$ binary sequence, with the black block indicating where the target lies.

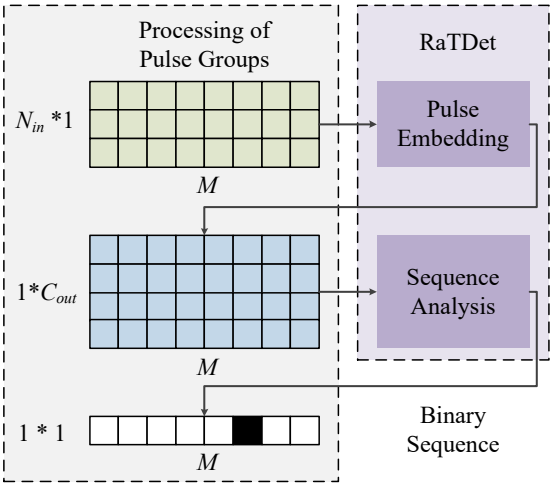


Figure 2. Dataflow of RaTDet.

4.2. Realization of RaTDet

Given the complex-valued nature of radar data, it is essential that the neural network is capable of processing such data directly. This approach ensures that all information inherent in the radar signals—including phase—is preserved and made available for analysis. To this end, complex-valued neural networks are adopted in this work.

The pulse embedding module is implemented as a complex-valued convolutional neural network (CV-CNN), in which all weights are complex numbers. For the activation function, we employ Cardioid ReLU, defined as follows:

$$f(z) = (\frac{1+\cos\phi}{2})z, \tag{2}$$

where z is the input complex number and ϕ is its argument. When ϕ ranges from 0 to π, the factor $(1+\cos\phi)/2$ decreases from 1 to 0, hence we call it an extended version of real-valued ReLU. Additionally, in this paper, as a demonstration, the number of a pulse group is 16. Therefore, four layers containing 1D complex-valued convolution operator are implemented for down-sampling. As an example, the complete structure of the pulse embedding module is shown as Figure 3. The module comprises four complex-valued convolutional layers which are each succeed by Cardioid ReLU as the activation function and whose number of output channels are 16, 64, 256, and 512 respectively. The kernel size, stride, and padding for all the convolution operators are 3×1, 2×1 and 1×0. Through the four-layer pulse embedding module, a 16-pulse group of data with a single channel are transformed into a 512-channel feature vector sequence along the dimension of range bins.

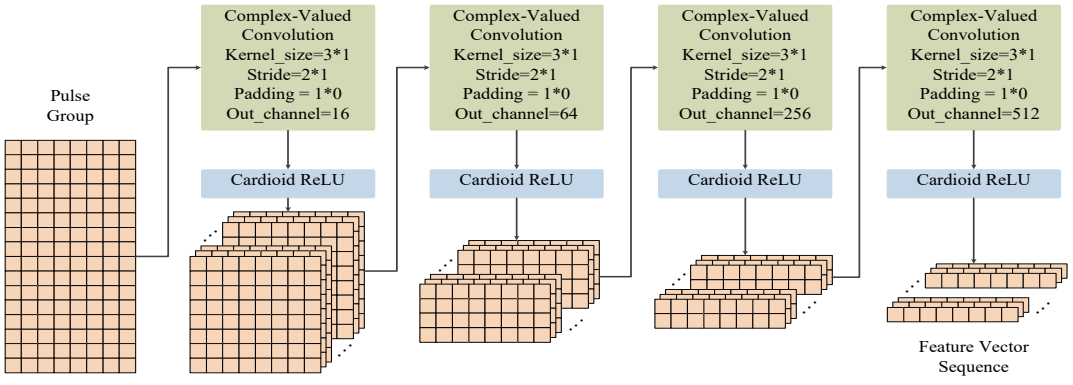


Figure 3. Structure of the pulse embedding module.

The sequence analysis module employs Transformer architecture, which has recently demonstrated exceptional performance across numerous domains. However, the original Transformer design introduced in [42] operates on real-valued data and is therefore incompatible with the complex-valued feature sequences produced by the pulse embedding module. A straightforward solution would be to apply a modulus operation to convert the complex-valued vectors into real-valued ones, thereby enabling standard Transformer processing. Yet, this method would inevitably result in substantial information loss.

Fortunately, recent work by Florian Eilers and Xiaoyi Jiang [43] introduced a set of foundational components for constructing complex-valued Transformer architecture. In RaTDet, we adopt these building blocks to implement a complex-valued Transformer encoder as shown in Figure 4. Unlike real-valued attention that discards phase by operating solely on magnitudes, the complex-valued multi-head attention mechanism preserves phase information inherently through its algebraic formulation. Specifically, the attention score between a query q and a key k is computed via the complex-valued inner product $q^H k$, which retains both the relative magnitude and the relative phase difference of the two complex vectors. This phase-sensitive compatibility measure is then normalized by a complex-valued Softmax function that preserves angular relationships while scaling the modulus. Subsequently, the weighted combination of complex-valued value vectors v is performed using complex multiplication and addition, ensuring that both amplitude and phase contributions from each sequence element are faithfully aggregated. By maintaining the full complex-valued representation throughout the attention mechanism, the encoder is able to exploit phase-coherent relationships within the feature sequence while harnessing the powerful representational capacity of the Transformer [43].

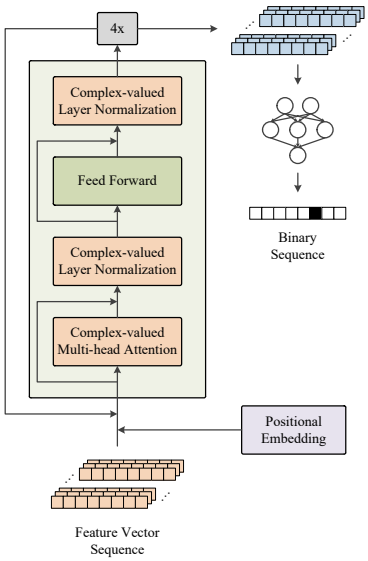


Figure 4. Sequence analysis module.

As a demonstration, four complex-valued Transformer encoders with an 8-head attention mechanism and a 512-dimensional feed-forward networks are stacked up and succeeded by a fully connected layer to convert the 512-channel feature vector sequence mentioned above into the binary sequence required. The layer normalization is the same as that is proposed in [42].

Therefore, the overall architecture of our proposed RatDeT is displayed below in Figure 5. This complex-valued convolutional neural network takes pulse group data as input and outputs target detection results. It primarily consists of two main components: pulse group feature extraction and feature sequence analysis, which is described in detail as above.

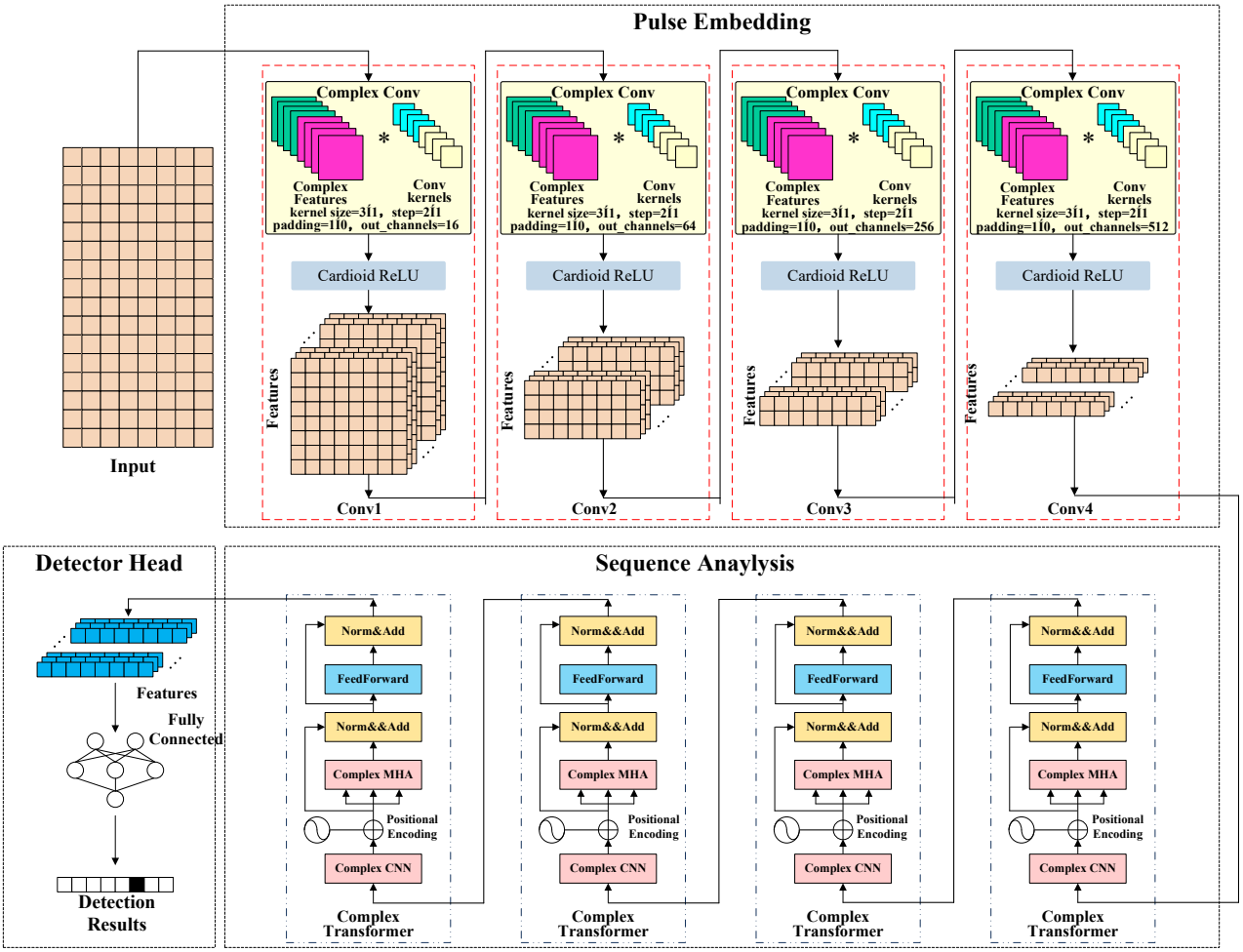


Figure 5. Network structure of RaTDet. Note: * represents convolution.

5. Experiments

5.1. Experiment Settings

In this paper, the proposed RaTDet utilizes 16 pulse echoes to construct range profiles. The input data of the detector is preprocessed using max-value normalization. The experimental hardware environment consists of an NVIDIA A6000 GPU and an Intel(R) Core(TM) i9-13900K CPU, while the software environment includes Python 3.10 and the deep learning framework PyTorch 2.0.0. The loss function employed is the diceloss function, and the Adam optimizer is used to minimize the loss function for optimizing the weights of RaTDet. The learning rate is set to 10^{-6}. For the experimental dataset, the batch-size are configured as 32, with the training iterations set to 400 epochs.

5.2. Dataset

The detection performance of RaTDet is evaluated using the publicly available radar maritime surveillance dataset from the Journal of Radars [44]. The radar maritime surveillance dataset was collected in Yantai, Shandong, China, and the parameters of the radar employed are also listed in Table 1. In this study, the files "20210104163843_01_staring.mat" and "20210106150614_01_staring.mat" are used to build the training set, while "2021010150614_03_staring.mat" is utilized to construct the test set.

Table 1. Property of the dataset.

Property	Value
Bandwidth(MHz)	25
Frequency(GHz)	9.30–9.50
Pulse Repetition Frequency(Hz)	1600
Pulse Width(us)	3.0

200,000 samples are generated randomly from the labeled dataset, 70 percent of which are utilized for training and the rest are for validation. Each sample comprises 16 pulses, 1 channel and 256 range bins. It is notable that the number of range bins may be increased if more cuda memory is available. Here Figure 6 is an example of the labeled sample. The first 16 rows are the original data selected from the files, while the 17th row is a binary label indicating the range bin in which the target lies.

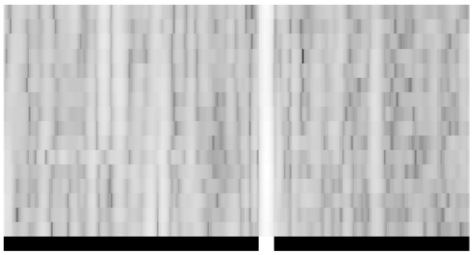


Figure 6. Demonstration of training sample.

5.3. Experiment Design

To explore the effectiveness of the proposed RaTDet, in the experiment stage we propose four variations of the aforementioned architecture, namely CNN-Real-RaT, CNN-CV-RaT, CNN-CV-RaT-X and CNN-CV-RaT-mini. The differences between these models are listed as in Table 2 below.

Table 2. Four variations of model architectures.

Model	Pulse Embedding	Sequence Analysis	Extra Processing
CNN-Real-RaT	Complex-valued CNN	Real-valued Transformer with four encoder blocks	--
CNN-CV-RaT	Complex-valued CNN	Complex-valued Transformer with four encoder blocks	--
CNN-CV-RaT-X	Complex-valued CNN	Complex-valued Transformer with four encoder blocks	With data enhancement in Doppler domain
CNN-CV-RaT-X-mini	Complex-valued CNN	Complex-valued Transformer with two encoder blocks	With data enhancement in Doppler domain

Among these four variational models, CNN-CV-RaT is the standard model proposed in Section 2.

In CNN-Real-RaT, the complex-valued Transformer is restored to a standard real-valued Transformer. When input into the real-valued Transformer, the complex-valued features generated by the complex-valued CNN are converted to real numbers by taking their modulo.

While in CNN-CV-RaT-X, a data enhancement method is employed to improve the performance of RaTDet. The samples for training are altered by adding a random Doppler steering vector to the target in the sample so as to decrease the sensitivity of RaTDet to the velocity of a target, where the velocity ranges in $[-v_q, v_q)$ with vq being the Nyquist velocity.

5.4. Ablation Study and Experiment Results

The pulse-range graph of the test data file "2021010150614_03_staring.mat" is shown in Figure 7. A target lies in range bins 1761 and 1762.

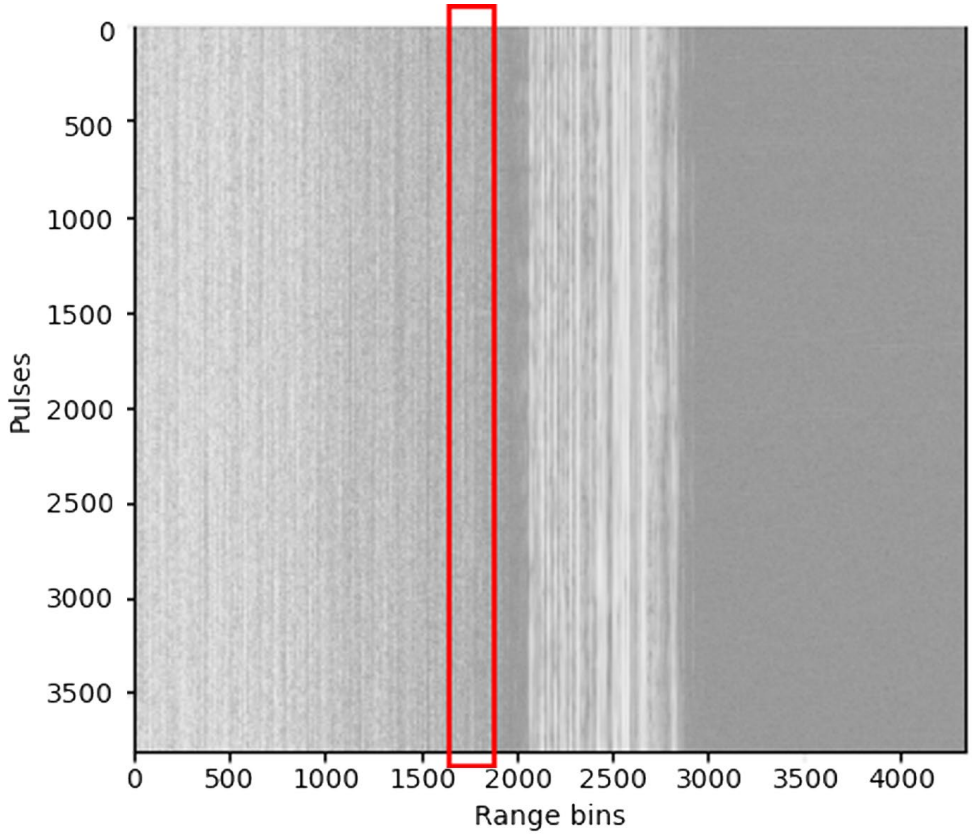


Figure 7. Visualization of test data file. The target lies in the red box.

When the false alarm rate is set around 2×10^{-3}, the detection results of the proposed models and other comparison models are demonstrated in Figure 8 and Table 3.

Table 3. Ablation study of different implementations under the test scenario.

Model	Pd (P_{fa}~10^{-3})	Accuracy	Specificity	Recall	Precision	F1 Score
CNN-Real-RaT	cannot converge	cannot converge	cannot converge	cannot converge	cannot converge	cannot converge
CNN-CV-RaT	70.5%	0.9979	0.9980	0.7050	0.0750	0.1356
CNN-CV-RaT-X	92%	0.9985	0.9985	0.9200	0.1237	0.2181
CNN-CV-RaT-X-mini	90.1%	0.9983	0.9983	0.9010	0.1087	0.1940

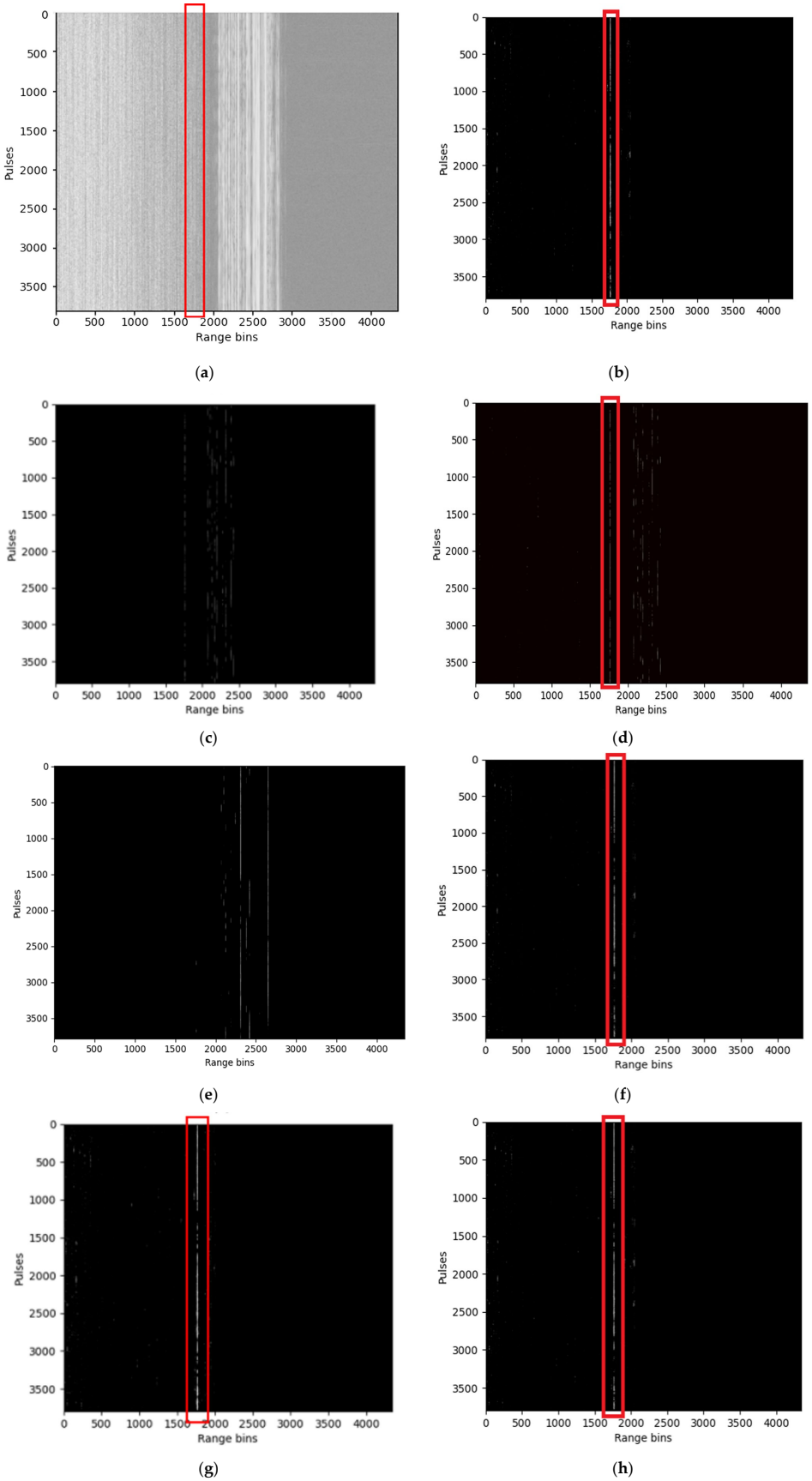


Figure 8. Detection results of (**a**) Original data; (**b**) CNN-CV-RaT-X; (**c**) traditional CA-CFAR; (**d**) MTD; (**e**) α-AMF and (**f**) MFF-Net; (**g**) U-Net; (**h**) Real-CNN-CV-RaT. The target lies in the red box.

During the training process, it was observed that the CNN-Real-RaT model failed to converge effectively. Consequently, as shown in Table 3, the CNN-Real-RaT model produced detection failures during testing, outputting all-zero results. This ablation study clearly demonstrates that connecting a real-valued network structure after complex-valued features disrupts the complex nature of the data, leading to information loss and resulting in ineffective model training. While the proposed CNN-CV-RaT-X converges successfully, with its loss decreasing smoothly from 0.9 to 0.09, this finding serves as a demonstration of the necessity of constructing an end-to-end complex-valued network.

In addition, the CNN-CV-RaT-X-mini reduced the number of complex-valued Transformer blocks from 4 to 2 and is trained on the augmented dataset. The result shows this operation only slightly decreases the detection performance, indicating the potential for further model simplification.

It should be noted that high accuracy and low precision is normal in a radar target detection scenario. The vast majority of range bins contain only clutter or noise (negative class), while targets (positive class) are exceedingly sparse. In such a setting, a classifier can achieve very high overall accuracy simply by correctly identifying the overwhelming number of negative samples.

Under the constraint of the specified false alarm probability, the traditional CA-CFAR method with 32 reference range cells and four protection cell ranges, although capable of performing detection, achieved a detection probability of only 0.04%, indicating significant performance limitations owing to model mismatching in a sea clutter background. In contrast, the RaTDet network proposed in this paper demonstrated outstanding performance: the standard model CNN-CV-RaT achieved a detection probability of 70.5%, while the data-augmented version CNN-CV-RaT-X further improved the detection probability to 83.1%. The classical coherent detection method MTD using 16 pulses and cascaded by a CA-CFAR detector with 32 reference range cells and four protection cell ranges achieved a detection probability of 7.6%. However, due to the limited number of coherent pulses and an inaccurate covariance matrix estimation, the α-AMF method [45], specifically designed for target detection in sea clutter environments, failed. Trained on the augmented dataset, the MFF-Net proposed in [46] and a plain U-Net proposed in [47] also exhibited a fair detection result. To further demonstration the necessity of a complex-valued model, a Real-CNN-CV-RaT with real-valued CNN and real-valued Transformer where the I/Q channels of input data are processed separately is also implemented for comparison. The specific results are summarized in Figure 8 and Table 4 below.

Table 4. Detection results of different algorithms under the test scenario.

Model	Pd (P_{fa}~10^{-3})	Accuracy	Specificity	Recall	Precision	F1 Score
CNN-CV-RaT-X	92%	0.9985	0.9985	0.9200	0.1237	0.2181
CA-CFAR	0.04%	0.9979	0.9981	0.0004	0.0000	0.0001
MTD	7.6%	0.9978	0.9980	0.0760	0.0087	0.0156
α-AMF	0%	0.0000	0.0000	0.0000	0.0000	0.0000
MFF-Net	87%	0.9941	0.9941	0.8700	0.0328	0.0633
U-Net	68.7%	0.9984	0.9985	0.6870	0.0954	0.1675
Real-CNN-CV-RaT	74.2%	0.9982	0.9983	0.7420	0.0913	0.1626

The Receiver Operating Characteristic (ROC) curves are plotted in Figure 9 for signal–clutter ratios (SCR) of 4 dB, 6 dB, 8 dB, and 10 dB. The results demonstrate that in all cases, our proposed RaTDet achieves the best detection performance, which is consistent with the observations from the test scenario discussed earlier.

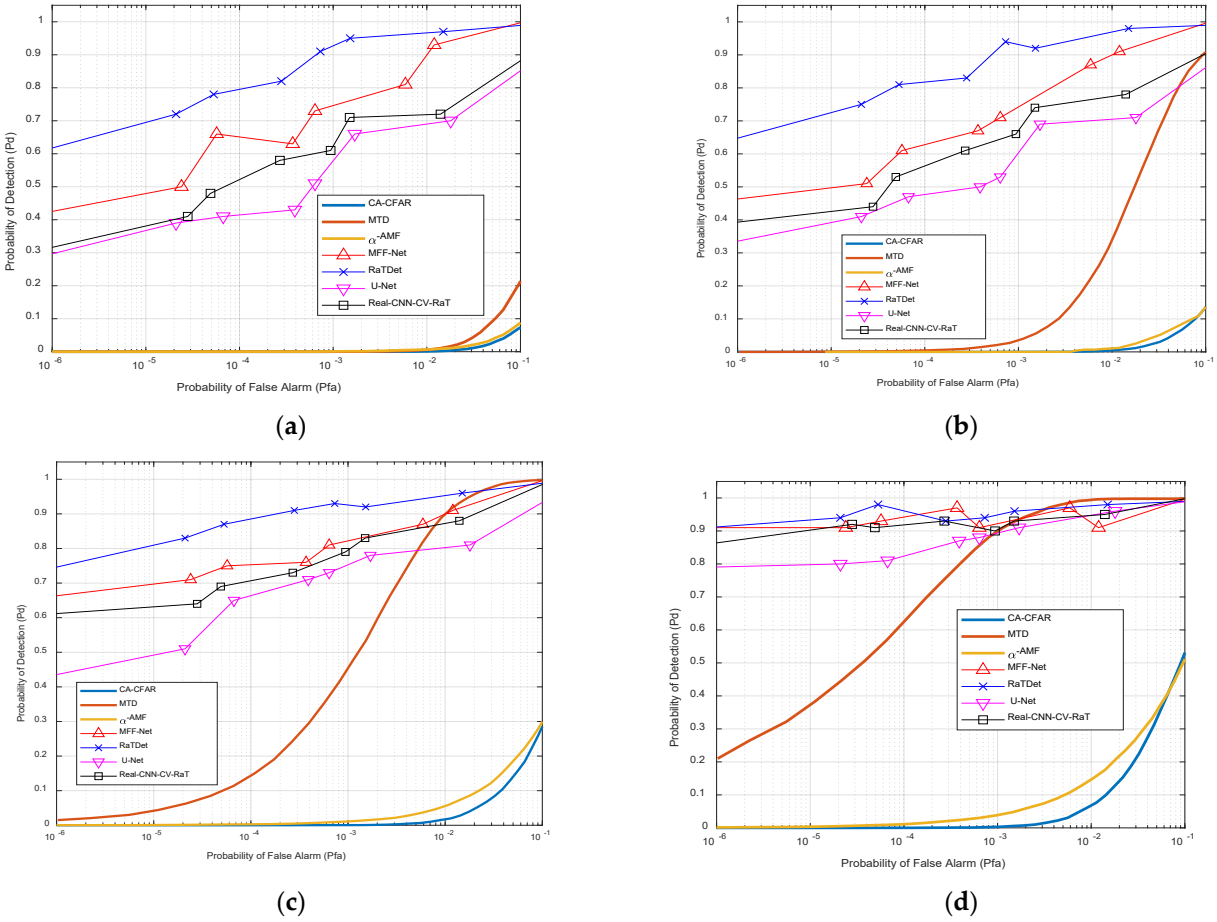


Figure 9. ROC curves under with (**a**) SCR = 4 dB; (**b**) SCR = 6 dB; (**c**) SCR = 8 dB; (**d**) SCR = 10 dB.

To further evaluate the performance of the detectors, we plot the Pd-SCR curves for false alarm rates of Pfa = 2×10^{-3} and Pfa = 2×10^{-4} in Figure 10. It is shown that RaTDet maintains robust detection performance across different SCR values in both cases.

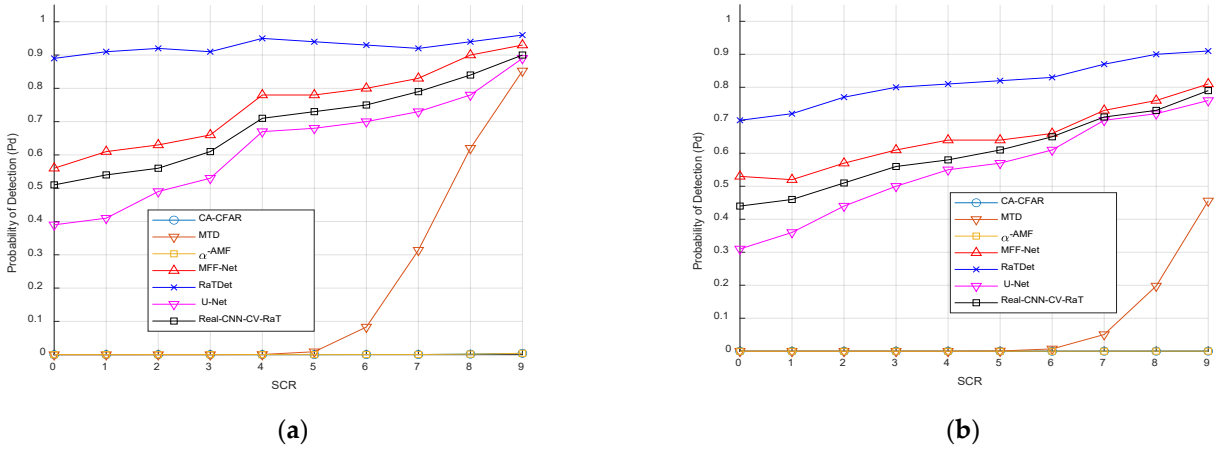


Figure 10. Pd-SCR curves under with (**a**) Pfa = 2×10^{-3}; (**b**) Pfa = 2×10^{-4}.

The experimental results have conclusively validated the efficacy of the proposed RaT-Det framework. The suboptimal performance of CNN-Real-RaT underscores the imperative for maintaining pure complex-valued topological configurations in radar signal processing architectures. Notably, the marked advancements in the probability of detection (Pd) metrics achieved by CNN-CV-RaT and CNN-CV-RaT-X relative to conventional CA-CFAR

implementations demonstrate their enhanced detection capabilities in intricate maritime environments. Furthermore, the state-of-the-art performance exhibited by CNN-CV-RaT-X substantiates the theoretical premise that the integration of Doppler-domain data augmentation techniques has demonstrably improved detection performance characteristics. This empirical evidence confirms the synergistic benefits derived from combining learnable feature representations with domain-specific signal processing priors in electromagnetic sensing systems.

6. Conclusions

The Radar Transformer for Detection (RaTDet) proposed in this work takes complex-valued pulse group data as input, offering two principal advantages: It fully preserves and processes the original information contained in radar I/Q data, while transforming the conventional fixed signal processing chain into a learnable, automated feature mining procedure within the neural network. By capturing implicit relationships among different range cells to localize targets, the model effectively integrates the powerful representational capacity of deep learning with the inherent inductive biases of radar signal processing and detection, while maintaining strong compatibility with traditional workflows.

The pulse group-based processing structure further ensures that the algorithm can be readily applied across diverse radar systems. This network also holds potential as a pre-training architecture for electromagnetic domain-specific large models. Future work will explore self-supervised pre-training strategies using more extensive and varied radar datasets to develop a large-scale radar electromagnetic model capable of understanding raw I/Q signals and adapting to multiple downstream tasks such as detection, recognition, and classification. Additionally, lightweight optimization of the model will constitute a key direction for subsequent research.

Author Contributions: Conceptualization, H.K. and H.Y.; methodology, H.K. and H.Y.; software, H.Y.; validation, H.Y.; formal analysis, H.Y.; investigation, H.K.; resources, H.K.; data curation, H.Y.; writing—original draft preparation, H.K.; writing—review and editing, L.Y.; visualization, H.Y. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The original data presented in the study are openly available in Dataset of Radar Detecting Sea at https://radars.ac.cn/web/data/getData?dataType=DatasetofRadarDetectingSea (accessed on 1 May 2025).

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

References

1. Wang, X.; Wang, Y.; Guo, S.; Zhao, W.; Zang, C.; Chen, X.; Cui, G. Intelligent Maritime Radar Target Detection In Unknown Environment Via Self-Learning. *IEEE Trans. Aerosp. Electron. Syst.* **2025**, *61*, 11172–11186. [CrossRef]
2. Li, Y.; Chen, L.; Li, W. Fine-Grained Ship Recognition with Spatial-Aligned Feature Pyramid Network And Adaptive Prototypical Contrastive Learning. *IEEE Trans. Geosci. Remote Sens.* **2025**, *63*, 5604313. [CrossRef]
3. Tan, J.; Sheng, W.; Zhu, H. Multiframe Disentangling and Unifying Graph Neural Network For Maritime Radar Target Detection. *IEEE Trans. Aerosp. Electron. Syst.* **2025**, *60*, 19239–19255. [CrossRef]
4. Chen, X.; Su, N.; Huang, Y.; Guan, J. False-Alarm-Controllable Radar Detection For Marine Target Based On Multi-Feature Fusion Via CNNs. *IEEE Sens. J.* **2021**, *21*, 9099–9111. [CrossRef]
5. Zhai, G.; Zhou, J.; Yang, H.; Zhang, Y. A Sea-Surface Radar Target-Detection Method Based On An Improved U-Net And Its FPGA Implementation. *Electronics* **2025**, *14*, 1944. [CrossRef]
6. Hu, J.; Liu, Z.; Ran, L.; Xie, R.; Xu, L. PCENet: A Method for SAR Ship Detection Based on Pixel Confidence Estimation. *IEEE Trans. Aerosp. Electron. Syst.* **2025**, *61*, 15540–15554. [CrossRef]
7. Li, J.; Huo, H.; Guo, L.; Zhang, D.; Feng, W.; Lian, Y.; He, L. Small Ship Detection Based on Improved Neural Network Algorithm and SAR Images. *Remote Sens.* **2025**, *17*, 2586. [CrossRef]

8. Wang, J.; Li, S. Maritime Radar Target Detection in Sea Clutter Based on CNN With Dual-Perspective Attention. *IEEE Geosci. Remote Sens. Lett.* **2023**, *20*, 3500405. [CrossRef]
9. Wen, L.; Ding, J.; Xu, Z. Multiframe Detection of Sea-Surface Small Target Using Deep Convolutional Neural Network. *IEEE Trans. Geosci. Remote Sens.* **2021**, *60*, 5107116. [CrossRef]
10. Jiang, W.; Wang, Y.; Li, Y.; Lin, Y.; Shen, W. Radar Target Characterization and Deep Learning in Radar Automatic Target Recognition: A Review. *Remote Sens.* **2023**, *15*, 3742. [CrossRef]
11. Yavuz, F.; Kalfa, M. Radar Target Detection via Deep Learning. In *Proceedings of the 2020 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey*; IEEE: New York, NY, USA, 2020. [CrossRef]
12. Jiang, W.; Ren, Y.; Liu, Y.; Leng, J. Artificial Neural Networks and Deep Learning Techniques Applied to Radar Target Detection: A Review. *Electronics* **2022**, *11*, 156. [CrossRef]
13. Wang, L.; Tang, J.; Liao, Q. A Study on Radar Target Detection Based on Deep Neural Networks. *IEEE Sens. Lett.* **2019**, *3*, 7000504. [CrossRef]
14. Salehi, A.; Imani, M.; Zaimbashi, A.; Yanikomeroglu, H. Learning and Model-Based Approaches for Radar Target Detection. *IEEE Trans. Cogn. Commun. Netw.* **2024**, *10*, 1817–1830. [CrossRef]
15. Pan, M.; Chen, J.; Wang, S.; Dong, Z. A Novel Approach for Marine Small Target Detection Based on Deep Learning. In Proceedings of the 2019 IEEE 4th International Conference on Signal and Image Processing (ICSIP), Wuxi, China, 19 July 2019. [CrossRef]
16. Xu, S.; Bai, X.; Guo, Z.; Shui, P. Status and prospects of feature-based detection methods for floating targets on the sea surface. *J. Radars* **2020**, *9*, 684–714. [CrossRef]
17. Wang, Y.; Zhang, Y.; Qu, H.; Tian, Q. Target Detection and Recognition Based on Convolutional Neural Network for SAR Image. In Proceedings of the 2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Beijing, China, 13 October 2018. [CrossRef]
18. Dong, Y. Optimal coherent radar detection in a K-distributed clutter environment. *IET Radar Sonar Navig.* **2012**, *6*, 283–292. [CrossRef]
19. Sangston, K.J.; Gini, F.; Greco, M.S. Coherent Radar Target Detection in Heavy-Tailed Compound-Gaussian Clutter. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 64–77. [CrossRef]
20. Shang, X.; Song, H. Radar detection based on Compound-Gaussian model with inverse gamma texture. *IET Radar Sonar Navig.* **2011**, *5*, 315–321. [CrossRef]
21. Shui, P.; Liu, M.; Xu, S. Shape-parameter-dependent coherent radar target detection in K-distributed clutter. *IEEE Trans. Aerosp. Electron. Syst.* **2016**, *52*, 451–465. [CrossRef]
22. Conte, E.; Lops, M.; Ricci, G. Asymptotically optimum radar detection in compound-Gaussian clutter. *IEEE Trans. Aerosp. Electron. Syst.* **1995**, *31*, 617–625. [CrossRef]
23. Zhou, W.; Xie, J.; Li, G.; Du, Y. Robust CFAR Detector With Weighted Amplitude Iteration in Nonhomogeneous Sea Clutter. *IEEE Trans. Aerosp. Electron. Syst.* **2017**, *53*, 1520–1535. [CrossRef]
24. Kelly, E.J. An Adaptive Detection Algorithm. *IEEE Trans. Aerosp. Electron. Syst.* **1986**, *AES-22*, 115–127. [CrossRef]
25. Robey, F.C.; Fuhrmann, D.R.; Kelly, E.J.; Nitzberg, R. A CFAR adaptive matched filter detector. *IEEE Trans. Aerosp. Electron. Syst.* **1992**, *28*, 208–216. [CrossRef]
26. Panagopoulos, S.; Soraghan, J.J. Small-target detection in sea clutter. *IEEE Trans. Geosci. Remote Sens.* **2004**, *42*, 1355–1361. [CrossRef]
27. Gao, S.; Liu, H. Performance Comparison of Statistical Models for Characterizing Sea Clutter and Ship CFAR Detection in SAR Images. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2022**, *15*, 7414–7430. [CrossRef]
28. Wen, B.; Lu, Z.; Zhou, B. Marine Radar Constant False Alarm Rate Detection in Generalized Extreme Value Distribution Based on Space-Time Adaptive Filtering Clutter Statistical Analysis. *Remote Sens.* **2024**, *16*, 3691. [CrossRef]
29. Ke, H.; Ke, X.; Zhang, Z.; Chen, X.; Xu, X.; Zhang, T. SLA-Net: A Novel Sea–Land Aware Network for Accurate SAR Ship Detection Guided by Hierarchical Attention Mechanism. *Remote Sens.* **2025**, *17*, 3576. [CrossRef]
30. Ke, H.; Ke, X.; Yan, Y.; Luo, D.; Cui, F.; Peng, H.; Hu, Y.; Liu, Y.; Zhang, T. Laplace & LBP feature guided SAR ship detection method with adaptive feature enhancement block. In Proceedings of the 2024 IEEE 6th Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), Chongqing, China, 2 July 2024. [CrossRef]
31. Ke, X.; Zhang, X.; Zhang, T. GCBANet: A Global Context Boundary-Aware Network for SAR Ship Instance Segmentation. *Remote Sens.* **2022**, *14*, 2165. [CrossRef]
32. Zhou, Z.; Li, S.; Lv, Y.; Guo, S.; Wang, H.; Yang, J.; You, J.; Zhang, K. LightKD-SAR: Lightweight Architecture With Knowledge Distillation for High-Performance SAR Object Detection. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2026**, *19*, 4935–4949. [CrossRef]
33. Chen, B.; Xue, F.; Song, H. A Lightweight Arbitrarily Oriented Detector Based on Transformers and Deformable Features for Ship Detection in SAR Images. *Remote Sens.* **2024**, *16*, 237. [CrossRef]

34. Yin, Y.; Yang, Z.; Shi, H.; Meng, F.; Li, W. Ship Detection Transformer in SAR Images Based on Key Scattering Points Feature Aggregation and Context Feature Refinement. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2025**, *18*, 17820–17836. [CrossRef]
35. Zhou, H.; Jiang, T. Decision Tree Based Sea-Surface Weak Target Detection With False Alarm Rate Controllable. *IEEE Signal Process. Lett.* **2019**, *26*, 793–797. [CrossRef]
36. Shi, Y.; Xie, X.; Li, D. Range Distributed Floating Target Detection in Sea Clutter via Feature-Based Detector. *IEEE Geosci. Remote Sens. Lett.* **2016**, *13*, 1847–1850. [CrossRef]
37. Xu, S.; Zheng, J.; Pu, J.; Shui, P. Sea-Surface Floating Small Target Detection Based on Polarization Features. *IEEE Geosci. Remote Sens. Lett.* **2018**, *15*, 1505–1509. [CrossRef]
38. Li, Y.; Xie, P.; Tang, Z.; Jiang, T.; Qi, P. SVM-Based Sea-Surface Small Target Detection: A False-Alarm-Rate-Controllable Approach. *IEEE Geosci. Remote Sens. Lett.* **2019**, *16*, 1225–1229. [CrossRef]
39. Haykin, S.; Li, X.B. Detection of signals in chaos. *Proc. IEEE* **1995**, *83*, 95–122. [CrossRef]
40. Hu, J.; Tung, W.; Gao, J. Detection of low observable targets within sea clutter by structure function based multifractal analysis. *IEEE Trans. Antennas Propag.* **2006**, *54*, 136–143. [CrossRef]
41. Shui, P.; Li, D.; Xu, S. Tri-feature-based detection of floating small targets in sea clutter. *IEEE Trans. Aerosp. Electron. Syst.* **2014**, *50*, 1416–1430. [CrossRef]
42. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *arXiv* **2017**, arXiv:1706.03762. Available online: https://arxiv.org/abs/1706.03762 (accessed on 12 June 2017).
43. Eilers, F.; Jiang, X. Building Blocks for a Complex-Valued Transformer Architecture. In Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes Island, Greece, 4 June 2023. [CrossRef]
44. Liu, N.; Li, J.; Wang, G.; Chen, B.; Cao, Z.; Dong, Y.; Guan, J.; Jiang, X.; Zhang, Z.; Xue, W. Sea-detecting radar experiment and target feature data acquisition for dual polarization multistate scattering dataset of marine targets. *J. Radars* **2023**, *12*, 456–469. [CrossRef]
45. Wang, Z.; Jian, T.; He, Y. Adaptive detector based on optimal control parameter in compound-Gaussian background. *Control. Decis.* **2018**, *33*, 1532–1536. [CrossRef]
46. Wang, X.; Wang, Y.; Chen, X.; Zang, C.; Cui, G. Deep learning-based marine target detection method with multiple feature fusion. *J. Radars* **2024**, *13*, 554–564. [CrossRef]
47. Ronneberger, O.; Fischer, P.; Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. In Proceedings of the Medical Image Computing and Computer-Assisted Intervention (MICCAI 2015), Munich, Germany, 5–9 October 2015; pp. 234–241. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.