

A Multiscale Convolution SAR Image Target Recognition Method Based on Complex-Valued Neural Networks

Guangyu Hou[✉], Zhihui Xin[✉], Guisheng Liao[✉], Senior Member, IEEE, Penghui Huang[✉], Member, IEEE, Yuhao Huang, and Rui Zou[✉]

Abstract—Recent advances in deep learning have driven significant success in synthetic aperture radar (SAR) automatic target recognition, particularly through convolutional neural network (CNN) based classification algorithms. However, SAR images possess distinctive physical scattering properties, owing to their unique imaging mechanism. Many deep learning algorithms rely solely on amplitude information, ignoring phase information, which may result in the loss of information in the original complex-valued SAR image and suboptimal performance. To tackle these problems, this article introduces a SAR target recognition approach based on complex-valued operations, designated as complex-valued residual mish activation and convolution block attention module (CBAM) net (CRMC-Net). The CRMC-Net effectively utilizes the amplitude and phase information in complex-valued SAR data. Specifically, first, the elements of CNN, including the input and output layers, the convolution layers, the activation functions, and the pooling layers, are extended to the complex-valued domain. Second, in order to further enhance the representation ability of the model, multiscale information of the target is extracted through different convolution kernel sizes. Finally, the network constructs many complex-valued operation blocks to enhance the robustness of the designed network, including the complex-valued residual block, complex-valued Mish activation function, and complex-valued CBAM. The experimental results obtained from the moving and stationary target capture and recognition dataset and OpenSARShip2.0 dataset demonstrate that the proposed network model outperforms the traditional real-valued models. It can further reduce the classification error and enhance performance.

Index Terms—Complex-valued convolution neural network (CVCNN), phase, synthetic aperture radar (SAR), target recognition.

Received 5 December 2024; revised 13 March 2025; accepted 7 April 2025. Date of publication 10 April 2025; date of current version 29 April 2025. This work was supported in part by the National Natural Science Foundation of China under Grant 62361058 and Grant 62171272, in part by Yunnan Provincial Natural Science Foundation under Grant 202201AT070027, and in part by Yunnan Expert Workstation under Grant 202305AF150012. (Corresponding author: Zhihui Xin.)

Guangyu Hou, Zhihui Xin, Yuhao Huang, and Rui Zou are with the Yunnan Key Laboratory of Opto-Electronic Information Technology, Yunnan Normal University, Kunming 650500, China (e-mail: hgy18730016665@163.com; xinzhihui.luncky@163.com; h18381078711@163.com; 18189598995@163.com).

Guisheng Liao is with the National Laboratory of Radar Signal Processing, Xidian University, Xi'an 710071, China (e-mail: liaogs@xidian.edu.cn).

Penghui Huang is with the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai 200240, China (e-mail: huangpenghui@sjtu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2025.3559656

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) is an active microwave imaging sensor that provides all-weather, day-and-night remote sensing data collection capabilities. It is extensively applied to both military and civilian settings, owing to its ability to capture considerable surface information of the Earth under various conditions [1], [2]. Automatic target recognition (ATR) is an important aspect of SAR image processing. Several algorithms for SAR target recognition have been developed in recent studies and can be broadly categorized into two methods: traditional and deep learning (DL) methods. Traditional methods generally include template matching-based, model-based, and machine-learning-based methods.

Template matching is a basic pattern recognition method that generates multiple templates and compares them with the target image. The target category is determined by calculating the similarity (such as correlation coefficient and mean square error) and matching the region of interest for classification [3], [4]. However, this method requires traversing all possible positions, resulting in high computational cost and slow processing speed. In addition, it relies on numerous samples to generate templates of varying angles and scales, which are susceptible to target deformation, rotation, and scale changes. Consequently, implementing this method in actual SAR recognition tasks is challenging. Model-based methods perform recognition by constructing physical and conceptual models [5]. They have high robustness and adaptability, enabling them to accommodate changes in the target pose and configuration while extracting intricate feature information. Nonetheless, the processes of model construction and feature extraction are relatively complicated, limiting their application in SAR recognition. Machine-learning method achieves automatic recognition and classification by extracting features from SAR images and using algorithms for training and classification [6], [7]. Their key task involves designing suitable feature extractors. However, this approach requires considerable labeled data, computational resources, and expertise. Traditional SAR recognition methods mainly rely on the distinguishability of artificially designed features. However, achieving optimal recognition results when processing complex SAR images is challenging. In addition, these methods often lose key information during the preprocessing and feature extraction stages, ultimately resulting in reduced performance.

In recent years, the rapid development of DL theory has resulted in the emergence of convolutional neural networks (CNNs), which have demonstrated significant advantages in image classification [8], [9], [10]. The DL method performs well in computer vision tasks; hence, it has become the prevailing method for SAR image classification and is instrumental in advancing intelligent information extraction [11], [12]. Traditional image classification algorithms require hand-crafted features and intricate physical models. However, DL methods can automatically extract target features without manual intervention, simplifying feature extraction and outperforming traditional methods [13], [14]. The concept of DL was proposed in [15], which enhanced the importance of CNN in automatic feature extraction. Many studies have shown that methods based on deep neural networks achieve excellent performances in various SAR target recognition tasks. For example, a novel CNN model based on multikernel-size feature fusion for SAR target classification was proposed in [16]. The classification accuracy of the ten-class target MSTAR datasets reached 97.44%. Chen et al. [17] proposed a fully convolutional network to improve the accuracy of recognition tasks. In [18], a novel global and local descriptor-based metric network for few-shot SAR target classification was proposed. A method based on multilayer self-encoders and hyperpixels was proposed in [19], which significantly improved the image classification accuracy and discrimination between different image classes. In [20], the ResNet18 model attained 99% overall classification accuracy on the MSTAR dataset. A multitarget domain adaptation approach named progressive prototype refinement was proposed in [21]. To further demonstrate the benefits of DL for SAR target recognition, Wang et al. [22] proposed an improved GoogLeNet structure network. LcFGC, which consists of a multiscale local classification network and a feature generation and calibration network, was proposed in [23] to improve the performance of few-shot SAR target classification.

However, although DL-based methods have performed well in SAR image classification tasks, they primarily rely on real-valued networks that use only amplitude information while neglecting phase information. This limits their ability to achieve optimal recognition results. These methods are mainly based on real-valued networks. When considering solely the amplitude information of the data as input, it can be observed that the network structures are all based on real-valued representations, ignoring the use of phase information. It is difficult to obtain satisfactory recognition results relying solely on amplitude information [24], [25], [26]. Phase information is a distinct characteristic of SAR images and a key component in many SAR applications [27], [28]. It provides information on the material and shape boundary of the object, whereas the amplitude information encodes energy and reflects the electromagnetic wave intensity of the object [29], [30]. A complex-valued convolution neural network (CVCNN) can simultaneously utilize amplitude and phase information to improve classification performance [31], [32], [33]. To effectively utilize phase information from SAR targets, an analysis of the mathematical motivation of the CVCNN was conducted in [34] and [35]. In recent years, many studies have exploited amplitude and phase information in SAR images, demonstrating

its superior performance over amplitude information. Phase information is particularly useful for identifying different types of scatterers in polarimetric SAR (PolSAR) and this has prompted the exploration of CVCNN for PolSAR image classification. In [36], a CVCNN was proposed for classifying the terrain of PolSAR images and achieved a low classification error rate. In [37], a multiscale depth contourlet filter bank was constructed, which demonstrated effective performance in the classification of PolSAR images. Cao et al. [38] designed a fully CNN and achieved satisfactory classification results in experiments based on PolSAR datasets. Meanwhile, numerous complex-valued classification networks have proven to be successful in SAR-ATR. In [39], a complex-valued convolutional neural network (CV-Net), for target recognition in single-channel SAR images, was proposed. A complex-valued CNN was proposed in [40], achieving a 98.97% recognition rate in MSTAR classification experiments, which represented a significant improvement over many real-valued CNN algorithms. A new complex-valued full CNN (CV-FCNN) was proposed in [41], where the pooling and fully connected layers were replaced by convolution layers. The results of the experiments indicated that the CV-FCNN outperformed the CVCNN in terms of recognition. A multistream complex-valued network was developed in [42] that achieved significant performance improvements on the MSTAR dataset. To develop the complex-valued information inherent in SAR images adequately, a target detection method for SAR images using the complex-valued network (CV-SAR-Det) was proposed in [43]. Wang et al. [44] proposed a subaperture decomposition-guided complex-valued network (CGS-net) to effectively exploit the phase information. The feasibility of phase information in SAR images was further verified by recent research methods in [45], [46], and [47].

The scattering of target pixels results from the existence of coherent speckle noise, thereby impairing the separability of targets in SAR images. Traditional SAR target recognition methods usually rely only on amplitude information and ignore phase information. Although amplitude information provides certain features of the image, its recognition ability is limited in the presence of target defocus and noise interference, making it difficult to achieve the accurate target recognition. The rich complex-valued information inherent in SAR images provides more physical scattering features. SAR phase information is related to the geometric, material, and dynamic properties of scattering targets. This offers great potential for improving target recognition accuracy. Therefore, by introducing complex-valued operations, the model can effectively capture and utilize the phase information in the image and enhance the expression of the target features, thereby improving the model's recognition ability when dealing with complex backgrounds and noise interference. However, the existing CVCNNs still has the problem of failing to fully extract target feature information and demonstrates suboptimal recognition accuracy. Therefore, this article proposes a SAR target recognition method based on complex-valued operations, called the complex-valued residual mish activation and convolution block attention module (CBAM) net (CRMC-Net). The CRMC-Net effectively utilizes both amplitude and phase information in complex-valued SAR

data and fully exploits the physical scattering characteristics of SAR images. It serves to enhance the model's capacity to represent the target features. The main contributions of this article are summarized as follows.

- 1) By introducing convolution kernels of different sizes to extract multiscale information in the scene, the representational ability of the target is further enhanced.
- 2) The combination of the complex-valued residual block (CReLUBlock) and complex-valued Mish (CMish) activation function lead to a notable enhancement in the model's feature extraction capability and the stability of the training process, further enhancing the efficiency of capturing complex features and information flow.
- 3) A complex-valued convolution block attention module (CCBAM) is constructed to assist the model that distinguish different targets and backgrounds better under the interference of coherent speckle noise, thereby achieving a higher recognition rate in complex SAR images.

The rest of this article is organized as follows. Section II provides a brief introduction to the complex-valued arithmetic building blocks. A detailed theory of the CRMC-Net is proposed in Section III. Section IV presents the experimental design and comparative analysis of the results. Section V presents a discussion. Finally, Section VI concludes this article.

II. COMPLEX-VALUED BUILDING BLOCKS

This section provides a comprehensive account of the complex-valued arithmetic building blocks in the CRMC-Net. Following the works of the authors in [32] and [38], we focus on describing the design and implementation of these complex-valued building blocks and establish a mathematical framework for implementing complex-valued arithmetic in deep neural networks. This section establishes the fundamental basis of the core content of our study.

A. Complex-Valued Convolution Block

The complex-valued convolution block consists of a complex-valued convolution layer, a complex-valued batch normalization (CBN) layer, and a complex-valued activation function layer, as shown in Fig. 1.

1) *Complex-Valued Convolution Layer:* A complex-valued convolution is an extension of real-valued convolution. In contrast to real-valued convolution, which extracts features solely from amplitude data, complex-valued convolution employs both amplitude and phase information inherent in complex-valued SAR images to extract target features. To facilitate the implementation of convolution operations within a complex-valued domain, complex-valued operations are simulated in the convolution of a real-valued network. In traditional convolution operations, when real-valued convolution is employed for feature extraction, the elements of each small area on the feature map are multiplied individually by the convolution kernel, and the resulting products are subsequently summed to obtain a value within the feature map. The same principles apply to

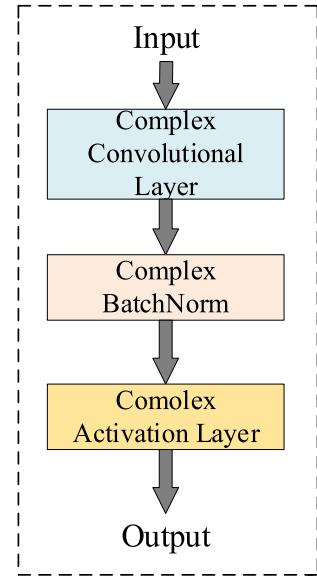


Fig. 1. Complex-valued convolution block.

complex-valued operations, which are expressed as follows:

$$\begin{aligned} W * H &= (A + iB) * (C + iD) \\ &= (A * C - B * D) + i(A * D + B * C). \end{aligned} \quad (1)$$

The matrix version of the convolution operation can be given by

$$\begin{bmatrix} \Re(W * H) \\ \Im(W * H) \end{bmatrix} = \begin{bmatrix} A & -B \\ B & A \end{bmatrix} * \begin{bmatrix} C \\ D \end{bmatrix} \quad (2)$$

where $W = A + iB$ represents the complex-valued convolution kernel, $H = C + iD$ represents the complex-valued feature layer, and A, B, C , and D represent the corresponding real and imaginary parts.

A new feature layer is generated after the complex-valued convolution operation. These complex-valued feature layers contained richer information. For a clearer description of the process, Fig. 2 illustrates the entire complex-valued convolution of the calculations [47].

2) *Complex-Valued BN:* BN is a widely adopted and efficient technique for enhancing model performance in deep neural networks. Typically, it regulates the intermediate outputs, accelerating the network learning process, improving the model convergence, and mitigating the likelihood of overfitting. However, traditional BN algorithms are mainly based on real-valued calculations, which are constrained in their capacity to process complex-valued data. In order to implement CBN, it is necessary to introduce a CBN operation as described in [32]. This allows the data results in the complex-valued neural network to be standardized in accordance with a standard normal complex distribution, with a mean of zero and a covariance of one. This normalization is expressed as

$$\tilde{x} = V^{-\frac{1}{2}}(x - E[x]) \quad (3)$$

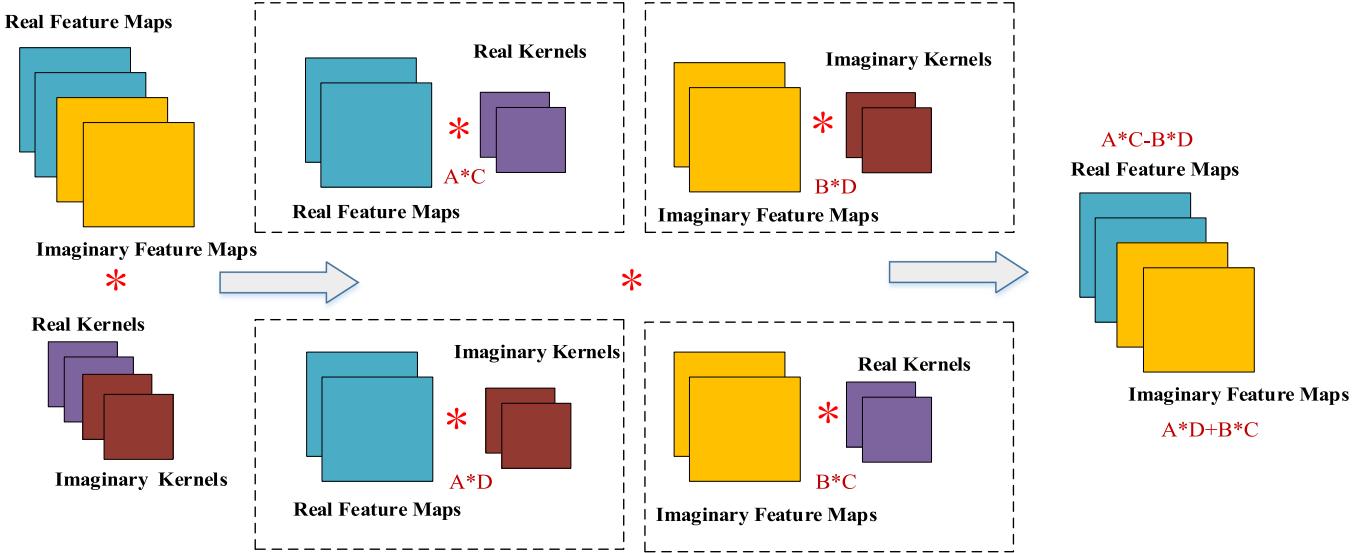


Fig. 2. Complex number convolution operation process.

$$\begin{aligned} V &= \begin{pmatrix} V_{rr} & V_{ri} \\ V_{ir} & V_{ii} \end{pmatrix} \\ &= \begin{pmatrix} \text{Conv}(\Re(x), \Re(x)) & \text{Conv}(\Re(x), \Im(x)) \\ \text{Conv}(\Im(x), \Re(x)) & \text{Conv}(\Im(x), \Im(x)) \end{pmatrix} \quad (4) \end{aligned}$$

where V is the covariance matrix, and $\text{Conv}(\cdot)$ and $E[\cdot]$ represent the covariance calculation and expectation operation, respectively.

Extend the real-valued BN to the complex-valued domain with two parameters: the scaling parameter γ and the translation parameter β . Then, the CBN can be defined as

$$\text{BN}(\tilde{x}) = \gamma\tilde{x} + \beta. \quad (5)$$

In particular, the translation parameter β is a complex number with learnable real and imaginary parts. The scaling parameter γ is a 2×2 symmetric positive semidefinite matrix with three learnable components γ_{rr} , γ_{ii} , and γ_{ri} , which are given as follows:

$$\gamma = \begin{pmatrix} \gamma_{rr} & \gamma_{ri} \\ \gamma_{ri} & \gamma_{ii} \end{pmatrix}. \quad (6)$$

Following the process of normalization, the variances of the real and imaginary parts of \tilde{x} are both 1. In order to ensure that the variance modulus of \tilde{x} is 1, γ_{rr} and γ_{ii} are initialized to $1/\sqrt{2}$, respectively. The real and imaginary parts of β and γ_{ri} are initialized to 0.

3) *Complex-Valued Activation Function Layer*: In traditional CNN, some common real-valued activation functions, such as ReLU and Sigmoid, are generally used. To address the issue of complex-valued representations, it is essential to consider the connection between the amplitude and phase information. In complex-valued CNN, the traditional ReLU activation function is extended to the complex-valued domain. Two distinct ReLU activation functions are then employed on the real and imaginary

parts

$$\text{CReLU}(z) = \text{ReLU}(\Re(z)) + i\text{ReLU}(\Im(z)). \quad (7)$$

B. Complex-Valued Pooling Layer

The complex pooling layer is used for dimensionality reduction and feature extraction in the CVCNN. This principle is similar to the pooling operation in real-valued neural networks, but it is necessary to process both real and imaginary parts of complex numbers. The application of complex-valued pooling to complex-valued neural networks enhances the networks performance. This is achieved by compressing the feature information and removing redundant parts while retaining key amplitude and phase information. This process guarantees the preservation of significant features and more effectively facilitates the capture and representation of complex-valued patterns in the data.

In complex-valued pooling, the commonly used operations are complex-valued maximum pooling (CMax-Pooling) and complex-valued average pooling (CAvg-Pooling). The CMax-Pooling selects the complex value with the largest magnitude in the feature patch, whereas the CAvg-Pooling calculates the average of the complex values in the feature patch. For a given input feature map, $Z = Z_r + iZ_i$, where Z_r and Z_i are the real and imaginary parts, respectively, the CMax-Pooling operation can be expressed as follows:

$$Z_{\max} = \max(|Z|) \quad (8)$$

where $|Z|$ represents the modulus of the complex number.

The formula for CAvg-Pooling is given as follows:

$$Z_{\text{avg}} = \frac{1}{n} \sum_{k=1}^n Z_k \quad (9)$$

where Z_k is the complex-valued feature map and n is the number of elements in the feature map.

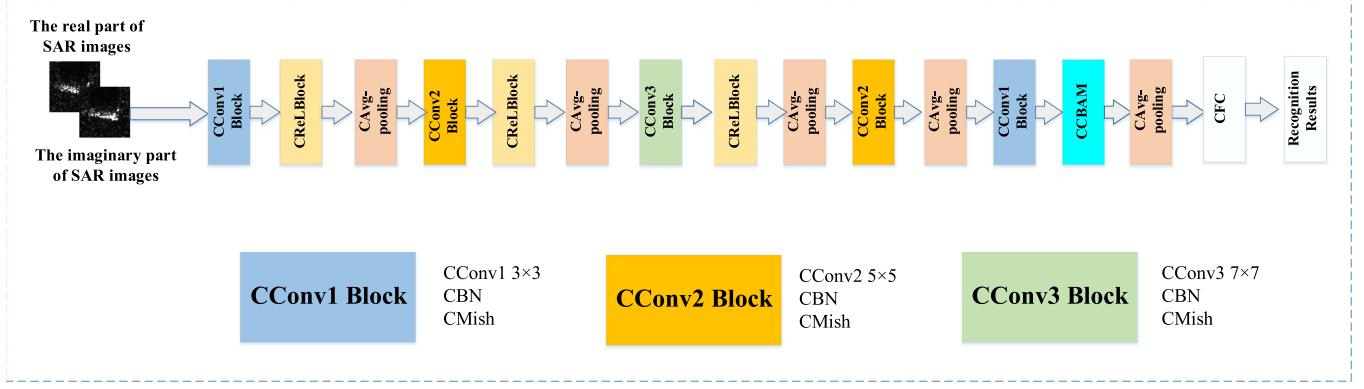


Fig. 3. Architecture of the proposed CRMC-Net.

C. Complex-Valued Fully Connected Layer

Similar to a complex-valued convolution operation, a complex-valued fully connected layer represents an extension of a real-valued fully connected layer in the complex field. In recognition tasks, it can process and retain the complex-valued information in these features by integrating complex-valued input features, thereby facilitating more efficient recognition output. Assuming that a complex weight matrix $W = W_r + iW_i$ and the input vector is $X = X_r + iX_i$, the formula of the complex-valued fully connected layer is as follows:

$$y = (W_r X_r - W_r X_i) + i(W_r X_i + W_i X_r) \quad (10)$$

where W_r and W_i represent the real and imaginary parts of the weight, and X_r and X_i represent the real and imaginary parts of the vector, respectively.

III. PROPOSED METHOD

In this section, we systematically explain the construction method of the CRMC-Net. Fig. 3 illustrates the proposed CRMC-Net architecture. The network consists of five complex-valued convolution blocks, three complex-valued residual modules, a CCBAM, five CAvg-Pooling layers, and a complex-valued fully connected layer. The proposed network employs convolution kernels of different sizes (3×3 , 5×5 , and 7×7) to extract multiscale information, thereby generating richer feature expressions. To enhance the network's capability to learn features effectively, we introduced CReLUBlock and the CMish activation function. CReLUBlock enhances the representation capability of the model through residual connections. The CMish activation function promotes information flow through a smooth nonlinear transformation and improves the stability of feature extraction. Their combination demonstrated an enhanced capacity for feature expression and classification performance. In addition, the CCBAM is utilized after the last convolution module of the network to further optimize the extraction of key features, which enhances the classification robustness of the model in complex scenarios. Table I provides a comprehensive account of the parameter configuration of the network, including key information, such as the convolution kernel size, stride, and feature map output size of each module. The CRMC-Net is

TABLE I
PARAMETER CONFIGURATION OF THE PROPOSED CRMC-NET

Image Name	Kernel size	Stride	Output size
Image			64×64
CConv1 Block	3×3	1	64×64
CRELBlock	3×3	1	64×64
CAvg-pooling	2×2	2	32×32
CConv2 Block	5×5	1	32×32
CRELBlock	3×3	1	32×32
CAvg-pooling	2×2	2	16×16
CConv3 Block	7×7	1	16×16
CRELBlock	3×3	1	16×16
CAvg-pooling	2×2	2	8×8
CConv2 Block	5×5	1	4×4
CAvg-pooling	2×2	2	2×2
CConv1 Block	3×3	1	2×2
CCBAM	3×3	1	2×2
CAvg-pooling	2×2	2	1×1
CFC			

a CVCNN designed specifically for processing complex data in SAR images. All layers of the network are complex-valued layers that can effectively extract the target features contained in the amplitude and phase information in the SAR image through the complex-valued convolution layers. Compared with real-valued network methods, complex-valued computation more accurately aligns with the physical scattering of targets, enabling the extraction of richer target features and significantly enhancing recognition performance. The substructures are detailed as follows.

A. Complex-Valued Residual Block

To effectively leverage the benefits of the residual structure in traditional real-valued CNNs, we extend it to a complex domain and introduce a complex-valued residual module into CRMC-Net. As illustrated in Fig. 4, the configuration of the CRELBlock is analogous to that of the conventional residual module. Both the convolution and activation layers are structured as complex valued. The backbone network consists of two 3×3 complex-valued convolution layers, which are used to extract features. The branch network modifies the input data

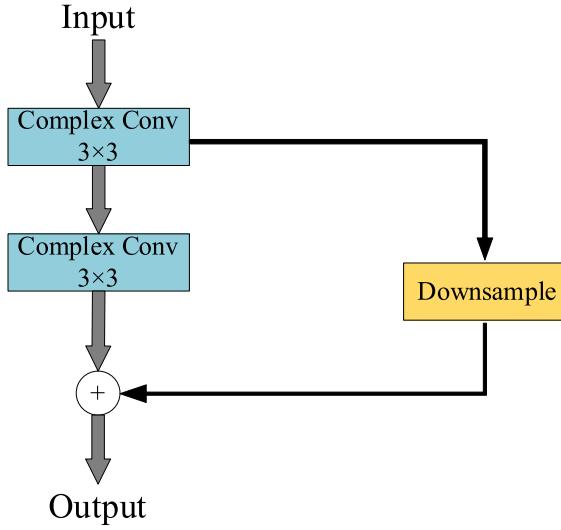


Fig. 4. Schematic diagram of the CReLBlock structure.

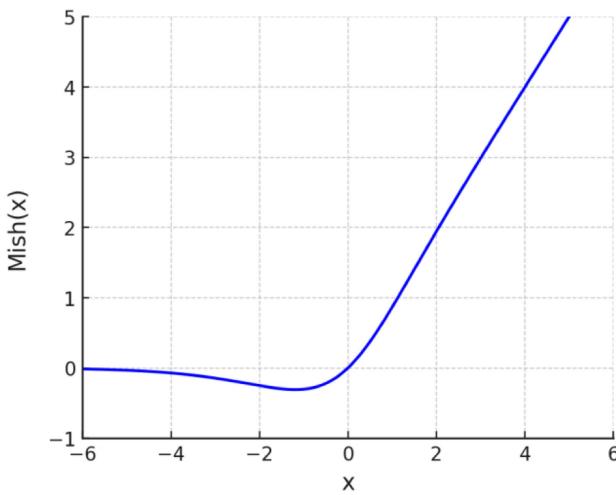


Fig. 5. Mish activation function image.

through the downsampling layer to ensure that it aligns with the output of the complex-valued convolution layer considering the channels' quantity and size. This process guarantees the normal functioning of CReLBlock.

The incorporation of skip connections serves to mitigate the issues of gradient vanishing and degradation that are inherent in deep networks. The network is capable of effectively integrating complex information, accurately extracting and processing the amplitude and phase features in complex data, ensuring the integrity of the features, and significantly improving the feature extraction effect. In general, the introduction of CReLBlock significantly enhances the robustness and recognition performance of the CRMC-Net, making it perform better in processing complex-valued SAR image tasks.

B. CMish Activation Function

An image of the Mish activation function is shown in Fig. 5. Unlike the ReLU function, which is not differentiable at the origin, the Mish function is smooth throughout its domain, facilitating network convergence. The Mish activation function helps maintain a stable gradient flow during training, owing to its smooth, nonlinear transformation and continuous differentiability. This mitigates gradient disappearance and dying ReLU issues that arise in the ReLU function. The CMish activation function is used to enhance the expressive power and training effect of complex-valued neural networks. The CMish function not only incorporates the benefits of the aforementioned Mish function by performing nonlinear transformations on the real and imaginary parts, respectively, but also enhances the fluidity of information flow and gradient transfer within the complex domain. The proposed design effectively enhances the training performance of complex-valued neural networks, facilitating a more precise extraction and retention of complex features. Among them, for $x \in R$, Mish function [48] can be expressed as

$$\text{Mish}(x) = x \cdot \tanh(\ln(1 + e^x)). \quad (11)$$

As with the CReLU activation function defined, the proposed CMish function applies the Mish function to the real and imaginary parts separately. Given that $z = x + iy$, the CMish function is defined as follows:

$$\begin{aligned} \text{CMish}(z) &= \text{Mish}(\Re(z)) + i \cdot \text{Mish}(\Im(z)) \\ &= x \tanh(\ln(1 + e^x)) + i \cdot y \tanh(\ln(1 + e^y)). \end{aligned} \quad (12)$$

C. Complex-Valued Convolution Block Attention Module

In the CRMC-Net, the CCBAM is introduced to further improve the capability of the model to represent and extract complex-valued features from SAR images. The CCBAM comprises two distinct components: complex-valued channel attention and complex-valued spatial attention, which perform weighted processing on complex features in the channel and spatial dimensions, respectively. As illustrated in Fig. 6, the complex-valued channel attention module (CCAM) initially processes the input features through max pooling and average pooling operations. Among them, CAvgpool calculates the average value of complex features and outputs both real and imaginary parts, while CMaxpool selects the complex number with the largest modulus and finally outputs a real number. In order to effectively fuse these three types of information, CCAM processes the real part, imaginary part, and magnitude results of CAvgpool and CMaxpool in a weighted manner through a shared MLP to compute the channel attention. Subsequently, channel weights are generated through sigmoid and act on the real and imaginary parts of the input features, respectively, to adaptively adjust the importance of different channels and highlight the key amplitude and phase information. The complex-valued spatial attention mechanism focuses on the spatial position of the feature map, enabling the network to capture key details in the image more accurately. By operating simultaneously in both the channel and spatial dimensions, the CCBAM enhances the network's ability

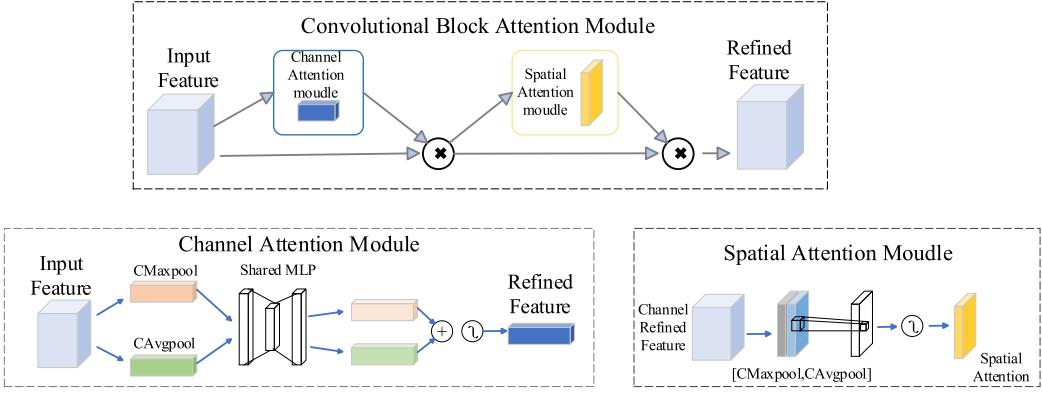


Fig. 6. Schematic diagram of the CCBAM structure.

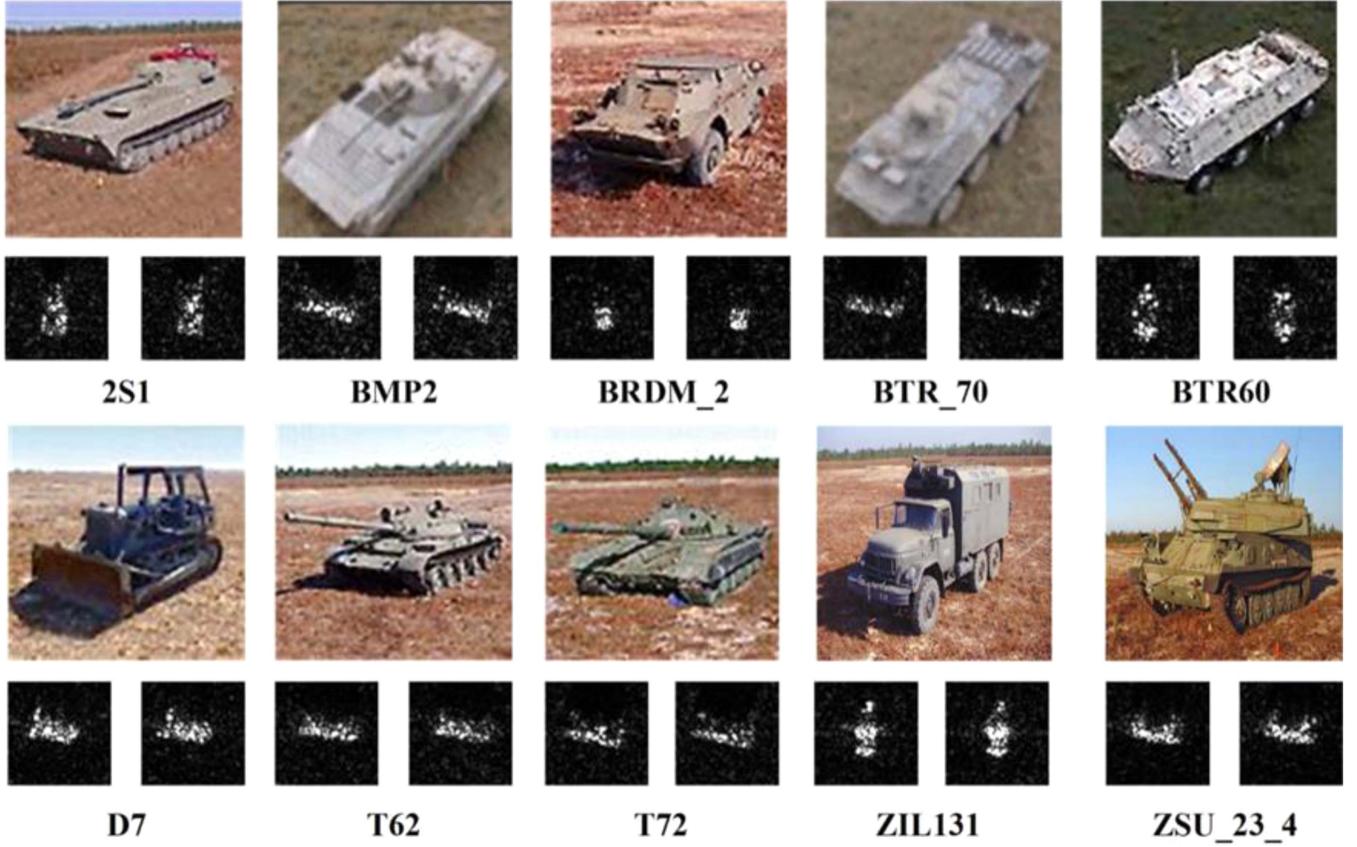


Fig. 7. Comparison of optical and complex SAR images: (Top) optical images; (Bottom) left the real part, right the imaginary part.

to identify and process salient features within complex datasets, ultimately improving classification performance.

IV. EXPERIMENTAL RESULTS ANALYSIS AND DISCUSSION

A. MSTAR Dataset Description and Platform

Using an X-band SAR sensor, we collected the MSTAR dataset for this study. It has a resolution of $0.3\text{ m} \times 0.3\text{ m}$ and includes various target information, such as category, serial number, quantity, azimuth, and depression angle, making it

a standard benchmark for evaluating SAR image recognition algorithms. In addition, it comprises ten distinct categories of ground vehicle targets. Notably, the experimental data employed in the complex-valued network consisted of the original MSTAR data with complex-valued components. Furthermore, the data used in the real-valued network were also processed from this dataset. The corresponding optical and SAR images are shown in Fig. 7. MSTAR target slices can be divided into standard operating conditions (SOC) and extended operating conditions (EOC). Based on the experimental settings described in [42], the

TABLE II
SPECIFIC DETAILS OF THE DATASET USED

Datasets	Training targets	Testing targets	Description
SOC-10	1-2s1, 2-BMP2(SN9563), 3-BRDM2, 4-BTR-60, 5-BTR-70, 6-D7, 7-T62, 8-T72(SN132), 9-ZIL131, 10-ZSU234		Training at 17° and testing at 15°
SOC-3	All variants within the categories BMP2, BTR70, and T72		Training at 17° and testing at 15°
EOC-VV-4	2-BMP2(SN9563), 4-BTR-60, 7-T62, 8-T72(SN132)	2-BMP2(SN9566), 2-BMP2(c21), 4-BTR-60, 7-T62, 8-T72(SN812), 8-T72(S7),	Training at 17° Testing at 15° against BMP2 and T72 version variants

TABLE III
EXPERIMENTAL CONFIGURATION OF SOC-10

Class	Serial No.	Training set 17°	Testing set 15°
1-2S1	B01	299	274
2-BMP2	SN9563	233	195
3-BRDM2	E-71	298	274
4-BTR-60	Kl0yt7532	256	195
5-BTR-70	C71	233	196
6-D7	92v13015	299	274
7-T62	A51	299	273
8-T72	SN132	232	196
9-ZIL131	E12	299	274
10-ZSU234	D08	299	274
Total		2747	2425

standard operating condition - ten classes (SOC-10), standard operating condition - three classes (SOC-3), and extended operating condition - variation in version 4 (EOC-VV-4) datasets are employed. Table II presents the detailed configurations of these datasets.

- 1) *SOC-10*: The SOC-10 dataset contains SAR images of all target types acquired under SOCs. Table III lists the SOC-10 experimental configuration, the training set comprises 2747 samples with a depression angle of 17°, whereas the test set consists of 2425 samples with a depression angle of 15°. The SOC-10 is employed to assess the model's overall performance across all target categories, providing a comprehensive standard test benchmark.
- 2) *SOC-3*: The SOC-3 is a simplified version of SOC-10, comprising three types of target vehicles: BMP2, BTR70, and T72. It is employed to assess the model's capacity for generalization when the number of categories is limited. Specific configurations of the SOC-3 dataset are illustrated

TABLE IV
EXPERIMENTAL CONFIGURATION OF SOC-3

Class	Serial No.	Training set 17°	Testing set 15°
2-BMP2	SN9563	233	195
	SN9566	232	196
	C21	233	196
5-BTR-70	C71	233	196
	SN132	232	196
	SN812	231	195
8-T72	S7	228	191
	Total	1622	1365

TABLE V
EXPERIMENTAL CONFIGURATION OF EOC-VV-4

	Class	Serial No.	No.	Total
Training set 17°	2-BMP2	SN9563	233	1020
	4-BTR-60	Kl0yt7532	256	
	7-T62	A51	299	
	8-T72	SN132	232	
Testing set 15°	2-BMP2	SN9566	196	1246
	2-BMP2	C21	196	
	4-BTR-60	Kl0yt7532	195	
	7-T62,	A51	273	
	8-T72	SN812	195	
	8-T72	S7	191	

in Table IV. The training and test sets comprised 1622 17° and 1365 15° target samples, respectively.

- 3) *EOC-VV-4*: The EOC dataset serves to evaluate the stability of the model under more complex conditions. Table V lists the experimental configuration of the EOC-VV-4. In contrast to the SOC dataset, the targets employed in the EOC-VV-4 test set had serial numbers or versions disparate from those observed in the training set. Owing to the

distinct imaging mechanisms of SAR, different versions of targets may exhibit subtle variations in appearance and radar reflection characteristics. This change makes the model recognition task more challenging. Therefore, it is essential to verify the effectiveness of the recognition approach in the context of target version variations.

In this article, Windows 10 was used as the operating system, PyTorch was employed as the DL framework, and the experiments were executed on a workstation with an Intel i9-10920X processor, 64 GB of memory, and an NVIDIA RTX 3080 Ti graphics processor. During the training process, the cross-entropy loss (CEL) function was employed for error evaluation, with the stochastic gradient descent optimizer utilized to optimize the loss function. The specific configuration is as follows: the initial learning rate was set to 0.0005, the model training batch size was 32, and the training cycle was set to 350 epochs.

To scientifically evaluate the experimental results, we employed the following indicators, accuracy, and average CEL. Its definition can be written as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (13)$$

$$\text{CEL} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C y_{i,c} \log(\hat{y}_{i,c}) \quad (14)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (15)$$

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (16)$$

$$F_1 = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (17)$$

where TN and TP denote the count of correct negatives and positives in the test sample, respectively; FN and FP represent the count of false negatives and false positives in the test sample, respectively. N denotes the count of samples (batch size), C denotes the count of categories, $y_{i,c}$ signifies the true label of sample i for category c , and $\hat{y}_{i,c}$ is the probability distribution of sample i for category c as predicted by the model, where P is the precision rate and R is the recall rate. F_1 is the harmonic average of precision and recall rate, which can measure the comprehensive performance of the detector.

B. Comparative Analysis With State-of-the-Art Methods

To validate the performance of the proposed CRMC-Net, several well-known and commonly adopted real-valued DL recognition methods have been chosen for comparison in the comparative experiments presented in this section. In particular, the selected methods are traditional CNN, VGG16, ResNet18, DenseNet121, CVCNN, CV-FCNN, and CV-Net. The traditional CNN method refers to a small real-valued network structure with four convolutional layers, and CVCNN is the complex-valued version of its corresponding network. The comparative experiments are conducted on the SOC-10, SOC-3, and EOC-VV-4 datasets, which reflect the model's capabilities

with respect to all categories, some specific targets, and version changes.

1) *SOC-10 dataset*: Table VI lists the quantitative test results of eight SAR target recognition models for the SOC-10 dataset. The recognition accuracy of these models ranged from high to low, with the highest accuracy being demonstrated by CRMC-Net (99.83%), followed by CV-Net (99.67), CV-FCNN (99.17%), CVCNN (98.59%), VGG16 (97.65%), DenseNet121 (97.32%), ResNet18 (96.82%), and traditional CNN (96.25%). The CRMC-Net model demonstrated superior performance on the SOC-10 dataset, with an accuracy rate exceeding that of the other seven models. With regard to the test loss, the CRMC-Net also shows a high level of performance, with a loss of 0.0114, which is well below the levels achieved by other models. The test loss reflects the degree of fit of the model when the target sample is processed. A lower loss value indicates a greater degree of accuracy in the model's ability to capture the characteristics of the target sample, thereby demonstrating a superior fit effect. A confusion matrix is used to further illustrate the effectiveness of the proposed method. As illustrated in Fig. 8, the confusion matrix of the CRMC-Net on the SOC-10 dataset demonstrates that the model attains 100% recognition accuracy for seven target categories: BMP2, BTR60, BTR70, T62, ZSU234, T72, and ZIL131, whereas the recognition accuracies for 2S1, BRDM2, and D7 are 99.64%, 99.64%, and 99.27%, respectively. The results demonstrate that the CRMC-Net achieves a superior level of accuracy in the recognition of all target classes. These results provide compelling evidence that the CRMC-Net is an effective method for target recognition.

2) *SOC-3 dataset*: The examination of the SOC-3 dataset enabled the evaluation of the model's capacity for learning and feature representation with respect to specific targets. Table VII and Fig. 9 present the quantitative test results and the corresponding confusion matrix on the SOC-3 dataset, respectively. The test results demonstrate that the CRMC-Net attained an accuracy of 100.00%, which is a markedly superior performance compared with the other seven advanced target recognition methods. The recognition accuracies of the comparison networks are as follows: CV-Net (99.83%), CV-FCNN (99.57%), traditional CNN (99.41%), CVCNN (99.26%), DenseNet121 (99.12%), ResNet18 (99.05%), and VGG16 (98.97%). Regarding the test loss, the CRMC-Net has the lowest error of only 0.0029. Fig. 10 depicts the confusion matrix, which further illustrates that CRMC-Net achieves 100.00% recognition accuracy for BMP2, BTR70, and T72 targets. Generally, the CRMC-Net demonstrates remarkable proficiency in both target recognition accuracy and fitting error, outperforming the existing methods. The results further confirm that incorporating phase information targets error outperforming existing methods. The results further confirm that incorporating phase information targets recognition performance. In addition, complex-valued networks proved highly effective in extracting target information from

TABLE VI
QUANTITATIVE EVALUATION RESULTS FOR THE SOC-10 DATASET

Datasets	Models	Model Type	Loss	Accuracy(%)
SOC-10	Traditional CNN	real-valued	0.3546	96.25
	VGG16	real-valued	0.0866	97.65
	ResNet18	real-valued	0.1469	96.82
	DenseNet121	real-valued	0.1084	97.32
	CVCNN	complex-valued	0.0655	98.59
	CV-Net [39]	complex-valued	0.0225	99.67
	CV-FCNN [41]	complex-valued	0.0439	99.17
	Proposed	complex-valued	0.0114	99.83

Best results are shown in bold.

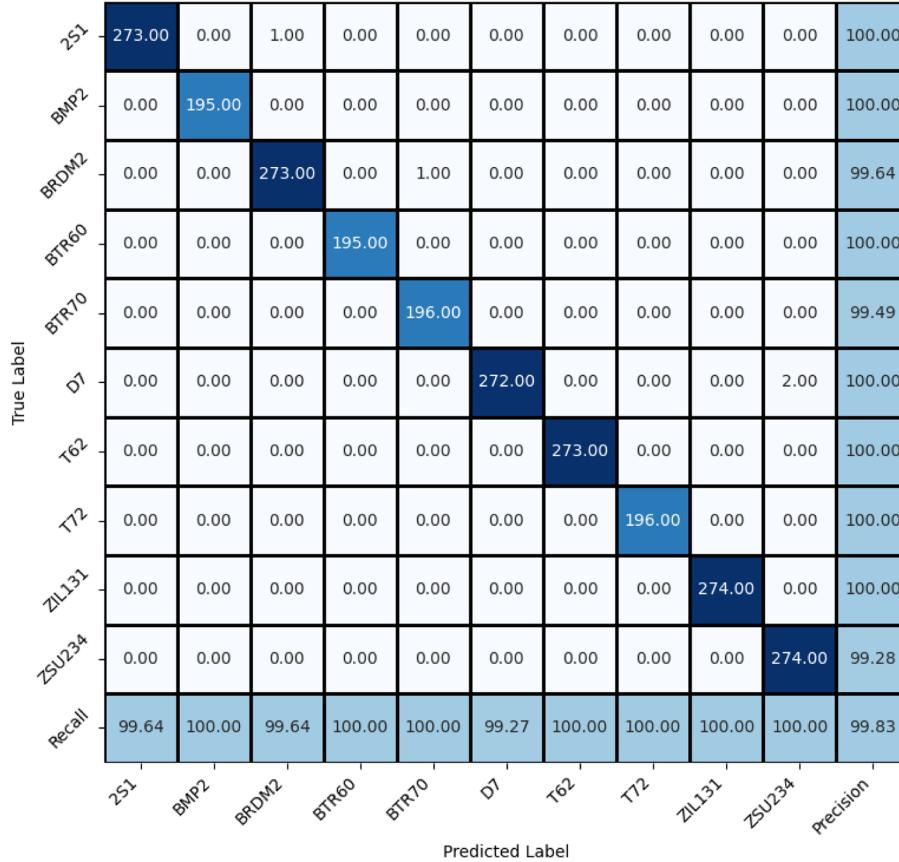


Fig. 8. Confusion matrix of CRMC-Net on SOC-10 dataset.

TABLE VII
QUANTITATIVE EVALUATION RESULTS FOR THE SOC-3 DATASET

Datasets	Models	Model Type	Loss	Accuracy(%)
SOC-3	Traditional CNN	real-valued	0.0273	99.41
	VGG16	real-valued	0.0315	98.97
	ResNet18	real-valued	0.0358	99.05
	DenseNet121	real-valued	0.0324	99.12
	CVCNN	complex-valued	0.0329	99.26
	CV-Net [39]	complex-valued	0.0076	99.83
	CV-FCNN [41]	complex-valued	0.0225	99.57
	Proposed	complex-valued	0.0029	100

Best results are shown in bold.

TABLE VIII
QUANTITATIVE EVALUATION RESULTS FOR THE EOC-VV-4 DATASET

Datasets	Models	Model Type	Loss	Accuracy(%)
EOC-VV-4	Traditional CNN	real-valued	0.3316	91.33
	VGG16	real-valued	0.3849	88.68
	ResNet18	real-valued	0.5477	89.97
	DenseNet121	real-valued	0.3471	91.89
	CVCNN	complex-valued	0.2451	93.09
	CV-Net [39]	complex-valued	0.1727	94.86
	CV-FCNN [41]	complex-valued	0.2226	93.26
Proposed		complex-valued	0.1383	95.02

Best results are shown in bold.

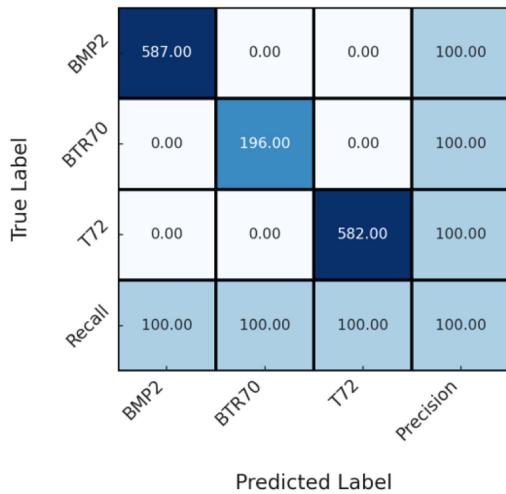


Fig. 9. Confusion matrix of CRMC-Net on SOC-3 dataset.

complex SAR data, significantly improving recognition outcomes.

- 3) *EOC-VV-4 dataset*: In SAR-ATR applications, the dynamic nature of battlefield conditions and evolving mission requirements present a significant challenge to SAR target recognition, particularly in the presence of VV. To address this issue, this section utilizes the EOC-VV-4 dataset to assess the impact of VV on various target recognition models. The degree of accuracy achieved by each model tested on the EOC-VV-4 dataset is presented in Table VIII. The accuracy of the traditional CNN model was 91.33%, with a loss function of 0.3316. The VGG16 model achieved an accuracy of 88.68%, with a loss function of 0.3849. The ResNet18 and DenseNet121 models achieved accuracies of 89.97% and 91.89%, respectively, with loss functions of 0.5477 and 0.3471, respectively. The accuracy of CVCNN is 93.09% and the loss is 0.2451, while the CV-FCNN model demonstrated an accuracy of 93.26% and a loss of 0.2226, indicating an enhanced model performance. The accuracy of the CV-Net model was 94.86%, with a loss function of 0.1727. Nevertheless, the proposed model demonstrated the best performance with an accuracy of 95.02% and a loss function of 0.1383.

TABLE IX
COMPARISON OF DIFFERENT COMPLEX ACTIVATION FUNCTIONS

	CTanh	CElu	CRelu	CMish
Accuracy (%)	98.56	98.80	99.59	99.83
loss	0.0535	0.0471	0.0210	0.0114

In contrast, the proposed CRMC-Net demonstrates superior generalization capabilities with regard to the VV of SAR targets.

C. Comparison of Different Complex Activation Functions

We compare several complex-valued activation functions (CTanh, CEElu, CReLU, and CMish) to evaluate their performance in complex-valued networks. These complex-valued activation functions are extensions of the real-valued activation functions Tanh, Elu, ReLU, and Mish to handle complex inputs. The performance of these four activation functions in SAR image recognition tasks was tested using the proposed model in this article. From the experimental results in Table IX, CTanh performs poorly, with a recognition accuracy of 98.56% and a corresponding loss value of 0.0535, which is significantly lower than other activation functions. Second, CEElu performs slightly better, with a recognition accuracy of 98.80% and a loss value of 0.0471. Although it is better than CTanh, it still fails to significantly optimize network performance. The CReLU activation function achieved relatively ideal results in the complex-valued network, with a recognition accuracy of 99.59% and a loss value of 0.0210, showing a relatively obvious advantage. However, the CMish activation function performed best in all tests, with a 99.83% recognition accuracy and a loss value of only 0.0114, which is significantly better than other activation functions. This shows that CMish as the activation function of complex-valued networks has better adaptability in capturing complex-valued features and improving model performance.

D. Comparative Analysis With Real-Valued Network of the Proposed Method

To further verify the superiority of complex-valued networks in SAR target recognition, we modified the model to its corresponding real-valued version, namely RMC-Net, and conducted comparative analyses using the SOC-10, SOC-3, and EOC-VV-4 datasets. The results are summarized in Table X. The results

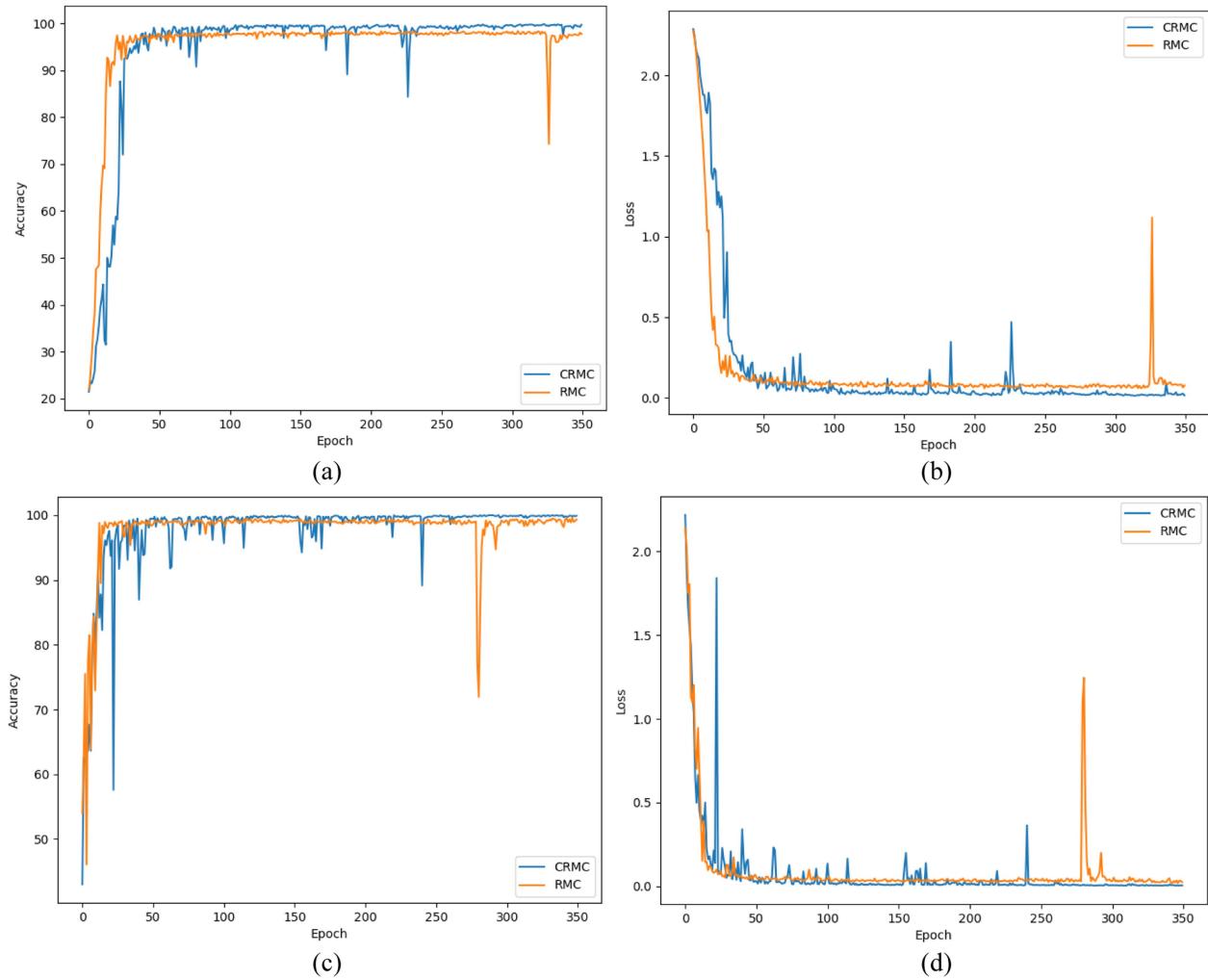


Fig. 10. CRMC-Net and RMC-Net accuracy curves and loss curves. (a) Accuracy curve based on SOC-10 dataset. (b) Loss curve based on SOC-10 dataset. (c) Accuracy curve based on SOC-3 dataset. (d) Accuracy curve based on SOC-3 dataset.

TABLE X
QUANTITATIVE EVALUATION RESULTS FOR THE THREE DATASETS

Datasets	Models	Loss	Accuracy(%)
SOC-10	RMC-Net	0.0504	98.97
	CRMC-Net	0.0114	99.83
SOC-3	RMC-Net	0.0109	99.78
	CRMC-Net	0.0029	100
EOC-VV-4	RMC-Net	0.2503	92.70
	CRMC-Net	0.1383	95.02

Best results are shown in bold.

show that the CRMC-Net model exhibits superior performance compared with the real-valued network (RMC-Net) across all test datasets. In particular, for the SOC-10 dataset, the accuracy of CRMC-Net reached 99.83%, which is notably higher than the accuracy of 98.97% achieved by RMC-Net. Furthermore, the loss value decreased from 0.0504 to 0.0114. Subsequently, on the SOC-3 dataset, the accuracy of CRMC-Net reached 100%, while the accuracy of RMC-Net was 99.78%. The loss

value was also reduced to 0.0029. Furthermore, on the complex EOC-VV-4 dataset, CRMC-Net demonstrated enhanced generalization capabilities, achieving an accuracy of 95.02% and a loss of 0.1383. In comparison, the accuracy of RMC-Net was 92.70% and the loss value was 0.2503. The accuracy and loss curves of CRMC-Net and RMC-Net are shown in Fig. 10. As shown in Fig. 10, all the accuracy curves show an upward trend, whereas the loss curves tend to decrease and eventually

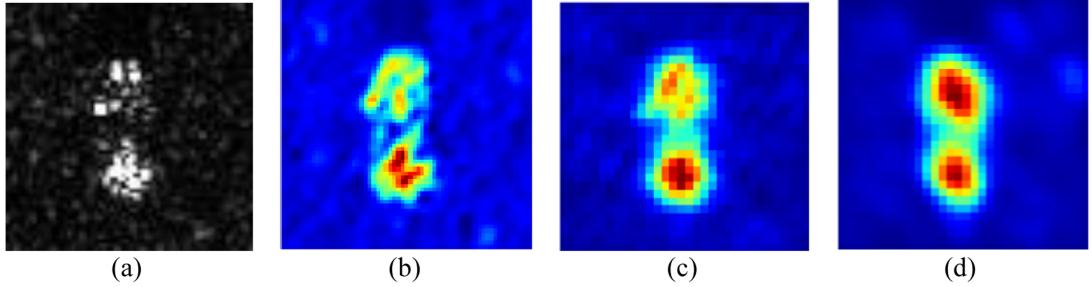


Fig. 11. (a) Original image. (b) RMC-Net feature map visualization. (c) Visualization of CRMC-Net real component feature map. (d) Visualization of CRMC-Net imaginary component feature map.

TABLE XI
COMPARISON OF THE PROPOSED METHOD WITH DIFFERENT MODEL PARAMETERS

Models	Params(MB)	Flops(G)
ResNet18	47.8	1.27
VGG16	138.4	1.42
RMC-Net	14.6	0.22
CRMC-Net	25.8	1.31

converge. However, compared with RMC-Net, CRMC-Net has a significantly higher accuracy and lower loss.

Fig. 11 shows the visualization of the second convolutional layer feature maps of RMC-Net and CRMC-Net. It can be seen from Fig. 11 that the features of the target area in the RMC-Net feature map are relatively fuzzy. However, the real part feature map of CRMC-Net is more prominent, and the features of the target area are significantly enhanced. The imaginary part map provides additional scattering information, further enhancing the identification of the target area. By combining real and imaginary features, the complex-valued network can better capture the physical scattering characteristics in SAR images, thereby improving the accuracy of target recognition. The utilization of complex-valued networks enables a more effective capture and utilization of the phase information inherent to complex-valued signals, thereby markedly enhancing the accuracy of the model, reducing errors, and demonstrating enhanced robustness.

We compare CRMC-Net with other models, such as RMC-Net, VGG16, and ResNet18, about the computation performance. Table XI lists the quantitative comparisons of the four networks tested on the SOC-10 dataset, including the weight parameters #Params and the floating-point operations (FLOPs). First, regarding the number of parameters (Params), CRMC-Net's parameter count stands at 25.8 MB, marginally surpassing RMC-Net's 14.6 MB yet significantly inferior to VGG16's substantial 138.4 MB and ResNet18's considerable 47.8 MB. This shows that although the complex-valued operation introduces certain additional parameters, the number of parameters of CRMC-Net remains within a reasonable range compared with the more classical deep real-valued model. With regard to computational complexity, the number of FLOPs of CRMC-Net is 1.31 G, which is comparable with that of ResNet18 and VGG16, and higher than that of RMC-Net (0.22 G). Although the computational overhead of CRMC-Net is higher than that of

RMC-Net, it can combine amplitude and phase information and significantly improve the target recognition accuracy. Compared with the classic real-valued network, RMC-Net and CRMC-Net both achieve effective optimization in the balance between performance improvement and computing cost.

E. Verification of the Proposed Method on the OPS2.0 Dataset

In addition, in order to evaluate the applicability of the proposed network to other datasets, as shown in Fig. 12, we selected the OpenSARShip2.0 (OPS 2.0) dataset to validate the proposed model. To evaluate the proposed CRMC-Net model on a SAR dataset with different polarization patterns, we selected the OPS 2.0 single-look composite product in dual-polarization VH/VV. We selected three types of ship targets with larger numbers, namely cargo ships, tankers, and other types of ships, for research and discarded the ship targets with smaller numbers. The three types of ship targets were screened and balanced in number. The total number of targets of each type is 510, and the number of samples in the test set is 51, so each type of ship was randomly divided into a training set and a test set in a ratio of 9:1.

In this experiment, we compared with four baseline models: VGG16, ResNet18, RMC-Net, and CVCNN to verify the performance of the proposed CRMC-Net model. The experimental results in Table XII show that the accuracy of CRMC-Net is 70.59%, slightly higher than ResNet18 (69.93%), CVCNN (69.84), and VGG16 (69.28%), and significantly better than RMC-Net (68.63%). These results show that the CRMC-Net model has stronger feature expression ability and better classification performance in complex-valued networks compared with classical real-valued networks. To further highlight the superiority of the complex-valued network, the precision, recall, and F_1 of three types of targets of RMC-Net and CRMC-Net are given in Table XIII. As can be seen from Table XIII, the F_1 scores of CRMC-Net and RMC-Net on Cargo and Other types of targets are similar, but the F_1 of Tanker is significantly higher than that of RMC-Net. By introducing complex-valued operations, amplitude and phase information can be used simultaneously, thereby enhancing the ability to capture the physical scattering characteristics of the target, thereby more effectively extracting target features under complex backgrounds and multipolarization information. Complex-valued operations not only improve the classification performance of the model but also

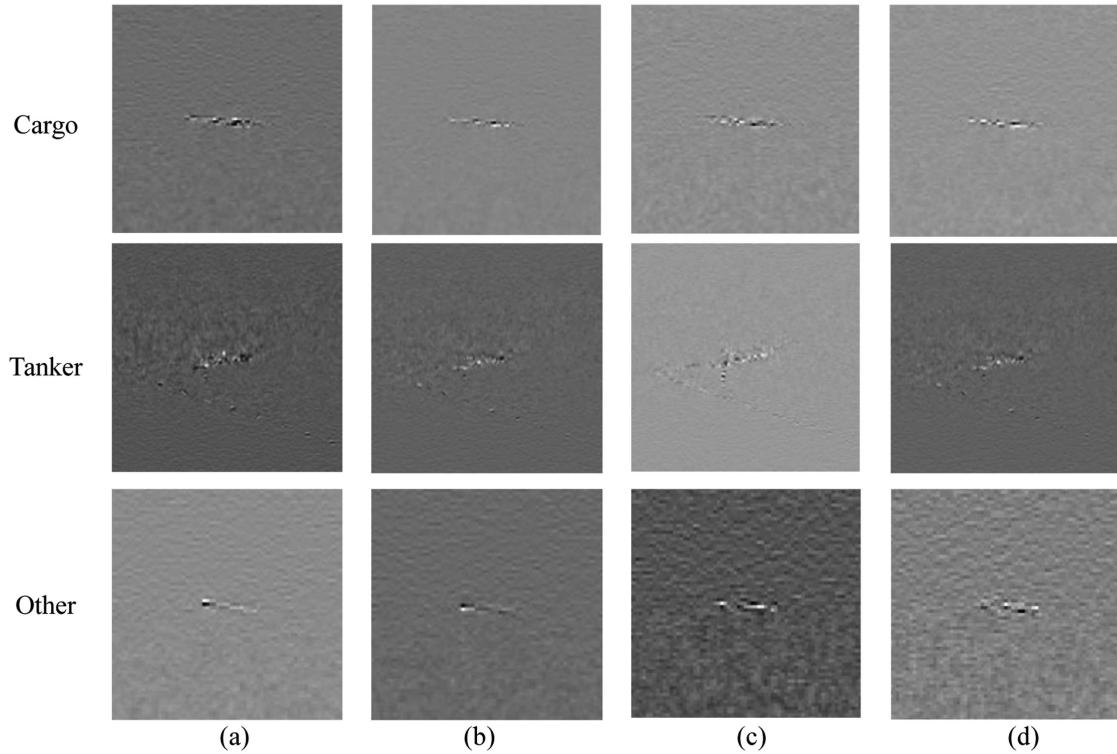


Fig. 12. Example images of OpenSAShip2.0 include three classifications. (a) VH polarization real component. (b) VH polarization imaginary component. (c) VV polarization real component. (d) VV polarization imaginary component.

TABLE XII
QUANTITATIVE EVALUATION RESULTS FOR THE OPS2.0 DATASETS

Datasets	Models	Model Type	Accuracy(%)
OPS 2.0	ResNet18	real-valued	69.93
	VGG16	real-valued	69.28
	RMC-Net	real-valued	68.63
	CVCNN	complex-valued	69.84
	CRMC-Net	complex-valued	70.59

Best results are shown in bold.

TABLE XIII
COMPARISON OF PRECISION, RECALL, AND F1 BETWEEN RMC-NET AND CRMC-NET ACROSS THREE CATEGORIES

Type	Precision(%)		Recall(%)		F1	
	RMC-Net	CRMC-Net	RMC-Net	CRMC-Net	RMC-Net	CRMC-Net
Cargo	64.18	59.49	84.31	92.16	72.84	72.31
Tanker	71.15	83.72	72.55	70.59	71.84	76.60
Other	73.53	80.65	49.02	49.02	58.96	60.98

The bold values represent the optimum values in each category.

increase the robustness and adaptability of the model in complex scenarios.

F. Ablation Experiments

To assess the impact of each module on target recognition, an ablation experiment is conducted using the SOC-10 dataset. The experimental results are presented in Table XIV. It demonstrates that the enhanced version of the CVCNN attained an accuracy of

99.62% with a loss of 0.0306. Following the introduction of the CReLUBlock, an increase in accuracy was observed, reaching 99.67%. The loss value is also reduced to 0.0198. This can be attributed to the effective alleviation of gradient vanishing and degradation issues in the deep network, which ensures the stability of training. In addition, compared with improvement 1, the incorporation of the CMish improves the accuracy of improvement 2 to 99.71%. The loss value is further decreased

TABLE XIV
ABLATION EXPERIMENTAL ANALYSIS OF THE PROPOSED CRMC-NET USING THE SOC-10 DATASET

Method	CReLBlock	CMish	CCBAM	Accuracy(%)	Loss
CVCNN Improvement				99.62	0.0306
Improvement1	✓			99.67	0.0198
Improvement2	✓	✓		99.71	0.0169
Proposed	✓	✓	✓	99.83	0.0114

Best results are shown in bold.

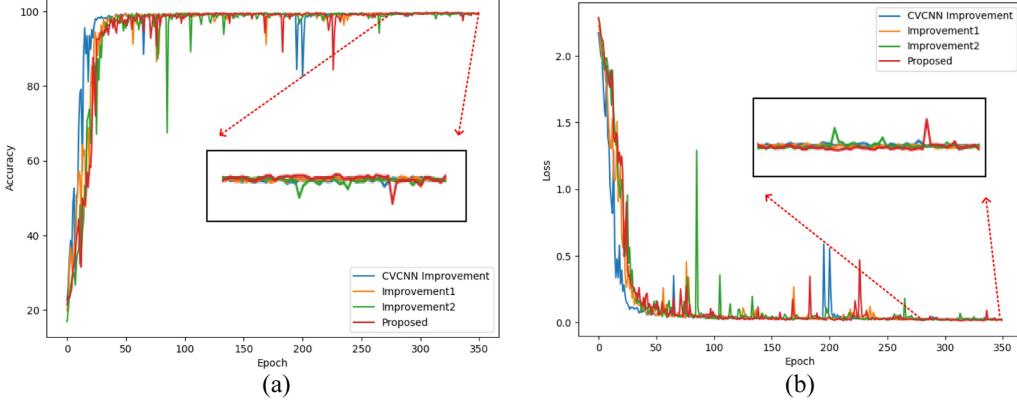


Fig. 13. (a) Test accuracy and the (b) loss over training epochs on the ablation experiment.

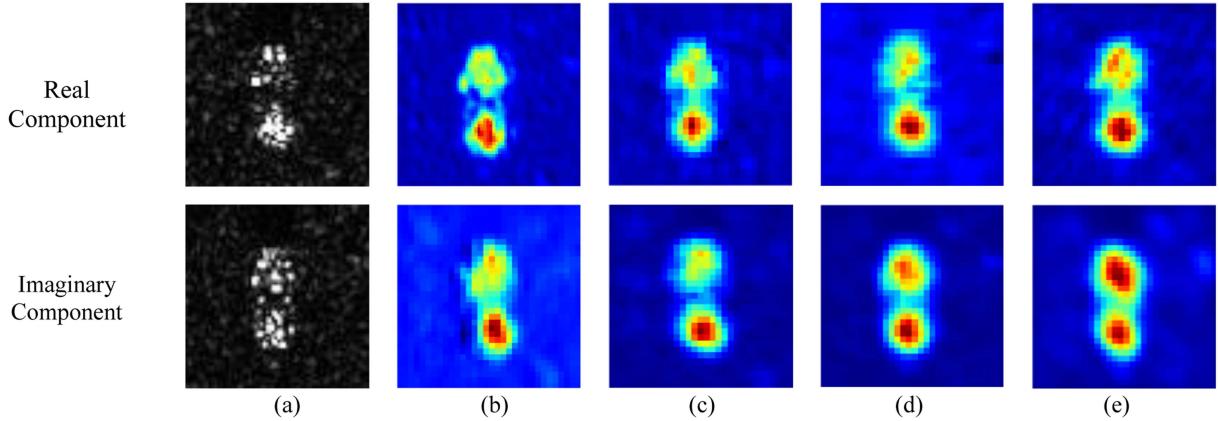


Fig. 14. Some visualization results. (a) Original image. (b) CVCNN improvement. (c) Improvement 1. (d) Improvement 2. (e) Proposed.

to 0.0169. This is primarily due to the fact that the CMish augments the nonlinear expression capability of the model, thereby improving the overall performance. Furthermore, in order to facilitate more precise localization and processing of salient features within intricate data sets, the CCBAM block has been introduced based on improvement 2. This improvement results in an increase in the accuracy of 0.12%, which is 0.21% higher than that observed in the improved version of the original CVCNN. Furthermore, the loss value decreased from 0.0306 to 0.0114.

Fig. 13 illustrates the test loss and accuracy curves obtained from the ablation experiment. In order to intuitively demonstrate the effectivity of our improved module, we selected BTR-70 to conduct visualization experiments on the feature maps of the second convolutional layer of the proposed model. Fig. 14 shows the visualization results, with the addition of each module, the model continues to strengthen the feature expression and ultimately achieves the accurate recognition of the target area. The results

demonstrate that the incorporation of each complex-valued module markedly enhances the recognition performance. Moreover, the efficacy of this method in achieving the highest accuracy can be attributed to its effective utilization of the physical scattering properties contained in complex-valued information inherent to the image.

V. DISCUSSION

It is obvious from the experiments in Section IV that the proposed CRMC-Net performs well in the SAR target recognition task. This is mainly based on the following reasons. First, compared with typical real-valued convolutional networks, the proposed method effectively exploits the physical scattering characteristics and complex-valued information. Second, the guided module significantly improves the feature extraction

capability of the model and further improves the accuracy of target recognition. However, CRMC-Net still has some limitations.

- 1) Currently, the performance of CRMC-Net has not been fully verified on large-scale SAR datasets. The publicly available large-scale complex-valued SAR dataset resources are very limited. Therefore, we need to verify and improve the model on big data in the future.
- 2) Although complex-valued operations can improve performance compared with the real-valued operations under the same net, they also increase the cost of computing. Therefore, in a resource-constrained environment, it is necessary to improve the complex-valued network to balance the computational overhead and the model performance.

Future work will address these limitations and further improve the proposed method to extend its application. In addition, we will optimize the method to increase the recognition accuracy and reduce the computational costs to make it more practical for applications.

VI. CONCLUSION

With the rapid development of DL technology, algorithms based on CNN have made considerable progress in the field of SAR-ATR. However, the unique physical scatter properties of SAR images are often underutilized, especially the phase information. A significant challenge in the field of SAR target recognition is the limited recognition performance of existing DL methods. This is largely due to the fact that many existing DL methods treat SAR images as if they were optical images and extract features from real-valued SAR images based solely on amplitude information. This approach imposes certain limitations on the model's recognition capabilities, which is a major challenge for SAR target recognition tasks. It also significantly hinders the advancement of DL methods. To put forward these challenges, this article proposes a SAR target recognition approach using complex-valued operations, called CRMC-Net. This method fully exploits the amplitude and phase information in complex-valued SAR data by extending all key components of CNN in the complex-valued domain. Furthermore, the CRMC-Net markedly enhances the model's representation capability and resilience through the incorporation of multiscale convolution kernels and the integration of CReLUBlock, CMish, and CCBAM. The experimental results demonstrate that the CRMC-Net model exhibits superior performance on the MSTAR dataset, outperforming traditional real-valued CNNs. This approach has been shown to significantly reduce classification errors and enhance recognition performance. By fully exploiting phase information in SAR images, complex-valued CNN has obvious advantages in target recognition tasks. In the future, we will expand complex-valued CNN to more fields, such as SAR target detection.

REFERENCES

- [1] S. Baraha and A. K. Sahoo, "Synthetic aperture radar image and its despeckling using variational methods: A review of recent trends," *Signal Process.*, vol. 212, Nov. 2023, Art. no. 109156.
- [2] K. El-Darymli, E. W. Gill, P. McGuire, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: A state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.
- [3] N. S. Hashemi, R. B. Aghdam, A. S. B. Ghiasi, and P. Fatemi, "Template matching advances and applications in image analysis," 2016, *arXiv:1610.07231*.
- [4] J. Cheng, Y. Wu, W. AbdAlmageed, and P. Natarajan, "QATM: Quality-aware template matching for deep learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 11553–11562.
- [5] F. A. Somogyi and M. Asztalos, "Systematic review of matching techniques used in model-driven methodologies," *Softw. Syst. Model.*, vol. 19, no. 3, pp. 693–720, Nov. 2020.
- [6] B. Balnarsaiah and G. Rajitha, "Denoising and optical and SAR image classifications based on feature extraction and sparse representation," 2021, *arXiv:2106.01896*.
- [7] F. Gao, T. Huang, J. Sun, J. Wang, A. Hussain, and E. Yang, "A new algorithm for SAR image target recognition based on an improved deep convolutional neural network," *Cogn. Comput.*, vol. 11, pp. 809–824, 2019.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.
- [9] C. Francois, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 1251–1258.
- [10] J. Pei et al., "Multiview deep feature learning network for SAR automatic target recognition," *Remote Sens.*, vol. 13, no. 8, Aug. 2021, Art. no. 1455.
- [11] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.
- [12] F. Su, Q. Lv, and R. Luo, "Review of image classification based on deep learning," *Telecommun. Sci.*, vol. 35, no. 11, pp. 58–74, 2019.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1026–1034.
- [14] X. Feng et al., "Deep learning as applied in SAR target recognition and terrain classification," *J. Radars*, vol. 6, no. 2, pp. 136–148, 2017.
- [15] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.
- [16] Y. Zhu, J. Ai, L. Wu, D. Guo, W. Jia, and R. Hong, "An active multi-target domain adaptation strategy: Progressive class prototype rectification," *IEEE Trans. Multimedia*, vol. 27, pp. 1874–1886, 2025.
- [17] S. Chen, H. Wang, F. Xu, and Y.-Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Apr. 2016.
- [18] J. Ai, Y. Mao, Q. Luo, L. Jia, and M. Xing, "SAR target classification using the multikernel-size feature fusion-based convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5214313.
- [19] B. Hou, H. Kou, and L. Jiao, "Classification of polarimetric SAR images using multilayer autoencoders and superpixels," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 7, pp. 3072–3081, Jul. 2016.
- [20] R. J. Soldin, "SAR target recognition with deep learning," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop*, 2018, pp. 1–8.
- [21] S. Wang, Y. Wang, H. Liu, Y. Sun, and C. Zhang, "A few-shot SAR target recognition method by unifying local classification with feature generation and calibration," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, 2024, Art. no. 5200319.
- [22] Z. Wang et al., "Land-sea target detection and recognition in SAR image based on non-local channel attention network," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Oct. 2022, Art. no. 5237316.
- [23] J. Zheng, M. Li, X. Li, P. Zhang, and Y. Wu, "Revisiting local and global descriptor-based metric network for few-shot SAR target classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 62, Feb. 2024, Art. no. 5205814.
- [24] I. Agadakos, N. Agadakos, J. Polakis, and M. R. Amer, "Deep complex networks for protocol-agnostic radio frequency device fingerprinting in the wild," 2019, *arXiv:1909.08703*.
- [25] B. Zhao, Q. Song, J. Li, W. Liu, G. Liu, and Y. Zhao, "High-frequency-link DC transformer based on switched capacitor for medium-voltage DC power distribution application," *IEEE Trans. Power Electron.*, vol. 31, no. 7, pp. 4766–4777, Jul. 2016.
- [26] Z. Wang, R. Wang, X. Fu, and K. Xia, "Unsupervised ship detection for single-channel SAR images based on multiscale saliency and complex signal kurtosis," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2022, Art. no. 4011305.
- [27] V. Macchiarulo, G. Giardina, P. Milillo, Y. D. Aktas, and M. R. Z. Whitworth, "Integrating post-event very high resolution SAR imagery and machine learning for building-level earthquake damage assessment," *Bull. Earthq. Eng.*, pp. 1–27, Mar. 2024.

- [28] H. Bi, G. Bi, B. Zhang, and W. Hong, "Complex-image-based sparse SAR imaging and its equivalence," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 9, pp. 5006–5014, Sep. 2018.
- [29] R. Chakraborty, "Manifoldnorm: Extending normalizations on Riemannian manifolds," 2020, *arXiv:2003.13869*.
- [30] B. O. Raeker and A. Grbic, "Compound metaoptics for amplitude and phase control of wave fronts," *Phys. Rev. Lett.*, vol. 122, no. 11, Mar. 2019, Art. no. 113901.
- [31] M. Wilmanski, C. Kreucher, and A. Hero, "Complex input convolutional neural networks for wide angle SAR ATR," in *Proc. IEEE Glob. Conf. Signal Inf. Process.*, 2016, pp. 1037–1041.
- [32] B. Ding and G. Wen, "Target reconstruction based on 3-D scattering center model for robust SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 7, pp. 3772–3785, Jul. 2018.
- [33] T. Li and L. Du, "Target discrimination for SAR ATR based on scattering center feature and K-center one-class classification," *IEEE Sensors J.*, vol. 18, no. 6, pp. 2453–2461, Mar. 2018.
- [34] N. Guberman, "On complex valued convolutional neural networks," 2016, *arXiv:1602.09046*.
- [35] C. Trabelsi et al., "Deep complex networks," 2017, *arXiv:1705.09792*.
- [36] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, Dec. 2017.
- [37] L. Li, L. Ma, L. Jiao, F. Liu, Q. Sun, and J. Zhao, "Complex contourlet-CNN for polarimetric SAR image classification," *Pattern Recognit.*, vol. 100, Apr. 2020, Art. no. 107110.
- [38] Y. Cao, Y. Wu, P. Zhan, W. Liang, and M. Li, "Pixel-wise PolSAR image classification via a novel complex-valued deep fully convolutional network," *Remote Sens.*, vol. 11, no. 22, Nov. 2019, Art. no. 2653.
- [39] R. Wang, Z. Wang, K. Xia, H. Zou, and J. Li, "Target recognition in single-channel SAR images based on the complex-valued convolutional neural network with data augmentation," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 2, pp. 796–804, Apr. 2023.
- [40] Z. Sun, X. Xu, and Z. Pan, "SAR ATR using complex-valued CNN," in *Proc. Asia-Pacific Conf. Image Process., Electron. Comput.*, 2020, pp. 125–128.
- [41] L. Yu, Y. Hu, X. Xie, Y. Lin, and W. Hong, "Complex-valued full convolutional neural network for SAR target classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 10, pp. 1752–1756, Oct. 2020.
- [42] Z. Zeng, J. Sun, Z. Han, and W. Hong, "SAR automatic target recognition method based on multi-stream complex-valued networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, May 2022, Art. no. 5228618.
- [43] Z. Wang, R. Wang, H. Kang, F. Luo, and J. Ai, "CV-SAR-det: Target detection for SAR images via deep complex-valued network," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 60, no. 6, pp. 8226–8238, Dec. 2024.
- [44] R. Wang, Z. Wang, Y. Chen, H. Kang, F. Luo, and Y. Liu, "Target recognition in SAR images using complex-valued network guided with sub-aperture decomposition," *Remote Sens.*, vol. 15, no. 16, Aug. 2023, Art. no. 4031.
- [45] D. Zhao, Z. Zhang, D. Lu, J. Kang, X. Qiu, and Y. Wu, "CVGG-Net: Ship recognition for SAR images based on complex-valued convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 20, Sep. 2023, Art. no. 4010805.
- [46] Q. Hua, Y. Zhang, Y. Jiang, and D. Xu, "CV-CFUNet: Complex-valued channel fusion UNet for refocusing of ship targets in SAR images," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 59, no. 4, pp. 4478–4492, Aug. 2023.
- [47] C. Li, L. Du, Y. Li, and J. Song, "A novel SAR target recognition method combining electromagnetic scattering information and GCN," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, Jun. 2022, Art. no. 4508705.
- [48] D. Misra, "Mish: A self regularized non-monotonic activation function," 2019, *arXiv:1908.08681*.



Guangyu Hou was born in Hebei, China, in 2000. He received the B.S. degree in electronic information from the School of Science and Engineering, Hebei University of Science and Technology, Shijiazhuang, China, in 2021. He is currently working toward the M.S. degree in electronic information with the School of Physics and Electronic Information, Yunnan Normal University, Yunnan, China.

His research interest focuses on SAR image detection.



Zhihui Xin was born in Shanxi, China. She received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 2012 and 2017, respectively.

In 2017, she joined the School of Physics and Electronic Information Technology, Yunnan Normal University, Kunming, China, where she is currently an Associate Professor. Her research interests include space-time adaptive processing, sea clutter modeling and simulation, and maritime target detection.



Guisheng Liao (Senior Member, IEEE) was born in Guilin, China. He received the B.S. degree in mathematics from Guangxi University, Nanning, China, in 1985, and the M.S. and Ph.D. degrees in signal and information processing from Xidian University, Xi'an, China, in 1990 and 1992, respectively.

He is currently a Professor with Xidian University, where he is also the Dean of the School of Electronic Engineering. His research interests include synthetic aperture radar (SAR), space-time adaptive processing, SAR ground moving target indication, and distributed small satellite SAR system design.



Penghui Huang (Member, IEEE) was born in Jiangxi, China. He received the B.S. and Ph.D. degrees in electrical engineering from Xidian University, Xi'an, China, in 2012 and 2017, respectively.

In 2017, he joined the School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University, Shanghai, China, where he is currently an Associate Professor. His research interests include space-borne radar system design and array signal processing.



Yuhao Huang was born in Sichuan, China, in 1998. He received the B.S. degree in electronic information from Chengdu Technological University, Sichuan, China, in 2022. He is currently working toward the M.S. degree in electronic information with the School of Physics and Electronic Information, Yunnan Normal University, Yunnan, China.

His research interests include synthetic aperture radar image change detection.



Rui Zou was born in Gansu, China, in 1999. She received the B.S. degree in information countermeasure technology from the Changchun University of Science and Technology, Changchun, China, in 2021. She is currently working toward the M.S. degree in optical engineering with the School of Physics and Electronic Information, Yunnan Normal University, Yunnan, China.

Her research interest focuses on multispectral remote sensing image detection.