A Novel SAR Automatic Target Recognition Method Based on Fully Complex-Valued Networks

Yuejun Zhu, Tao Li, Dongliang Peng, *Member, IEEE*, Haoran Wang, and Sainan Shi

***Abstract*—The existing automatic target recognition (ATR) methods for synthetic aperture radar (SAR) images mainly utilize the real-valued magnitude information while often ignoring the phase information. However, the phase information also provides important details, which can be utilized to improve the ATR performance. To address this issue, a fully complex-valued light-weight network (CVLWNet) is proposed based on complex-valued operations, such as complex-valued convolution and complex-valued batch normalization. Besides, to achieve reduced parameters and enhanced robustness of the designed network, many complex-valued blocks of operations are built, including the CMish activation function, the complex-valued residual link block (CVReLBlock), the lightweight complex-valued cross stage partial block (LC-CSPBlock). In the designed CVLWNet, the input, output, and weight parameters are all complex-valued, which makes it possible to sufficiently exploit the complex-valued characteristics of SAR data. Comparative experiments are conducted with the moving and stationary target acquisition and recognition dataset. Compared with the state-of-the-art real-valued and complex-valued models under both standard and extended operating conditions, the performance of proposed method is verified.**

***Index Terms*—Complex-valued network, deep learning, radar target recognition, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) has unique advantages in remote sensing with the capability of all-day observation under severe weather conditions. In recent years, it has played an increasingly important role in every corner of the military field, especially in the automatic target recognition (ATR) application. In the earlier stage, radar ATR methods generally recognize targets based on the feature handcrafted with domain knowledge by principles of template matching or machine learning classifiers. However, the performance of target recognition suffers severe degradation with nonrobustness of feature extraction methods under varying imaging scenarios.

Since single-channel SAR amplitude information can be presented as grayscale images and multipolarized SAR data as pseudocolor images, different kinds of convolutional neural networks (CNNs) such as AlexNet [1], VGGNet [2], and ResNet [3] were transferred from networks pretrained on the natural image dataset to SAR images with fine-tuning. However, due to the obvious difference between the imaging mechanism of SAR images and that of optical images, the fine-tuned models pretrained with optical image datasets cannot achieve good performance for SAR ATR [4].

To utilize the abundant information embedded in SAR data and exploit the potential of CNN, many modified CNNs were proposed in recent researches. In view of the problem that the traditional CNN cannot effectively use the component information of the target that is stable to the local change, Li et al. [5] proposed a multiscale CNN (CA-MCNN), which comprehensively used the global features and components of the target features, yielding good performance with the well-learned target information. Though it is often ignored by the ordinary CNNs, the shape information is more effective than the texture information for SAR ATR. Therefore, Huang et al. [6] built a dataset based on the extracted target and their shadow features, which enabled the CNN to pay more attention to the shape of targets. To improve the CNN's recognition capability in the limit of small sample sizes, Lang et al. [7] proposed the hinge loss CNN module to build a lightweight cascaded multidomain attention network, which extracted both the frequency and wavelet transform domain features of feature maps. Chen et al. [8] used 1×1 convolution instead of fully connected layers to build the new all-convolutional networks (A-ConvNet), achieving a breakthrough target recognition accuracy on the moving and stationary target acquisition and recognition (MSTAR) dataset. The existence of speckle noise, difficulty in feature expression and lack of sample data pose a severe challenge to the performance of SAR target recognition. Mohammadimanesh et al. [9] efficiently solved the above problem using the fully convolutional network (FCN) with an encoder–decoder paradigm.

The aforementioned methods were based on the real-valued network architectures and SAR amplitude information. It is noted that SAR images are one typical kind of naturally complex-valued data, of which both the magnitude and phase contain useful information for the ATR task. In particular, the prior encodes the quantity of energy, while the latter denotes

Manuscript received 14 April 2023; revised 31 May 2023 and 29 June 2023; accepted 2 July 2023. Date of publication 5 July 2023; date of current version 12 July 2023. This work was supported in part by the National Natural Science Foundation of China under Grant 62201184 and Grant 61901224, and in part by the Zhejiang Provincial Natural Science Foundation of China under Grant LZ23F030002. *(Corresponding authors: Tao Li; Dongliang Peng.)*

Yuejun Zhu and Dongliang Peng are with the School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China (e-mail: zhuyuejun@hdu.edu.cn; dlpeng@hdu.edu.cn).

Tao Li is with the School of Automation, Hangzhou Dianzi University, Hangzhou 310018, China, and also with the Nanjing Research Institute of Electronics Technology, Nanjing 210039, China (e-mail: litao@hdu.edu.cn).

Haoran Wang is with the University of Science and Technology of China, Hefei 230026, China, and also with the Sungrow Power Supply Company, Ltd., Hefei 230088, China (e-mail: whr.cloud9@gmail.com).

Sainan Shi is with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China (e-mail: snshi@nuist.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2023.3292315

the material of the object and forms boundaries [10], [11]. However, in the majority of existing methods for target detection and recognition in SAR images, the magnitude information is widely applied while the phase variation is paid with less attention. To fully utilize the abundant information contained in complex-valued SAR data, two typical approaches have been studied in recent literatures. One type of methods was based on the integration of electromagnetic scattering feature (ESF) into the real-valued neural network architecture to enhance the interpretability, while the other one explored the complex-valued version of neural networks. Following the first pattern, the part convolution was utilized by Feng et al. [12] to learn the characteristic of parts model obtained from the attributed scattering centers (ASCs), then the bidirectional convolutional-recurrent network was modified to realize local feature extraction. Then, Feng et al. [13] designed an ASC parameter extractor using the CNN structure, based on which the extracted ESF characteristics were incorporated into the classification network. Furthermore, a part attention network based on the modeling of ASC model was proposed in [14], which facilitated the interpretability of deep learning network. Similar works have been done by Zhang et al. in [15], Li et al. in [16], and Liu et al. in [17], and excellent performance can be obtained.

With respect to the modification of networks, the complex-valued neural networks (CVNNs) are developed to fully investigate the complex-valued representation of SAR data. Since the polarimetric SAR (PolSAR) images are generally represented by the covariance or coherency matrices with naturally complex-valued property, many complex-valued CNNs were first explored in the application of PolSAR image classification. To apply the convolution structure and weight parameters of the real-valued networks to the complex-valued domain, Zhang et al. [18] proposed a complex-valued CNN for the terrain classification task of PolSAR images. A novel FCN architecture was proposed for the PolSAR data in [19], which consisted of the complex-valued max-unpooling and complex-valued downsampling and used a novel average cross-entropy loss function to accelerate the learning ability of the network. To take full advantage of the polarization dispersion contained in the complex-valued covariance and coherence matrix, Tan et al. [20] proposed a complex-valued 3-D CNN that can extract both spatial feature information and polarization scattering information.

To sum up, there are currently two typical forms of CVNNs. The first solution is to apply real-valued networks to the isomorphic two-way representation of complex-valued data, which were started with split-CVNNs. Specifically, split-CVNNs can be roughly divided into two classes, both of which have real-valued activation functions, while the weights can be either real- or complex-valued. [21]. Based on the multistream structure, Zeng et al. [22] proposed the MSCVNets by constructing convolution modules with different convolution sizes, and fused the multistream feature maps output by the modules. The experimental results on the MSTAR dataset show that MSCVNets achieved excellent recognition results. Since it uses real-valued backpropagation for gradient updates, it can cause a phase distortion of complex-valued input data and a significant deviation of the output values of the network during network training [23].

The second form of CVNNs extends the real-valued network to the fully complex-valued domain, including the convolution operation, batch normalization, activation function, and pooling function, etc. In order to process SAR data with phases, the pure complex-valued CNNs were presented by [24], where weight parameters and activation functions were both learned and calculated in the complex-valued domain. In complex-valued convolutional neural networks, too many weight parameters can also cause overfitting problems. To address this issue, Yu et al. [25] replaced the fully connected and pooling layers with complex-valued 1×1 and 3×3 convolutions, respectively. Therefore, the designed network consisted only of convolution operations, which achieved improved classification accuracy on the MSTAR dataset. Wilmanski et al. [26] inputted the complex-valued data into the first complex-valued convolution layer of the network while the next convolution module used real-valued convolution. A performance jump of 87.9% to 99.21% was achieved on the real collected wide angle SAR data with the use of complex-valued features. In [27], two major problems were discovered for the CVNNs, one of which is limited training data and another is the mismatch between measurement data and synthetic data. Besides, to realize the convergence analysis for CVNNs, Zhang et al. [28] obtained the convergence theorem of the fully complex-valued minibatch gradient algorithm applicable to any weight initialization method when the training samples were input into the network in a fixed order.

To thoroughly utilize the abundant information contained in complex SAR data, a novel complex-valued lightweight network (CVLWNet), is proposed to enhance the feature extraction ability of CNNs for the SAR ATR. In addition, the complex-valued version of MSTAR dataset [21], [27] with a complex-valued signal similar to the original radar echo is used in the experiment. The key contributions of the proposed method can be summarized as follows.

1) A new complex-valued activation function CMish is adopted to perform complex domain derivation in all quadrants [29], which inherits excellent performance of Mish function in the real number domain [30]. By comparison, the complex-valued version of ReLU activation function, i.e., CReLU, cannot be derived in the second and fourth quadrants, resulting in a large amount of information loss.
2) The complex-valued residual link block called complex-valued residual link block (CVReLBlock) is proposed, which greatly improves the robustness of the network in the recognition process. When its convolution kernel is large enough, it can suppress clutter and improve recognition performance in complex-valued scenes.
3) A fully complex-valued CNN is constructed, where the input, output, and weight parameters are all complex-valued. Referring to the existing real-number domain lightweight network structure with excellent performance [31], [32], [33], [34], [35], the lightweight complex cross stage partial block, represented by complex-valued cross stage partial block (LC-CSPBlock), is proposed.

The rest of this article is organized as follows. A brief introduction of related works on the basic complex-valued CNN blocks is provided in Section II. Section III presents a detailed

theory for the proposed CVLWNet. The experimental setting and comparative analysis of results are provided in Section IV. Finally, Section V concludes the article.

II. Related Works on Complex-Valued Operation Building Blocks

Trabelsi et al. [29] proposed a general framework for CNNs based on complex-valued operations, which consists of the major modules of the complex-valued CNNs, followed by Zeng et al. [22] who built a general complex-valued framework for SAR ATR. In the Cartesian coordinate system, a complex-valued number can be denoted as $z = x + i \cdot y$ with $i = \sqrt{-1}$, yielding the magnitude $U = \sqrt{x^2 + y^2}$ and phase angle $\theta = \tan^{-1}(y/x)$.

A. Derivation of Wirtinger Calculus in the Process of Backpropagation

The Cauchy–Riemann differential equations are two partial differential equations that provide sufficient and necessary conditions for differentiable functions to be holomorphic functions in the open set [36]. The conventional complex-valued function derivative method can be used for holomorphic functions. However, it is no longer applicable to the nonholomorphic functions. For instance, the cross-entropy loss function widely used in the classification process is nonholomorphic, which does not meet the differential derivative conditions of the Cauchy–Riemann equation. Therefore, it cannot be derived using the traditional complex-valued number derivative rule.

To address this issue, the Wirtinger calculus can be utilized. Consider $\varphi(z) : \mathbb{C} \to \mathbb{C}$ with a format

$$\varphi(z) = \mathcal{U}(x, y) + i \cdot \mathcal{V}(x, y) \tag{1}$$

where $z = x + i \cdot y$. Substitute

$$x = \frac{z + \bar{z}}{2}, y = \frac{z - \bar{z}}{2i} \tag{2}$$

into complex-valued (1), then φ can be rewritten as a bivariate function of z and $\bar{z}$. If the complex-valued function $\varphi(z)$ is a holonomic function, it can be written as $\frac{\partial \varphi}{\partial x}$ and $\frac{\partial \varphi}{\partial y}$ under the traditional complex-valued number derivative rule. However, if the complex-valued function $\varphi(z)$ is a nonholonomic function, it does not satisfy the Cauchy–Riemann differential equation and cannot be differentiated under the traditional complex-valued number derivative rule. Under the derivation of Wirtinger calculus, the nonholonomic function $\varphi(z)$ can be differentiated and therefore, can be written as $\frac{\partial \varphi}{\partial z}$ and $\frac{\partial \varphi}{\partial \bar{z}}$. The derivative chain rule is shown below as follows:

$$\begin{cases} \dfrac{\partial \varphi}{\partial z} = \dfrac{1}{2}\left(\dfrac{\partial \varphi}{\partial x} - i\dfrac{\partial \varphi}{\partial y}\right), \\ \dfrac{\partial \varphi}{\partial \bar{z}} = \dfrac{1}{2}\left(\dfrac{\partial \varphi}{\partial x} + i\dfrac{\partial \varphi}{\partial y}\right). \end{cases} \tag{3}$$

The following property holds:

$$\begin{cases} \overline{\dfrac{\partial \varphi}{\partial z}} = \dfrac{\partial \bar{\varphi}}{\partial \bar{z}}, \ \mathcal{V}(x, y) \neq 0, \\ \dfrac{\partial \varphi}{\partial z} = \dfrac{\partial \varphi}{\partial \bar{z}}, \ \mathcal{V}(x, y) = 0. \end{cases} \tag{4}$$

It can be seen that the Wirtinger calculus provides a convenient method for computing two partial derivatives of a nonholonomic function. Therefore, the Wirtinger calculus provides the basis for designing gradient-based optimization algorithms to minimize the cost function of the real-valued variables, allowing the CVNNs to learn the features of the data.

B. Complex-Valued Convolution

To realize convolution operations on the complex-valued domain, the convolutions of real-valued networks are utilized to simulate that of complex-valued ones. To be specific, the complex-valued feature map $\mathbf{h} = \mathbf{x} + i \cdot \mathbf{y}$ can be multiplied by the complex-valued filter $\mathbf{w} = \mathbf{A} + i \cdot \mathbf{B}$ where $\mathbf{A}$, $\mathbf{B}$, $\mathbf{x}$, and $\mathbf{y}$ are real-valued matrix. Since the convolution operator satisfies the distributive law of the four arithmetic operations, the result of feature map $\mathbf{h}$ convoluted through the filter $\mathbf{w}$ can be denoted as follows:

$$\mathbf{w} * \mathbf{h} = (\mathbf{A} * \mathbf{x} - \mathbf{B} * \mathbf{y}) + i \cdot (\mathbf{B} * \mathbf{x} + \mathbf{A} * \mathbf{y}). \tag{5}$$

The overall calculation process of complex-valued number convolution is shown in Fig. 1. The matrix version of the convolution operation can be given by

$$\begin{bmatrix} \Re(\mathbf{W} * \mathbf{h}) \\ \Im(\mathbf{W} * \mathbf{h}) \end{bmatrix} = \begin{bmatrix} \mathbf{A} & -\mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{bmatrix} * \begin{bmatrix} \mathbf{x} \\ \mathbf{y} \end{bmatrix}, \tag{6}$$

where $\Re(\cdot)$ and $\Im(\cdot)$ represent the real and imaginary operations, respectively.

C. Complex-Valued Batch Normalization

Batch normalization (BN) [37] can transform data with different distributions in a neural network into uniform Gaussian distributed data, which can accelerate the learning process of networks, achieving the network convergence as soon as possible. To a certain extent, the BN can prevent the problems of gradient disappearance and explosion during the training process. However, the current BN algorithm can only transform real-valued data into Gaussian distribution. In this section, a brief introduction to complex-valued BN is provided. For a complex-valued matrix $\mathbf{x}$, denoted as $\mathbf{x} = \Re(\mathbf{x}) + i \cdot \Im(\mathbf{x})$, its standard normal complex-valued distribution $\tilde{\mathbf{x}}$ can be calculated by multiplying the data $\mathbf{x} - \mathrm{E}[\mathbf{x}]$ centered at 0 and the matrix $\mathbf{V}$ with reciprocal square roots of dimension 2, denoted as

$$\tilde{\mathbf{x}} = \mathbf{V}^{-\frac{1}{2}}(\mathbf{x} - \mathrm{E}[\mathbf{x}]) \tag{7}$$

where

$$\begin{aligned} \mathbf{V} &= \begin{pmatrix} \mathbf{V}_{rr} & \mathbf{V}_{ri} \\ \mathbf{V}_{ir} & \mathbf{V}_{ii} \end{pmatrix} \\ &= \begin{pmatrix} \mathrm{Cov}(\Re(\mathbf{x}), \Re(\mathbf{x})) & \mathrm{Cov}(\Re(\mathbf{x}), \Im(\mathbf{x})) \\ \mathrm{Cov}(\Im(\mathbf{x}), \Re(\mathbf{x})) & \mathrm{Cov}(\Im(\mathbf{x}), \Im(\mathbf{x})) \end{pmatrix} \end{aligned} \tag{8}$$

and $\mathrm{Cov}(\cdot)$ and $\mathrm{E}(\cdot)$ represents the covariance calculation and expectation operation, respectively.

The real-valued BN is extended to the complex-valued domain with two parameters, i.e., the shift parameter β and scaling

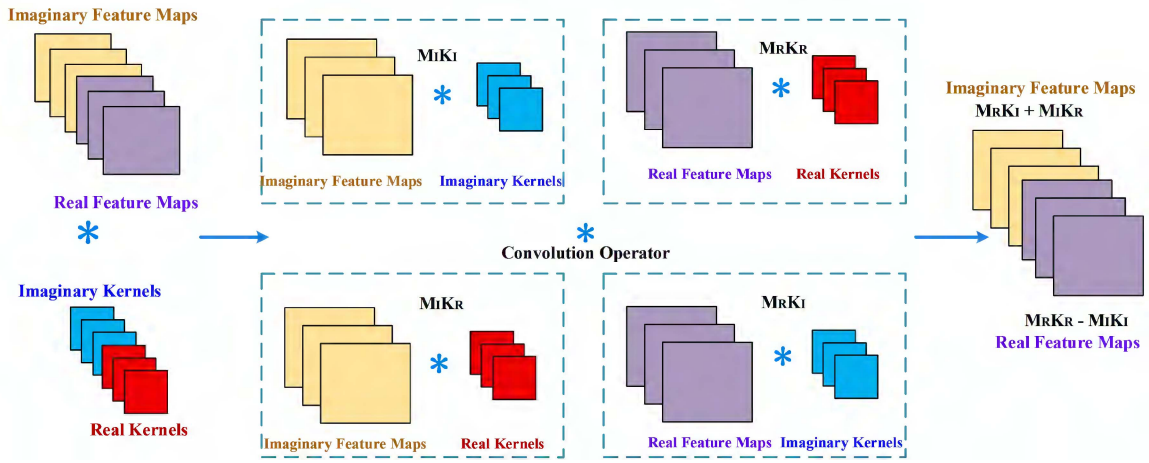


Fig. 1. Complex number convolution operation process.

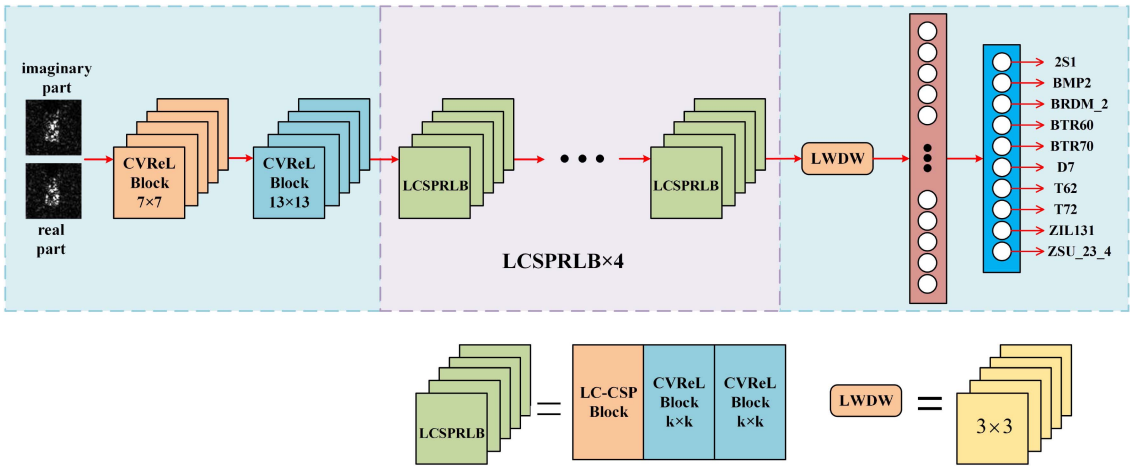


Fig. 2. Architecture of the proposed CVLWNet.

parameter γ. The corresponding complex-valued BN is defined

$$\mathrm{BN}\left(\tilde{\mathbf{x}}\right) = \gamma\tilde{\mathbf{x}} + \beta. \tag{9}$$

Specifically, β is a complex-valued parameter whose real and imaginary parts are learnable parameters. Besides, γ is a 2×2 symmetric matrix which contains three learnable parts, including γ_{rr}, γ_{ii}, and γ_{ri}, which is given by

$$\gamma = \begin{pmatrix} \gamma_{rr} & \gamma_{ri} \\ \gamma_{ri} & \gamma_{ii} \end{pmatrix}. \tag{10}$$

After the normalization, both real and imaginary variances of $\tilde{\mathbf{x}}$ are 1. Therefore, to ensure the variance modulus of $\tilde{\mathbf{x}}$ to be 1, both γ_{rr} and γ_{ii} are initialized to $1/\sqrt{2}$. Besides, γ_{ri}, $\Re(\beta)$, and $\Im(\beta)$ are set with an initialization value of 0.

III. Modified Complex-Valued Neural Networks

The CVLWNet is systematically formulated in this section, whose architecture is shown in Fig. 2. The CVLWNet is constructed with basic blocks including the CMish activation, CVReLBlock, and LC-CSPBlock. The detailed construction and parameter configuration of the proposed network are given in Table I, where the output size of feature map, the kernel size and stride of convolution, the number of repeats, and the output channels are provided for each module. First, two CVReLBlocks are designed with kernel sizes 7×7 and 13×13 (denoted as CVReLBlock 7×7 and CVReLBlock 13×13), respectively. Besides, the two basic blocks CVReLBlock and LC-CSPBlock are further adopted to design the module called LCSPRLB. In the proposed network, four LCSPRLB modules are designed with the LC-CSPBlock and CVReLBlock blocks. Specifically, the first LCSPRLB modules consist of one LC-CSPBlock and two CVReLBlocks with kernel size 7×7 in series. The second and third LCSPRLB modules consist of one LC-CSPBlock and two CVReLBlocks with kernel size 5×5 in series. The last LCSPRLB consists of one LC-CSPBlock and one CVReLBlock with kernel size 5×5 in series. Then, the feature map extracted from the backbone part is sent to the LWDW block to realize

TABLE I
PARAMETER CONFIGURATION OF THE PROPOSED CVLWNET

Layer	Output size	Kernel size	Stride	Repeat	Output channels
Image	64×64				1
CVReLBlock	64×64	7×7	1	1	1
CVReLBlock	64×64	13×13	1	1	32
LC-CSPBlock	32×32	3×3	2	1	64
CVReLBlock	32×32	7×7	1	2	64
LC-CSPBlock	16×16	3×3	2	1	128
CVReLBlock	16×16	5×5	1	2	128
LC-CSPBlock	8×8	3×3	2	1	256
CVReLBlock	8×8	5×5	1	2	256
LC-CSPBlock	4×4	3×3	2	1	512
CVReLBlock	4×4	5×5	1	1	512
LWDW	2×2	3×3	2	1	768
CVGlobalPool	1×1	2×2			768
CVFC					10

downsampling. This LWDW block consists of a lightweight 3 × 3 convolution with a stride size of 2, followed by the CV-GlobalPool layer to realize the complex-valued global pooling. Finally, the downsampled feature map is transformed into a one-dimension vector and input to the complex-valued fully connected (CVFC) layer to realize the classification of targets.

A. CMish Activation Function

Different activation functions have different application scopes according to their characteristics. For instance, the sigmoid and tanh functions are suitable for probability value processing since their outputs are limited to $(0, 1)$ and $(-1, 1)$, respectively. Since ReLU has no maximum limit and may have a large value, it is suitable for deep network training. In this case, the sigmoid and tanh functions are less superior due to the possible gradient disappearance. On the contrast, with a derivative of constant 1, the gradient of ReLU will remain in the chain reaction. However, the ReLU function is easy to cause the death of neurons during training. For $z \in \mathbb{C}$, the CReLU [29] can be obtained by extending real-valued ReLU function to complex-valued field, as shown in the following formula:

$$\mathrm{CReLU} = \mathrm{ReLU}\left(\Re(z)\right) + i \cdot \mathrm{ReLU}\left(\Im(z)\right). \tag{11}$$

It is noted that the Cauchy–Riemann differential equations do not hold for the CReLU in the second and fourth quadrants. For the complex-valued image, when the real part is greater than 0 and the imaginary part is smaller than 0, the CReLU is nondifferentiable. In this situation, some weight parameters cannot be gradient updated in the back-propagation process, resulting in the massive loss of information. To solve this problem, the CMish activation function is adopted, which is obtained by extending the Mish activation function from the real number domain to the complex one. The Mish activation function image is shown in Fig. 3. Compared with the ReLU function which is not smooth at point 0, the Mish is smooth everywhere which facilitates the convergence of network.

For $x \in \mathbb{R}$, the Mish function [30] can be denoted as

$$\mathrm{Mish}(x) = x \cdot \tanh\left(\ln\left(1 + e^{x}\right)\right). \tag{12}$$

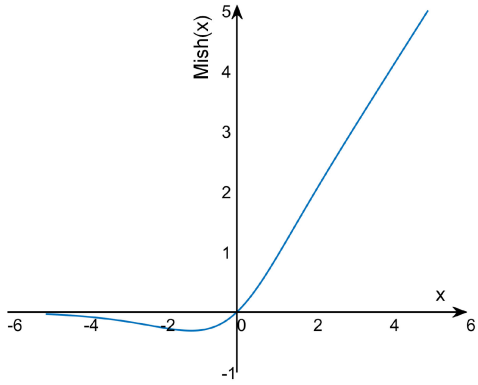


Fig. 3. Mish activation function image.

Similar to the CReLU activation function defined in [16], the proposed CMish actually applies the Mish function to the real and imaginary parts, respectively. For $z = x + i \cdot y \in \mathbb{C}$, the CMish function is given by

$$\begin{aligned}\mathrm{CMish}(z) &= \mathrm{Mish}\left(\Re(z)\right) + i \cdot \mathrm{Mish}\left(\Im(z)\right) \\ &= x\tanh\left(\ln\left(1 + e^{x}\right)\right) + i \cdot y\tanh\left(\ln\left(1 + e^{y}\right)\right).\end{aligned} \tag{13}$$

Satisfying the Cauchy–Riemann equations, the CMish can be differentiable in all quadrants to improve the expression ability of networks. Without upper limit, there will be no saturated area for CMish, avoiding the problem of gradient disappearance in the training process. Meanwhile, with a lower limit, the standardization effect of CMish can be guaranteed, which is a good characteristic for the network. Besides, the CMish is nonmonotonic. To be specific, the amplitude of CMish can be enlarged for small input and reduced for large input, which can accelerate the training speed of the network.

Another complex-valued version of Mish function denoted as cvMish was provided in [38] as follows:

$$\begin{aligned}\mathrm{cvMish}(z) &= z \cdot \tanh\left(\ln\left(1 + e^{z}\right)\right) \\ &= x\tanh\left(\ln\left(1 + e^{x}\right)\right) - y\tanh\left(\ln\left(1 + e^{y}\right)\right) \\ &\quad + i \cdot \left(x\tanh\left(\ln\left(1 + e^{y}\right)\right)\right. \\ &\quad \left. + y\tanh\left(\ln\left(1 + e^{x}\right)\right)\right).\end{aligned} \tag{14}$$

However, the above expression was deduced based on the following formula defined in [38].

$$\begin{aligned}\tanh\left(\ln\left(1 + e^{x+i\cdot y}\right)\right) &= \tanh\left(\ln\left(1 + e^{x}\right)\right) \\ &\quad + i \cdot \tanh\left(\ln\left(1 + e^{y}\right)\right)\end{aligned} \tag{15}$$

which cannot be derivated with rigid mathematical formulas.

B. CVReLBlock

In the process of object recognition, large convolution kernels can greatly increase the receptive field of object detection, while small ones increase the depth to achieve the same goal, yielding too deep a network to avoid gradient disappearance during the training process. The aforementioned phenomenon is especially

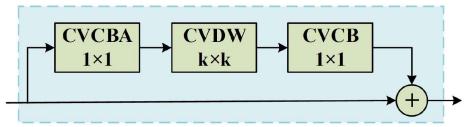


Fig. 4. Structure of the CVReLBlock.

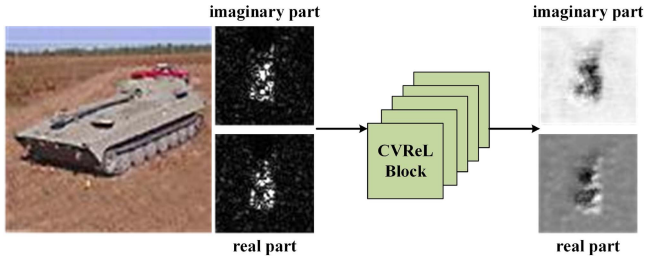


Fig. 5. Visualization of 7 × 7 CVReLBlock processing results.

true for the recognition network, whose performance will greatly degrade due to increased network depth.

With respect to the excellent classification, segmentation, and detection performance of RepLinkBlock [39] in the real-valued field, a complex-valued RepLinkBlock (denoted as CVReLBlock) is constructed, which is composed of three convolution operations with kernel sizes of 5 × 5, 7 × 7, and 13 × 13, respectively. Because the traditional large convolution leads to substantially increased parameters of the convolution, we use a point convolution plus a depth-separable large convolution to replace it, which can reduce the parameters in the network. Shown in Fig. 4, the CVReLBlock module is constructed with three parts, i.e., the CVCBA, CVDW, and CVCB. The CVCBA is constructed with a complex-valued BN layer, complex-valued convolution with kernel size 1 × 1, and CMish activation function. The CVDW stands for grouped large complex-valued convolution with kernel size $k \times k$, and CVCB consists of the complex-valued 1 × 1 convolution and complex-valued BN.

The results of sensitivity experiments show that the CVLWNet performs best when its first layer consists of CVReLBlock with $k = 7$. Taking the result of 2S1 as an example, it can be visually observed in Fig. 5 that the speckle noise and clutter in the complex-valued images can be filtered out after the CVReLBlock, making characteristics of the feature map more obvious. It is helpful for the backbone part to extract effective features and improve the classification accuracy of downstream tasks.

C. LC-CSPBlock

The lightweight networks [31], [32], [33], [34], [35], [40] have achieved excellent performance in the real-valued domain, not only with fewer weight parameters, but also with high accuracy. By extending the real-valued lightweight network structure to the complex-valued domain, the lightweight complex-valued cross stage partial block (denoted as LC-CSPBlock) is designed, whose structure is shown in Fig. 6. The Primary Conv represents 1 × 1 complex-valued convolution, which is mainly used for information fusion to adjust the number of channels in the output characteristic graph. The CVDP Conv denotes the complex-valued separable convolution with downsampling rate of 2 and the CVSELayer refers to the complex-valued squeeze-and-excitation channel attention mechanism. Besides, the output feature map size after Shortcut is the same as that from the upper branch.

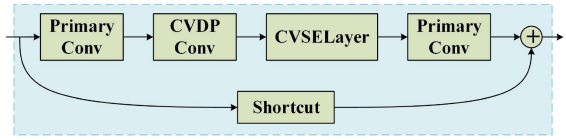


Fig. 6. Structure of the LC-CSPBlock.

TABLE II
MSTAR DATASET INFORMATION

Target	Serial No.	train17°	test15°	test30°	test45°
2S1	b01	299	274	288	288
BMP2	9563	233	195	-	-
	9566	232	196	-	-
	c21	233	196	-	-
BRDM2	c71	298	274	288	288
BTR60	k10yt7532	256	195	-	-
BTR70	c71	233	196	-	-
D7	-	299	274	-	-
T62	a51	299	273	-	-
T72	132	232	196	-	-
	812	231	195	-	-
	s7	228	191	-	-
ZIL131	e21299	299	274	-	-
ZSU234	d08	299	274	288	303

IV. EXPERIMENTAL RESULTS AND ANALYSIS

A. Dataset Description

The MSTAR dataset [8], [41] was collected using a 10 GHz center frequency X-band SAR sensor, achieving a minimum resolution of 0.3 m in both range and azimuth direction. This dataset contains ten different categories of ground targets, consisting of various military vehicles with different attributes, such as armored personnel carriers, tanks, bazookas, antiaircraft units, trucks, and bulldozers. The visual comparison between the optical images and complex-valued SAR images is provided in Fig. 7. The information of targets in the MSTAR dataset [22], including the category, quantity, azimuth, and serial number, is provided in Table II.

According to previous studies on the MSTAR dataset [22], [46], [47], [48], [49], we divide the dataset into two different types SOC and EOC, which denote the standard and extended operating conditions, respectively. To be specific, the SOC means that the training and testing are conducted with targets of the same serial number at different angles. On the contrast, the EOC refers to training and testing with targets of varying serial numbers or at significantly different angles. In order to

Fig. 7. Ten classes of military objectives. (Top) Optical image. (Bottom left) Real part. (Bottom right) Imaginary part.

TABLE III
DETAILED CONFIGURATION OF THE ADOPTED DATASET

Datasets	Training targets	Testing targets	Description
SOC-10	All categories, if there is more than one series, take the first series.		Training at 17° and testing at 15° under this category of this series.
SOC-3	All series in the categories BMP2, BTR70, T72.		Training at 17° and testing at 15° under this category of this series.
EOC-VV-4	First series of BMP2, BTR60, T62, T72.	The first series of BTR60, T62, the last two series of BMP2, T72.	Training at 17°, however tested at 15° on the other two variants of the same series.

better test the proposed network, referring to the experimental setting in [22], we use three complex-valued datasets, namely SOC-10, SOC-3, and EOC-VV-4, whose detailed configurations are shown in Table III.

1) SOC-10 test: SOC-10 contains all categories of the MSTAR dataset, of which each category contains only one series. In the experiment, training is performed with 2747 samples at an indentation angle of 17° while testing with 2425 samples at 15°. The SOC-10 dataset can be tested in a comprehensive manner comparative model recognition performance across all categories in the MSTAR dataset.

2) SOC-3 test: Aiming to validate the ability of the designed CVLWNet model to recognize a specific category, the SOC-3 dataset is consisted of all series of three target types, namely BMP2, BTR70, and T72. To be specific, shown in Table II, both T72 and BMP2 are composed of three series. On the contrast, the BTR70 consists of only one series. In total, training is conducted with 1622 samples at 17° and testing with 1365 target chips at 15°, respectively.

3) EOC-VV-4 test: Typically, there are more than one series variation of the same target type, which present slight difference in terms of the appearance and radar reflection characteristic. Therefore, it is necessary to demonstrate the performance of recognition methods under the condition of version variation of targets. The EOC-VV-4 dataset consists of four categories of targets in the MSTAR dataset, among which both BMP2 and T72 have three serial number variants. The proposed CVLWNet model is trained on the target collections, where BMP2 is with the serial number of 9563 and T72 with 132, while the testing process is done with changed series of BMP2 and T72 targets.

B. Training Process

PyTorch version 1.9 is used as the deep learning framework for our experiments, which are conducted on a Windows 10 personal computer with an Intel Core i9-10900 K CPU of 3.7 GHz and an Nvidia GeForce GTX 3090. In the training process, the cross-entropy loss function and the stochastic gradient descent

TABLE IV
ABLATION EXPERIMENTAL ANALYSIS OF THE PROPOSED CVLWNET WITH SOC-10 DATASET

%	Accuracy	Precision	Recall
CVLWNet	**98.23**	**98.39**	**98.39**
noCMish	97.73	97.76	97.75
noCVReLBlock	97.28	97.54	97.44
noLC-CSPBlock	93.86	94.67	94.42

TABLE V
ABLATION EXPERIMENTAL ANALYSIS OF THE PROPOSED CVLWNET WITH EOC-VV-4 DATASET

%	Accuracy	Precision	Recall
CVLWNet	**95.02**	**94.75**	**95.70**
noCMish	94.53	94.61	94.59
noCVReLBlock	94.24	93.85	94.74
noLC-CSPBlock	91.76	91.90	91.94

(SGD) [50] are used for the error evaluation of backward propagation and optimization of loss function, respectively. The batch size is set as 32. The initial learning rate is set to 0.01, and the Cosine annealing algorithm is adopted to update the learning rate. The training period is set as 300 epochs, and the network convergence is determined by the loss function decreasing to a certain range and tending to be constant.

C. Performance Analysis Metrics

In the experiment, we adopt accuracy, precision, and recall as evaluation indicators, whose definitions can be written as

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}}, \tag{16}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{17}$$

and

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{18}$$

where TN and TP represent the number of correct negatives and positives in the testing samples, respectively; FN and FP denote the number of false negatives and positives in the testing samples, respectively.

D. Ablation Experiments

In order to investigate the sensitivity of target recognition capability of the proposed CVLWNet, a series of ablation experiments are carried out by changing the main modules of the model, including CMish, CVReLBlock, and LC-CSPBlock. Specifically, three experiments named noCMish, noCVReLBlock, and noLC-CSPBlock are conducted, among which noCMish and noLC-CSPBlock indicate that CVLWNet is formed with the lack of CMish activation function and LC-CSPBlock modules, respectively. Besides, noCVReLBlock represents that CVLWNet is composed of CVReLBlock modules that lack the first layer of convolution with a kernel size of 7×7.

Ablation experiments are conducted on both SOC-10 and EOC-VV-4 datasets and the corresponding results are analyzed with three metrics, i.e., accuracy, precision, and recall, shown in Tables IV and V. It can be seen that all three metrics of the complete CVLWNet are higher than other configurations, followed by noMish, noCVReLBlock, and noLC-CSPBlock successively. In detail, for the SOC-10 dataset, the accuracy rates for the noMish, noCVReLBlock, and noLC-CSPBlock are 97.73%, 97.28%, and 93.86%, respectively. According to the analysis, it can be seen that the importance of CMish is similar to that of CVReLBlock, while the lack of LC-CSPBlock modules pose more significant effect on the recognition performance than the other two modules. The same conclusion can be drawn for the ablation experiment performed with the EOC-VV-4 dataset. In Table V, the accuracy rates of noCMish and noCVReLBlock degrade with about 0.49% and 0.72%, while that of noLC-CSPBlock decreases with an approximate value of 3.26%.

TABLE VI
NUMBER OF PARAMETERS AND ACCURACY ON THE SOC-10 DATASET

Models	Model type	#Params (MB)	FLOPs (G)	Accuracy (%)
RepLKNet-31B	real-valued	78.8	0.75	96.98
ResNet18	real-valued	42.7	**0.40**	96.54
DenseNet121	real-valued	27.1	0.58	96.37
AlexNet	real-valued	98.2	0.53	93.61
VGGNet	real-valued	131.5	0.63	95.75
A-ConvNet	real-valued	68.1	0.68	97.86
CV-CNN	complex-valued	73.0	0.42	92.70
CVLWNet	complex-valued	**18.7**	0.67	**98.23**

E. Comparative Experiments

In this section, we refer to six high-performing real-valued CNNs and one high-performing purely complex-valued CNN as comparison experiments, including RepLKNet-31B [39], ResNet18 [3], DenseNet121 [51], AlexNet [1], VGGNet [2], A-ConvNet [8], and CV-CNN [18], for experimental comparison with the proposed CVLWNet. Since the RepLinkBlock of RepLKNet in [39] has been referred in the proposed CVLWNet, the comparative results of RepLKNet are provided. According to the experimental results in [39], the RepLKNet-31B with as large a kernel as 31×31 achieves the best performance against the ImageNet dataset. Therefore, the RepLKNet-31B is applied for comparison in the following experiments. To be specific, the comparison experiments are carried out on three datasets that reflect the model's ability to all classes, some specific targets, and version changes.

1) Comparative Experiments on SOC-10 Dataset: To provide direct visual effect of the precision rate, recall rate, and accuracy rate for all classes, the confusion matrix is adopted. Shown in Fig. 8, the detailed confusion matrix obtained by testing the CVLWNet with the SOC-10 dataset is provided. It is clear from the confusion matrix that the CVLWNet achieves 100% precision and 100% recall for the BMP2. Besides, it obtains 100% precision for 2S1, D7, and T72 while a 100% recall can be achieved for three targets, including BTR60, BTR70, and ZSU234.

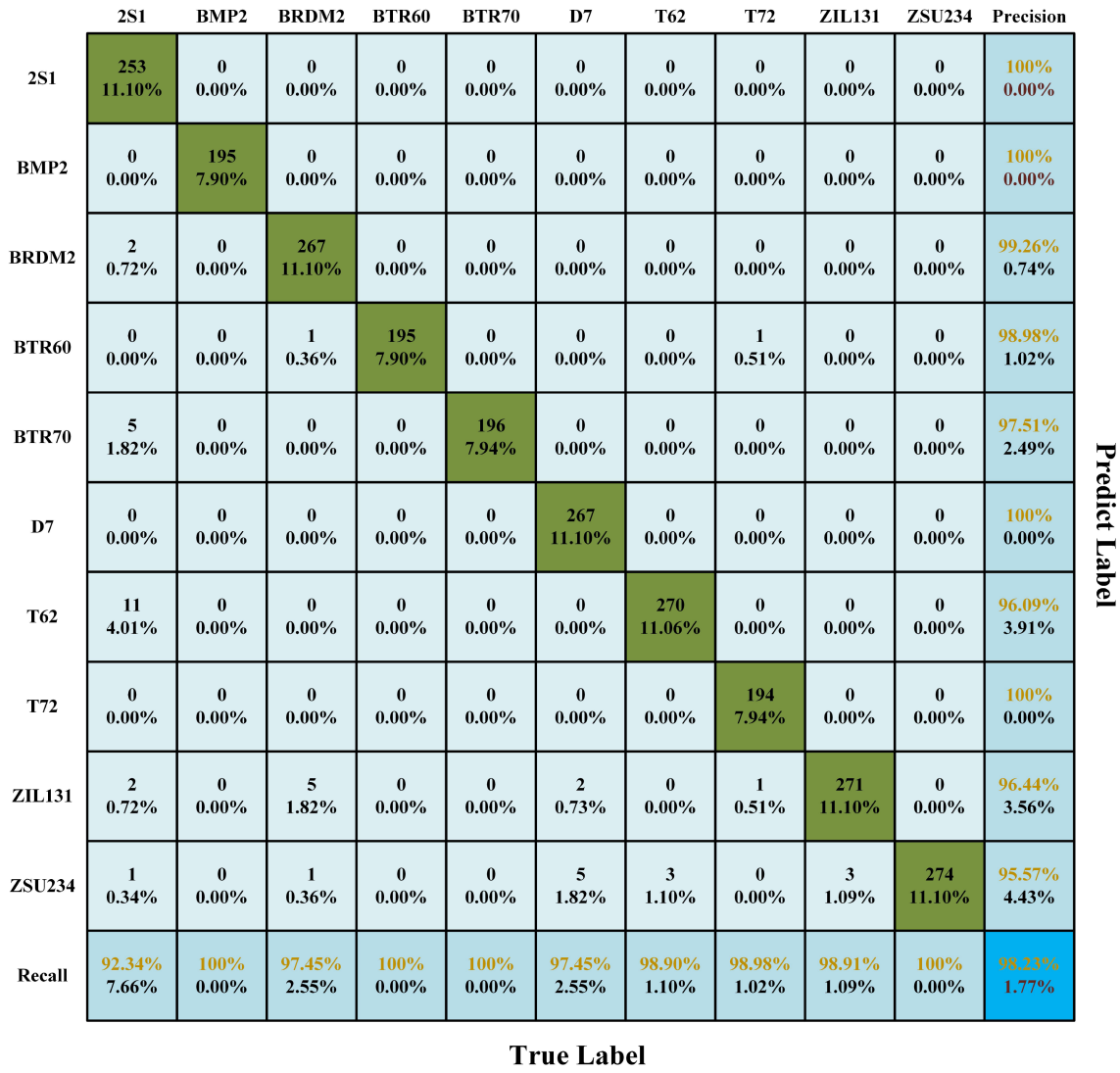


Fig. 8. SOC-10 classification confusion matrix.

Table VI shows quantitative comparisons of the eight networks tested on the SOC-10 dataset, including the weight parameters #Params, the floating-point operations (FLOPs), and the Accuracy. With respect to the amount of weight parameters, the proposed CVLWNet ranks top with only 18.7 MB, followed by DenseNet121 with 27.1 MB, ResNet18 with 42.7 MB, and A-ConvNet with 68.1 MB. On the contrast, the VGGNet obtains the greatest weight parameters with 131.5 MB, followed by AlexNet with 98.2 MB, and RepLKNet-31B with 78.8 MB. It is shown that the weight parameters of RepLKNet-31B are acceptable though a large convolutional kernel size 31×31 is adopted. In terms of FLOPs, the ResNet18 achieves the best performance with 0.40 G, followed by CV-CNN with 0.42 G, and AlexNet with 0.53 G. The proposed CVLWNet obtains 0.67 G FLOPs, which ranks medium among all comparative methods. By contrast, due to the large convolution kernel size, the FLOPs of the RepLKNet-31B is the greatest among all comparative methods. Shown in Table VI, the proposed CVLWNet achieves best performance with a recognition accuracy rate of 98.23%, marking in bold. Besides, the accuracy rates of other models are listed, namely CV-CNN-92.70%, AlexNet-93.61%, VGGNet-95.75%, DenseNet121-96.37%, ResNet18-96.54%, RepLKNet-31B-96.98%, and A-ConvNet-97.86%, ranking from low to high. It can be seen that the performance of DenseNet121 and ResNet18 are similar. However, DenseNet121 requires only half as many parameters to achieve the same accuracy as ResNet18. Compared to these proposed comparison models, our proposed CVLWNet network achieves excellent accuracy results among the SOC-10 dataset with the fewest weight parameters and the highest accuracy.

2) Comparative Experiments on SOC-3 Dataset: To validate the ability of the designed CVLWNet model to recognize some specific targets, the confusion matrix is adopted. Shown in Fig. 9, the detailed confusion matrix obtained by testing the CVLWNet with the SOC-3 dataset is provided. It can be seen that the CVLWNet obtains the best performance for BTR70 with 100%

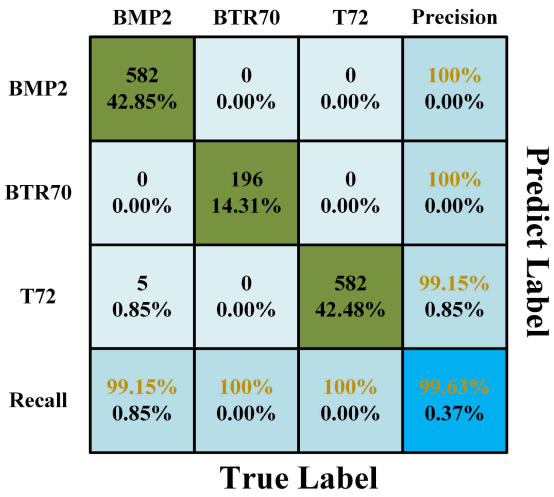


Fig. 9. SOC-3 classification confusion matrix.

TABLE VII
ACCURACY ON THE SOC-3 DATASET

Dataset	Models	Model Type	Accuracy(%)
SOC-3	RepLKNet-31B	real-valued	99.56
	ResNet18	real-valued	99.49
	DenseNet121	real-valued	99.34
	AlexNet	real-valued	99.05
	VGGNet	real-valued	99.27
	A-ConvNet	real-valued	99.56
	CV-CNN	complex-valued	98.24
	CVLWNet	complex-valued	**99.63**

precision and 100% recall. Besides, it achieves 100% precision for BMP2 and 100% recall for T72, respectively.

Grouped by their model type, Table VII presents the recognition accuracy of eight models testing with the SOC-3 dataset. It can be seen that the CVLWNet obtains a high recognition rate of 99.63%, which exceeds the other seven excellent SAR target recognition models. Besides, the RepLKNet-31B and A-ConvNet tie for the second place with a 99.56% recognition accuracy. The recognition accuracies of other five comparative networks are AlexNet-99.05%, CV-CNN-99.24%, VGGNet-99.27%, DenseNet121-99.34%, and ResNet18-99.49%, ranking from low to high. Therefore, compared to other models, excellent learning ability and representation ability for specific targets of the proposed CVLWNet model can be verified.

3) Comparative Experiments on EOC-VV-4 Dataset: The accuracy of each model tested on the EOC-VV-4 dataset is shown in Table VIII. In detail, the accuracy rates of seven comparative models are CV-CNN-83.31%, AlexNet-89.81%, A-ConvNet-89.25%, VGGNet-90.69%, ResNet18-91.33%, RepLKNet-31B-92.05%, and DenseNet121-92.38%, ranking from low to high. It can be seen that the CVLWNet obtains a high accuracy of 95.02%, outperforming the other seven comparative models with excellent recognition capability. By comparison, the proposed CVLWNet has excellent robustness until the situation of target version changes.

TABLE VIII
ACCURACY ON THE EOC-VV-4 DATASET

Dataset	Models	Model Type	Accuracy(%)
EOC-VV-4	RepLKNet-31B	real-valued	92.05
	ResNet18	real-valued	91.33
	DenseNet121	real-valued	92.38
	AlexNet	real-valued	89.81
	VGGNet	real-valued	90.69
	A-ConvNet	real-valued	89.25
	CV-CNN	complex-valued	83.31
	CVLWNet	complex-valued	**95.02**

V. CONCLUSION

To sufficiently investigate the abundant information contained in the complex-valued SAR images, the CVLWNet is proposed for target recognition based on the fully complex-valued CNN architecture, where the input, output, and weight parameters are all complex-valued. Based on the basic complex-valued building blocks, including the complex-valued convolution, complex-valued BN, and CMish activation function, two modules, i.e., the CVReLBlock and LC-CSPBlock, are well designed to enhance the representation ability of the proposed network. Therefore, the CVLWNet is constructed with the capability of extracting distinguishable features for the SAR target recognition task, which can be demonstrated by the comparative experimental results.

In the future, the research will be concentrated on the design of more lightweight and faster-reasoning complex-valued networks. In the inference stage, the reasoning speed of a large and integrated network is much higher than that of a small and scattered network [40]. Since RepVGG [52] has achieved excellent results in the real-valued domain, a possible solution will be the complex-valued structural reparameterization, including the equating of all convolution kernels into 3×3 convolution kernels in the complex-valued domain and the equivalent transformation of multibranch models. On the one hand, the cuda calculation density of 3×3 convolution is the highest, and the calculation efficiency of the underlying acceleration library for 3×3 convolution is much higher than convolution kernels of other sizes, such as $1 \times 1, 5 \times 5$, etc. On the other hand, merging the parameters of BN into the convolution operation to reduce the number of layers can reduce the cost of memory access in the complex-valued domain, thus fastening the calculation speed.

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.
[2] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Int. Conf. Learn. Representations*, 2015, pp. 1–14.
[3] 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.
[4] Z. Huang, Z. Pan, and B. Lei, "What, where, and how to transfer in SAR target recognition based on deep CNNs," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2324–2336, Apr. 2020.

[5] Y. Li, L. Du, and D. Wei, "Multiscale CNN based on component analysis for SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5211212.

[6] M. Huang, F. Liu, and X. Meng, "Few samples of SAR automatic target recognition based on enhanced-shape CNN," *J. Math.*, vol. 2021, pp. 9141023, Dec. 2021.

[7] P. Lang, X. Fu, C. Feng, J. Dong, R. Qin, and M. Martorella, "LW-CMDANet: A novel attention network for SAR automatic target recognition," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 6615–6630, 2022.

[8] S. Z. Chen, H. P. 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, Aug. 2016.

[9] F. Mohammadimanesh, B. Salehi, M. Mandianpari, E. Gill, and M. Molinier, "A new fully convolutional neural network for semantic segmentation of polarimetric SAR imagery in complex land cover ecosystem," *ISPRS J. Photogrammetry Remote Sens.*, vol. 151, pp. 223–236, May 2019.

[10] R. Chakraborty, Y. Xing, and S. X. Yu, "SurReal: Complex-valued learning as principled transformations on a scaling and rotation manifold," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 3, pp. 940–951, Mar. 2022.

[11] Q. Liu and L. Lang, "MMFF: Multi-manifold feature fusion based neural networks for target recognition in complex-valued SAR imagery," *ISPRS J. Photogrammetry Remote Sens.*, vol. 180, pp. 151–162, Oct. 2021.

[12] S. Feng, K. Ji, L. Zhang, X. Ma, and G. Kuang, "SAR target classification based on integration of ASC parts model and deep learning algorithm," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 10213–10225, 2021.

[13] S. Feng, K. Ji, F. Wang, L. Zhang, X. Ma, and G. Kuang, "Electromagnetic scattering feature (ESF) module embedded network based on ASC model for robust and interpretable SAR ATR," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5235415.

[14] S. Feng, K. Ji, F. Wang, L. Zhang, X. Ma, and G. Kuang, "PAN: Part attention network integrating electromagnetic characteristics for interpretable SAR vehicle target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5204617.

[15] J. Zhang, M. Xing, and Y. Xie, "FEC: A. feature fusion framework for SAR target recognition based on electromagnetic scattering features and deep CNN features," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 3, pp. 2174–2187, Mar. 2021.

[16] 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, 2022, Art. no. 4508705.

[17] J. Liu, M. Xing, H. Yu, and G. Sun, "EFTL: Complex convolutional networks with electromagnetic feature transfer learning for SAR target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5209811.

[18] Z. M. Zhang, H. P. 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.

[19] Y. C. Cao, Y. Wu, P. Zhang, W. K. Liang, and M. Li, "Pixel-wise PolSAR image classification via a novel complex-valued deep fully convolutional network," *Remote Sens.*, vol. 11, no. 22, 2019.

[20] X. F. Tan, M. Li, P. Zhang, Y. Wu, and W. Y. Song, "Complex-valued 3-D convolutional neural network for PolSAR image classification," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1022–1026, Jun. 2020.

[21] C. Lee, H. Hasegawa, and S. Gao, "Complex-valued neural networks: A comprehensive survey," *IEEE/CAA J. Automatica Sinica*, vol. 9, no. 8, pp. 1406–1426, Aug. 2022.

[22] 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, 2022, Art. no. 5228618.

[23] R. Savitha, S. Suresh, N. Sundararajan, and P. Saratchandran, "A new learning algorithm with logarithmic performance index for complex-valued neural networks," *Neurocomputing*, vol. 72, no. 16, pp. 3771–3781, Oct. 2009.

[24] A. Hirose, "Proposal of fully complex-valued neural networks," in *Proc. IJCNN Int. Joint Conf. Neural Netw.*, 1992, pp. 152–157.

[25] 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.

[26] 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.

[27] T. Scarnati and B. Lewis, "Complex-valued neural networks for synthetic aperture radar image classification," in *Proc. IEEE Radar Conf.*, 2021, pp. 1–6.

[28] H. Zhang, Y. Zhang, S. Zhu, and D. Xu, "Deterministic convergence of complex mini-batch gradient learning algorithm for fully complex-valued neural networks," *Neurocomputing*, vol. 407, pp. 185–193, Sep. 2020.

[29] C. Trabelsiet al., "Deep complex networks," 2017, *arXiv:1705.09792*.

[30] D. Misra, "Mish: A self regularized non-monotonic activation function," 2019, *arXiv:1908.08681*.

[31] K. Han, Y. Wang, Q. Tian, J. Guo, C. Xu, and C. Xu, "GhostNet: More features from cheap operations," 2019, *arXiv:1911.11907*.

[32] A. Howardet al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*.

[33] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," 2018, *arXiv:1801.04381*.

[34] A. Howardet al., "Searching for MobileNetV3," 2019, *arXiv: 1905.02244*.

[35] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," 2017, *arXiv:1707.01083*.

[36] A. Kreutz and K. Delgado, "The complex gredient operator and CR-Calculus," 2009, *arXiv:0906.4835*.

[37] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," 2015, *arXiv:1502.03167*.

[38] Z. Chen, Y. Huang, Y. Hu, and Z. Chen, "Phase recovery with deep complex-domain priors," *IEEE Signal Process. Lett.*, vol. 29, pp. 887–891, 2022.

[39] X. Ding, X. Zhang, Y. Zhou, J. Han, G. Ding, and J. Sun, "Scaling up your kernels to 31: Revisiting large kernel design in CNNs," 2022, *arXiv:2203.06717*.

[40] N. Ma, X. Zhang, H. T. Zheng, and J. Sun, "ShuffleNet V2: Practical guidelines for efficient CNN architecture design," 2018, *arXiv:1807.11164*.

[41] Z. Zeng, J. Sun, C. Xu, and H. Wang, "Unknown SAR target identification method based on feature extraction network and KLD-RPA joint discrimination," *Remote Sens.*, vol. 13, no. 15, 2021, Art. no. 2901.

[42] S. T. Seydi, M. Hasanlou, and M. Amani, "A new end-to-end multi-dimensional CNN framework for land cover/land use change detection in multi-source remote sensing datasets," *Remote Sens.*, vol. 12, no. 12, 2020, Art. no. 2010.

[43] R. H. Shang, J. H. He, J. M. Wang, K. M. Xu, L. C. Jiao, and R. Stolkin, "Dense connection and depthwise separable convolution based CNN for polarimetric SAR image classification," *Knowl. Based Syst.*, vol. 194, 2020.

[44] J. P. Zhao, M. Datcu, Z. H. Zhang, H. L. Xiong, and W. X. Yu, "Contrastive-regulated CNN in the complex domain: A method to learn physical scattering signatures from flexible PolSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 12, pp. 10116–10135, Dec. 2019.

[45] J. Q. Ai, M. Huang, F. F. Wang, X. M. Yang, and Y. L. Wu, "Completed local binary patterns feature integrated convolutional neural network-based terrain classification algorithm in polarimetric synthetic aperture radar images," *J. Appl. Remote Sens.*, vol. 16, no. 1, Jan. 2022.

[46] O. Kechagias-Stamatis and N. Aouf, "Automatic target recognition on synthetic aperture radar imagery: A survey," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 36, no. 3, pp. 56–81, Mar. 2021.

[47] L. M. Novak, G. J. Owirka, and A. L. Weaver, "Automatic target recognition using enhanced resolution SAR data," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 35, no. 1, pp. 157–175, Jan. 1999.

[48] B. Ding, G. Wen, C. Ma, and X. Yang, "An efficient and robust framework for SAR target recognition by hierarchically fusing global and local features," *IEEE Trans. Image Process.*, vol. 27, no. 12, pp. 5983–5995, Dec. 2018.

[49] Y. Guo, L. Du, D. Wei, and C. Li, "Robust SAR automatic target recognition via adversarial learning," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 716–729, 2021.

[50] S. Ruder, "An overview of gradient descent optimization algorithms," 2016, *arXiv:1609.04747*.

[51] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 4700–4708.

[52] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding, and J. Sun, "RepVGG: Making VGG-style ConvNets great again," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 13728–13737.

Yuejun Zhu was born in 1998. He received the B.S. degree in electronic engineering from the Inner Mongolia University of Science and Technology, Baotou, China, in 2021. He is currently working toward the M.S. degree in control science and engineering with the School of Automation, Hangzhou Dianzi University, Hangzhou, China.

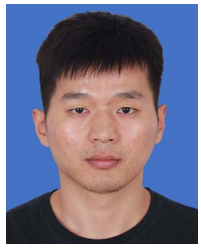
His research interests include deep learning and SAR target detection and recognition.

Tao Li was born in 1992. She received the B.S. degree in electrical engineering and the Ph. D. degree in pattern recognition and intelligent system from Xidian University, Xi'an, China, in 2013 and 2019, respectively.

She is currently a Lecturer with the School of Automation, Hangzhou Dianzi University, Hangzhou, China, and a Postdoctoral Researcher with the Nanjing Research Institute of Electronics Technology, Nanjing, China. Her research interests include SAR image processing, target detection, and recognition.

Dongliang Peng (Member, IEEE) was born in 1977. He received the B.S. and M.S. degrees in flight vehicle design and engineering from the Harbin Institute of Technology, Harbin, China, in 1998 and 2000, respectively, and the Ph.D. degree in control science and engineering from Zhejiang University, Hangzhou, China, in 2013.

He is currently a Professor with the School of Automation, Hangzhou Dianzi University, Hangzhou, China. His research interests include information fusion and estimated theory.



Haoran Wang was born in 1992. He received the B.S degree and the Ph.D degree in electrical engineering from Xi'an Jiaotong University, Xi'an, China, in 2013 and 2019, respectively.

He is currently a Postdoctoral Researcher with the University of Science and Technology of China, Hefei, China, and working as a Director in Sungrow Power Supply Company, Ltd, Hefei, China. His research interests include deep learning, power electronics, and battery safety.

Sainan Shi was born in 1990. She received the B.S. degree in electrical engineering and the Ph. D. degree in information and signal processing from Xidian University, Xi'an, China, in 2013 and 2018, respectively.

She is currently an Associate Professor with the School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjing, China. Her research interests include radar signal processing and weak target detection in sea clutter.