EMC$^2$A-Net: An Efficient Multibranch Cross-Channel Attention Network for SAR Target Classification

Xiang Yu, Ying Qian, Zhe Geng, Member, IEEE, Xiaohua Huang, Senior Member, IEEE, Qinglu Wang, and Daiyin Zhu, Member, IEEE

Abstract—In recent years, convolutional neural networks (CNNs) have demonstrated significant potential for synthetic aperture radar (SAR) target recognition. SAR images possess a strong sense of granularity and contain texture features of varying scales, including speckle noise, dominant scatterers, and target contours, which are not typically considered in traditional CNN models. This article proposes two residual blocks, termed multibranch cross-channel attention (EMC$^2$A) blocks, with multiscale receptive fields (RFs) based on a multibranch structure and designs an efficient isotopic architecture deep CNN (DCNN) called EMC$^2$A-Net, whose structure is interpretable from a probability and mathematical statistics perspective. EMC$^2$A blocks employ parallel dilated convolution with different dilation rates to effectively capture multiscale contextual features without significantly increasing the computational load. To further enhance the efficiency of multiscale feature fusion, this article presented a multiscale feature cross-channel attention module, known as the EMC$^2$A module, which adopts a local multiscale feature interaction strategy without dimensionality reduction. This strategy adaptively adjusts the weights of each channel using efficient one-dimensional (1-D)-circular convolution and sigmoid function to guide attention at the global channel-wise level. Comparative results on the moving and stationary target acquisition and recognition (MSTAR) dataset demonstrate that EMC$^2$A-Net outperforms the other available models of the same type and possesses a relatively lightweight network structure. The ablation experimental results further demonstrate that the EMC$^2$A module significantly enhances the model's performance by utilizing only a few parameters and appropriate cross-channel interactions.

Index Terms—Channel attention, deep convolutional neural network (DCNN), isotropic architecture, multiscale feature fusion, synthetic aperture radar (SAR), target classification.

Manuscript received 1 August 2022; revised 20 November 2022 and 29 March 2023; accepted 4 June 2023. Date of current version 28 June 2023. This work was supported in part by the Natural Science Foundation of Jiangsu Province under Grant BK20200420, in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2020B1515120060, in part by the National Natural Science Foundation of China Grant 62076122 and Grant 62271252, and in part by the Nanjing Institute of Technology (NJIT) through the Talent Startup Project under Grant YKJ201982.

Xiang Yu, Ying Qian, Xiaohua Huang, and Qinglu Wang are with the School of Computer Engineering, Nanjing Institute of Technology, Nanjing 211106, China (e-mail: 419781300@qq.com).

Zhe Geng and Daiyin Zhu are with the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China. Digital Object Identifier 10.1109/TGRS.2023.3285037

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) imaging technology can obtain high-resolution images of geographical objects under all-day and all-weather conditions. Benefitting from the flexibility of its platform and efficient remote sensing information acquisition capability, SAR imaging technology has been widely employed in military and civil fields.

In contrast to optical imagery, SAR imagery is characterized by a pronounced sense of granularity. In SAR target imagery, smaller textural features correspond to local details of the target and speckle noise, while larger textural features are typically present in the contours of the target. The aforementioned characteristics of SAR imagery lead to low efficiency in the manual interpretation of high-resolution SAR imagery. Consequently, SAR automatic target recognition (SAR-ATR), as proposed by the MIT Lincoln Laboratory, has garnered significant attention.

Detection, discrimination, and classification are the three stages involved in SAR-ATR. Among these stages, SAR target classification holds significant strategic importance in the military field and has attracted considerable interest. In classical SAR target classification methods, feature extraction [1], [4] and classifier design [5], [8] are two distinct aspects. Commonly used feature extraction methods include mathematical transformation features [1], [2], computer vision features [3], [9], [10], and electromagnetic features [4]. Presently, mainstream classifiers used in SAR-ATR include template matching [5], support vector machines [6], [11], boosting [7], sparse representation [8], and more. Notably, conventional machine learning approaches are unable to effectively and automatically extract the features that accurately represent SAR targets. To achieve this aim, some researchers have introduced the deep convolutional neural network (DCNN) method [12], [14], which can automatically extract the features of target images and enable end-to-end SAR target classification.

As a branch of deep learning, the application of DCNN in target classification and detection has garnered considerable attention and has yielded remarkable results. Nonetheless, two major problems still exist in applying the DCNN model for SAR target classification. First, the intraclass difference of image features is significant, and the number of samples...
is limited. The SAR target image is highly sensitive to the azimuth and depression angles of imaging. Even slight changes in these two angles can lead to significant variations in the target backscattering coefficient, and the same targets in SAR images can exhibit different features. In addition, compared to the optical image dataset, labeled samples of SAR target images are scarce. With a large number of trainable parameters in the network, the generalization ability of the DCNN model is greatly tested under a small-scale SAR target training dataset. Second, the presence of speckle noise interference. Due to the small backscattering coefficient of natural scenery and isotropic and uniform scattering, the background of SAR images has strong speckle noise [15]. The DCNN model is required to learn the local details of SAR targets in speckle noise, in which both have small textural features, thus posing a challenge for the design of the DCNN model.

The notion of the local receptive field (RF) in CNNs originated from research on neurons in the primary visual cortex of cats in the last century [16]. DCNNs employ convolution kernels with different RF sizes to gather multiscale spatial information in the same processing stage. This mechanism has been widely adopted in recent DCNNs. A typical implementation is the multibranch structure, which employs various RF convolution kernels in the same layer [17], [19]. Several relevant studies [15], [20] have indicated that convolution kernels with different RF sizes can enhance the accuracy of SAR target recognition, particularly under the influence of speckle noise.

However, several experimental studies have suggested that the RF size of neurons in the visual cortex is not fixed but is modulated by the stimulus [21]. Inspired by this evidence, some nonlinear approaches (e.g., channel attention modules) [19], [20] were proposed to aggregate information from adaptively selected features with different scales. However, these channel attention modules have some weaknesses. In most current approaches, channel weight activation is either within channels with the same RF [22], [24] or between local channels with different RFs [19]. Cross-channel mutual information, which is the mutual information between channels with the same and different RFs, has not received much attention in the process of designing channel attention modules. It is believed that the channel weight could reflect the competitiveness of channel features in helping the model achieve better performance. Therefore, the interaction of all channels in the same layer should be reflected not only in the feature fusion steps but also in the channel weight activation step.

Inspired by the studies mentioned above, this article presented a pair of novel and efficient multibranch cross-channel attention (EMC 2-A) blocks and further proposed a lightweight EMC 2-A-Net. The primary contributions of this study are as follows.

1) Based on the multibranch structure, this article presented a novel residual convolution framework comprising the EMC 2-A Type-A block and Type-B block that leverages multiple RFs. The proposed EMC 2-A blocks employ parallel dilated convolution with varying dilation rates to effectively capture multiscale texture features while maintaining computational efficiency. The residual structure integrated into the EMC 2-A Type-A and Type-B blocks mitigates the issues of gradient vanishing and exploding in training caused by block-deep stacking. Leveraging the pair of EMC 2-A blocks, an efficient isotropic architecture DCNN, called EMC 2-A-Net, is proposed. It is noteworthy that the parameterization of the network structure is achieved from the fundamental blocks (the pair of EMC 2-A blocks) and extends to the main part of EMC 2-A-Net. Through the utilization of structural quantification and a network design methodology inspired by prior research [25], the EMC 2-A-Net structure has interpretability within the framework of probability and mathematical statistics.

2) The EMC 2-A module is proposed as a fully convolutional multiscale feature cross-channel attention module. To mitigate the impact of dimensionality reduction on channel attention prediction [26], a direct association between the channel and its weight is established within the EMC 2-A module. Furthermore, leveraging the multibranch cross-channel interaction structure, the adaptively learned channel weights reflect the global contribution of feature maps across both the scale and type dimensions. These properties make the EMC 2-A module efficient and lightweight.

Contrast and ablation experiments were conducted using the moving and stationary target acquisition and recognition (MSTAR) SAR target dataset. The contrast experimental results, conducted without any data augmentation, demonstrate that EMC 2-A-Net outperforms previous state-of-the-art (SOTA) models (see [19], [27]) while maintaining a relatively small total number of trainable parameters and floating-point operations (FLOPs). In addition, the author further evaluated the effectiveness of the EMC 2-A module through ablation experiments conducted on the EMC 2-A-Net architecture.

The rest of this article is organized as follows. Section II gives an overview of the related work. Section III introduces the structure and components of EMC 2-A-Net. Section IV discusses the dataset, implementations, and evaluation of the proposed method. Finally, conclusions are presented in Section V.

II. RELATED WORKS

A. Multibranch Convolutional Networks

Convolutional network structures that incorporate multiple information transmission paths are commonly referred to as multibranch convolutional networks. These structures can be categorized into two distinct classes. The first class consists of structures that incorporate a main path alongside a bypass structure. The second class comprises structures that feature parallel paths that are characterized by a multiscale RF structure.

The main objective of the first category of multibranch convolutional networks is to address the problem of gradient vanishing and exploding during DCNN training. This was first introduced by highway networks [28] and then further developed by ResNet [29], [30], with the bypassing path utilizing pure identity mapping. In addition, DenseNet [31] built upon this idea by incorporating level-wise multiple bypasses that enhance the integration of shallow features with high resolution and high-level features with rich semantic information.
The second category focuses on enhancing the multiscale feature representations of layers in a neural network. For instance, InceptionNets [17], [34] have been designed meticulously, with each branch featuring different RF convolution kernels. This design improves the width of the network and reduces the risk of overfitting. Although such a structure with multi-RF kernels can learn multiscale texture features, parallel two-dimensional (2-D) convolution operations with large kernels may consume excessive computing resources during training and testing. Dilated convolution, fortunately, can significantly reduce computational costs and increase the RF of the convolution kernel simultaneously [35]. More recently, dilated inception network (DINet) was proposed for visual saliency prediction [18]. This network employs parallel dilated convolutions with varying dilation rates within the same layer, making multiscale information extraction and fusion more efficient. Our proposed EMC²-A-Net follows the idea of ResNet and DINet. Moreover, it efficiently integrates both identity mapping and multiscale feature fusion.

B. Attention Mechanisms

The attention mechanism has been recognized as a promising means for enhancing DCNNs and has been applied in various research fields, including semantic segmentation [36], [39], object detection [40], [41], and target recognition [20], [42], [43]. In the task of target recognition, the incorporation of an attention mechanism can enhance the expressiveness of relevant features, thereby facilitating an improvement in recognition accuracy. At the same time, the attention mechanism is capable of selectively suppressing less informative features.

Squeeze-and-excitation network (SENet) [22], as a milestone of a single-scale RF channel attention network, represents an efficient gating mechanism for self-recalibrating channel weights and has demonstrated promising performance. Bottleneck attention module (BAM) [23] and convolutional block attention module (CBAM) [24] entail self-contained adaptive attention modules in both the spatial and channel axes.

With an increased focus on the parallel multiscale RF structure, selective kernel network (SKNet) [19] followed the SENet approach and introduced an adaptive channel attention mechanism into the multiscale RF kernel structure for the first time. However, further studies of efficient channel attention (ECA)-Net reveal that channel-wise dimensionality reduction in the above networks may have side effects on weight prediction, indicating that using one-dimensional (1-D) convolution for appropriate cross-channel interaction is a more effective and efficient weight prediction method. As a result, our designed EMC²-A module adopts 1-D circular convolution to ensure a one-to-one mapping relationship between the channel and its weight. The design of the entire module, particularly in the weight activation step, embodies the multiscale RF cross-channel interaction, which was overlooked in SKNet.

C. Grouped/Shuffled/Dilated Convolution

In the standard convolutional layer, the numbers of input and output channels (feature maps) are denoted by $C$ and $N$, respectively. When the kernel size is $K \times K$, the total number of parameters of the kernel is $K \times K \times C \times N$; hence, the computational cost becomes expensive if $C$ is large. To reduce this cost, group convolution partitions the input channels into $G$ mutually exclusive groups, compressing the number of parameters to $K \times K \times (C/G) \times N$ [44].

However, such outputs from grouped convolution only relate to inputs within the same group, resulting in reduced information exchange among groups. To address this limitation, the channel shuffle operation can be utilized to preserve cross-group information and improve the performance of grouped convolution. By allowing grouped convolution to access input data from different groups, the shuffle operation can improve information exchange and enhance the overall effectiveness of the convolutional layer [45].

Dilated convolution was originally developed as a signal analysis algorithm for wavelet decomposition [46]. In the design of DCNNs, pooling and down-sampling layers are often employed to reduce image size and increase RFs. However, these methods may cause a loss of image structure and spatial hierarchical information. To mitigate this problem, dilated convolution was introduced as an alternative approach that increases the dilation rate to enlarge the RF of the convolution kernel without reducing the image size.

In DINet [18], a technique known as “atrous spatial pyramid pooling (ASPP)” was developed to capture multiscale features and aggregate context information using parallel dilated convolutions with varying dilation rates in the same layer. Similarly, in our proposed EMC²-A-Net, dilated and grouped convolutions are utilized to efficiently capture multiscale features. In addition, a channel shuffle operation with 1-D circular convolution is employed to fuse information across multiscale RF channels within the EMC²-A module, further improving the effectiveness of the module.

III. OUR METHOD

This article presented EMC²-A-Net, a SAR target classification network that employs a four-stage isotropic architecture and incorporates a novel full convolution multibranch cross-channel attention module. The network’s parallel multibranch structure is used to extract texture features of different scales from the same resolution feature map with fewer parameters, thus enabling efficient feature extraction. The lightweight network structure is effective in mitigating the overfitting problem associated with small-sample learning. Feature maps have different contributions to the recognition accuracy, and the proposed EMC²-A module aims to suppress the expression of features with less contribution (e.g., speckle features) while enhancing the expression of features with more contribution (e.g., target local features). This is achieved through the adaptive adjustment of channel weights based on feature maps of different scales and types. In brief, this paper’s contributions are twofold: first, the design of a channel weight self-recalibration module—EMC²-A module, and second, the proposal of an SAR target classification network—EMC²-A-Net. Next, the author will introduce the EMC²-A module and network in two parts.
A. EMC²A Module

1) Compression: The proposed EMC²A module utilizes multiscale features obtained from a parallel multibranch convolution layer with different RFs, as illustrated in Fig. 1. To simplify the description of the function of the module, only two groups of feature maps with two different feature scales, \( U \in \mathbb{R}^{C \times H \times W} \) and \( V \in \mathbb{R}^{C \times H \times W} \), are considered as two inputs to the module. Here, \( C, H, \) and \( W \) represent the number of channels and the vertical and horizontal sizes of the feature map, respectively. Since channel weights are determined by all feature maps, it is necessary to reduce the dimensions of feature maps while maintaining their independence. To achieve this, we employ global average pooling (GAP) to compress the global information of each feature map into channel-wise statistics. As depicted in Fig. 1, the channel-wise statistic vector \( u(1 \times C) \) is obtained by shrinking \( U \) through spatial dimensions \( H \times W \)

\[
u = \text{GAP}(U) = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} U(i, j).
\]

(1)

Similarly, the vector \( v \) can be obtained.

2) Shuffle and Fuse: The primary objective of the EMC²A module is to enable neurons to adaptively adjust their interests from multiple branches carrying different scales of features according to the stimulus content. To achieve this goal, we take two main steps: shuffle and fuse.

In the shuffle step, we gather all channel features and shuffle them. Specifically, we concatenate all GAP results, which are channel-wise statistics of multiscale feature maps, and apply batch normalization (BN) to smooth the landscape of the entire loss function [47]. Then, we shuffle all these channel-wise statistics, allowing channel information with multiscale features to be integrated into the next convolution layer.

For the problem of adaptive estimation of channel weights, a study in [26] indicates that avoiding dimensionality reduction of channel-wise statistical vectors is beneficial in learning effective channel attention, and it is inefficient and unnecessary to capture dependencies across all channels. Inspired by this idea, we proposed a channel-wise statistics fusion and competition method in the fusion step for the parallel multiscale RF structure. In this method, a learnable and lightweight 1-D-circular convolution is used. As shown in (2) and Fig. 1, the vector \( w \) is shuffled by channel-wise statistics \( u \) and \( v \), \( h \) is the learnable 1-D-circular convolution kernel, and \( R_N \) refers to the principal value interval with lengths \( N \) and \( N = 2 \times C \). Therefore, without padding, the dimension of the convolution result \( x \) is \( 2 \times C \), which is consistent with the dimension of the input data

\[
x(n) = w(n) \odot h(n) = \left[ \sum_{m=0}^{N} w(m)h((n-m)N) \right] R_N(n).
\]

(2)

Let \( k \) be the size of the convolution kernel, and we utilize \( R_{\text{ncks}} \) to represent the ratio of the number of channels to the kernel size of the 1-D-circular convolution, which has been verified and discussed in Section IV-E. By collaborating with the previous shuffle step, the convolution kernel can perceive the channel-wise statistic and its \( k - 1 \) neighbors with multiscale features during multiplication and addition operations. Thus, the convolution result represents an adaptive weighted average of channel-wise statistics with multiscale features and reflects the interaction of all channels.

3) Activation and Regrouping: In the EMC²A module, to improve the performance of the model, all channels should adaptively adjust their weights through mutual competition during the network training process. Significantly, this is a nonmutually exclusive competitive relationship since multiple channels should be allowed to be emphasized instead of enforcing a one-hot activation. To meet these criteria, we employed a gating mechanism with a sigmoid activation function to determine the weights of all channels, as shown in (3). Notably, this activation operation fully utilizes the mutual information of channels with multiscale RFs. Thus, the channel weight integrates both the type and scale information of the feature maps in the EMC²A module

\[
\text{Sigmoid}(x) = \frac{1}{1 + \exp(-x)}.
\]

(3)

To ensure a direct correspondence between the channel and its weight [26], we designed a regroup operation that reassigns the channel weight vector back to two groups. This operation is essentially the inverse operation of the shuffle step.

B. EMC²A-Net

EMC²A-Net adopts a four-stage isotopic architecture, which is mainly composed of two kinds of blocks, namely, EMC²A block-A and block-B. Next, the author introduces EMC²A-Net from these two blocks.
Fig. 2. Structures of the EMC2A blocks. The structural parameters are described as follows: \( r \) and \( w \) represent the resolution of the feature maps and number of channels, \( b \) indicates the bottleneck ratio, the number of branches and the RF of each branch are controlled by \( f_n \) and \( f_s \), \( g \) represents the number of groups in the group convolution, and \( s \) denotes the stride of convolution. (a) Type-A block. (b) Type-B block.

1) Blocks and Stages: The structures of the EMC2A Type-A block and EMC2A Type-B block are depicted in Fig. 2, where \( r \) and \( w \) represent the size of the feature maps and number of channels, respectively, and the subscript \( i \) denotes the \( i \)-th stage. Both blocks follow a structure similar to the classical ResBlock, with some notable differences. First, the \( 1 \times 1 \) convolution layer is designed not to alter the number of channels \( w_i \), which means that the bottleneck ratio is \( b_i = 1 \). This design allows for an increase in the depth of the network and strengthens its ability to express complex features. Second, relevant studies [15], [20] have demonstrated that multiscale features can improve the accuracy of SAR target recognition under speckle noise interference. In light of this, both blocks adopt multibranch and multiscale RF convolution structures based on the \( 3 \times 3 \) dilated convolution. The number of branches and the RF of each branch (dilation rate) are determined by \( f_n_i \) and \( f_s_i \), respectively. Another structural parameter \( g_i \) is introduced to specify the number of groups in the group convolution. Finally, the convolution output from all branches is fed into the EMC2A module to obtain channel weights, and the final fused feature maps are obtained through elementwise summation of weighted feature maps from each branch.

As shown in Fig. 2(a) and (b), the Type-A block and Type-B block have two differences. First, in the Type-A block, the numbers of input and output channels \( (w_{i-1} - 1) \) and \( w_i \) are different. As a result of utilizing a stride of \( 2(s = 2) \) and \( 3 \times 3 \) convolution in the main pathway of the Type-A block, the resolutions of the input and output feature maps (denoted by \( r_{i-1} \) and \( r_i \), respectively) are also different (where \( r_i = r_{i-1}/2 \)). In contrast, the Type-B block preserves the same input and output dimensions. Second, the skip connection utilized in both blocks plays a critical role in facilitating the transfer of gradients to the shallower layers during backpropagation. This mechanism effectively mitigates the issues of gradient vanishing and exploding to a certain extent. Specifically, Type-A and Type-B blocks employ a stride 2 \((s = 2)\) \( 1 \times 1 \) convolution and an identity mapping as the skip connection, respectively.

Based on the aforementioned two block types, the stage was constructed. As illustrated in Fig. 3, the stage’s depth is represented by the structural parameter \( d_i (d_i \geq 1) \). A single Type-A block is employed as the connection block between the two stages. Following the passage of the input feature map through the Type-A block, the resolution and channel number were altered from \( r_{i-1} \) and \( w_{i-1} \) to \( r_i \) and \( w_i \), respectively. The remaining section of the stage can be built by stacking Type-B Blocks, and the number of Type-B blocks equals \( d_i - 1 \). This implies that when \( d_i = 1 \), there is only one Type-A block but no Type-B block present in the stage, as shown in stage 4 of the forthcoming EMC2A-Net.

2) EMC2A-Net Structure: EMC2A-Net is a DCNN with a four-stage isotopic architecture that can be decomposed into three components: the stem, body, and head, as depicted in Fig. 4. The stem component comprises a \( 3 \times 3 \) convolutional
Fig. 3. Stage structure. The depth of each stage is determined by the structural parameter $d_i (d_i \geq 1)$. The stage structure consists of only one Type-A block, followed by Type-B blocks connected in series. The number of Type-B blocks in each stage is $d_i - 1$.

Fig. 4. EMC$^2$A-Net structure diagram. The input of the network is a single channel ($w = 1$) gray image, and the output is a vector consisting of $L$ elements, where $L$ represents the number of categories.

The middle component of EMC$^2$A-Net is the body, which comprises four stages, as illustrated in Fig. 3. The structural parameters of each stage are explicitly detailed in Table I, and when combined with the block structure (see Fig. 2) and stage structure (see Fig. 3), the body component of the network can be clearly elucidated. Notably, the values of the structural parameters have a substantial impact on the performance of EMC$^2$A-Net. To design the network, we leveraged the EMC$^2$A Type-A and Type-B blocks to define a 6-D design space spanned by the structural parameters specified in Table I. We followed and improved the approach proposed in [25] to optimize the design space incrementally, resulting in the discovery of the design rule for EMC$^2$A-Net. Table I presents the structural parameter values obtained based on this rule, which were constrained by the complexity of the model.

The final classification output of EMC$^2$A-Net is generated by the head component, which comprises a fully connected layer and a softmax activation function. The output is a $1 \times L$ dimensional vector, and each element of the vector corresponds to the probability $\phi_i$ of each category, as shown in the following equation:

$$
\phi_i = \text{Softmax}(z_i) = \frac{\exp(z_i)}{\sum_{l=1}^{L} \exp(z_l)}, \quad i = 1, 2, \ldots, L.
$$

The cost function of EMC$^2$A-Net adopts a cross-entropy function, which is defined in (5). Where $\Phi_i$ represents the $i$th element of a sample label in the form of a one-hot vector. The weights of the network are updated by backpropagation according to the result of the cost function during the training process [48]

$$
CE = - \sum_{i=1}^{L} \Phi_i \cdot \log \phi_i
$$

IV. EXPERIMENTS

The experimental dataset employed in this study has been derived from the MSTAR program [49], which was launched in the mid-1990s with financial support from the U.S. Defense Advanced Research Projects Agency (DARPA). The dataset is acquired through the employment of the synthetic aperture radar target location and recognition system (STARLOS) sensor, an X-band horizontal polarization SAR. The MSTAR program conducted a comprehensive collection of SAR images of diverse military vehicles affiliated with the former Soviet Union. Data, included target occlusion, camouflage, configuration changes, and other expansibility conditions, forming a more systematic and comprehensive measured database. Only a limited subset of this extensive dataset is accessible to the public via the website [50], and only the experimental dataset used in this study is available. For simplicity, we shall refer to this specific subset as the MSTAR dataset hereafter.

A. Dataset and Implementation Details

The MSTAR dataset includes ten different categories of military vehicles (armored personnel carrier: BMP-2, BRDM-2, BTR-60, and BTR-70; tank: T-62 and T-72; rocket launcher:
Fig. 5. Types of targets in the MSTAR dataset. (Top) Optical images versus (bottom) SAR images.

2S1; air defense unit: ZSU-234; truck: ZIL-131; and bulldozer: D7) in spotlight mode within full aspect, and the resolution is \(0.3 \times 0.3\) m. Examples of optical and SAR images of these targets are shown in Fig. 5.

To comprehensively evaluate the performance of SAR target recognition algorithms, according to different acquisition conditions and target types, the MSTAR dataset is divided into three subsets: the standard operating condition (SOC) and extended operating conditions 1 and 2 (EOC-1 and EOC-2). In the SOC subset, the test and training sets comprise the same target types, albeit with differing azimuth and depression angles. The EOC subset has a more significant disparity between the training and test sets, which includes a greater variation in depression angle (EOC-1) or a difference in the target series version (EOC-2). It is pertinent to note that the above three subsets were artificially subjected to a covariate shift problem, where covariate refers to the input variables (images) of the model. Covariate shift implies that the training and test sets have distinct data distributions [51]. The introduction of a covariate shift in the dataset tests the model’s extrapolation (generalization ability). All experiments were conducted using two Intel Xeon Gold 6240 CPUs and one Nvidia Tesla V100S GPU (32.0 GB). The images in the MSTAR dataset utilized in this study have been resized to \(158 \times 158\) pixels and are in JPG format, except for A-ConvNet, where the input image size is \(88 \times 88\) pixels, as per the original article. During the training process, the stochastic gradient descent (SGD) optimizer was employed, with the learning rate (LR) utilizing the cosine annealing descent method over 100 epochs, with the initial LR set to 0.005 and the final LR set to 0.00001. The batch size during both the training and testing phases was set to 64. No data augmentation was applied.

### B. Classification Results With SOC and Analysis

The training and test sets of SOC consist of ten types of military vehicles that are completely consistent, while the imaging depression and azimuth angles are different, as shown in Fig. 5. Specifically, the azimuth angle is uniformly distributed between \(0^\circ\) and \(360^\circ\), while the depression angles of the training and test sets are \(17^\circ\) and \(15^\circ\), respectively, as shown in Table II. It is widely recognized that SAR image features are highly sensitive to the incident angle of radar beams. Consequently, a slight covariate shift exists between the training and test sets, which can be utilized to evaluate the generalization capabilities of the models.

Upon completion of 100 training epochs, our proposed EMC2A-Net achieved a remarkable accuracy of 99.7%, which is the mean value of five test results reported in this article. Furthermore, Fig. 6 illustrates the confusion matrix of our proposed network, where each row corresponds to the true category labels of the target, and each column represents the predicted category labels of targets. It is noteworthy that the main diagonal elements of the confusion matrix are substantially greater than those in other positions. This indicates that EMC2-A-Net achieved high classification accuracy in the SOC experiment.

To provide a comprehensive analysis of EMC2A-Net’s performance, we conducted a comparison with seven representative CNN models. Table III presents a comparison of the performance and complexity of each model. All models...
achieved an accuracy of over 92%, with DenseNet-121 performing relatively well, indicating that the use of dense connections at various levels can strengthen learned image features. However, EMC\textsuperscript{2}A-Net outperformed all other models and had a relatively small number of trainable parameters and FLOPs. It is noteworthy that EMC\textsuperscript{2}A-Net demonstrated superior performance compared to SKNet and ResNext-50, which are also multibranch networks.

C. Classification Results With EOC and Analysis

In the EOC-1 subset, the test set and training set comprise four identical categories of military vehicles. The azimuth angle distribution of this subset is the same as that of the SOC, ranging from 0\degree to 360\degree. However, the difference between the depression angles of the training and test sets is much larger than that in the SOC, with depression angles of 17\degree and 30\degree, respectively, as indicated in Table IV. As a result, a significant covariate shift characteristic of the EOC subset will inevitably affect the test results.

Remarkably, even with the EOC-1 subset, EMC\textsuperscript{2}A-Net achieves the highest accuracy among all the aforementioned models, albeit with a slight decrease in accuracy (99.5\%). As evidenced by the confusion matrix (see Fig. 7), misclassified samples are predominantly concentrated between 2S1 and ZSU234. In contrast to the results in Tables III and V, the test accuracy of the aforementioned models declined with EOC-1, except for VGG-11, SENet, and SKNet. Notably, DenseNet-121, which is representative of the dense connection model, exhibited the largest drop in test accuracy. These outcomes demonstrate that increasing connections at various levels of the model cannot enhance the model’s generalization ability, while the channel attention mechanism has greater potential for addressing the covariate shift issue.

In the EOC-2 subset, as presented in Tables VI and VII, there exists a partial difference in the depression angles between the training and test sets, although the difference is relatively small. The training set consists of four types of military vehicles, while the test set includes five series of
T72 tanks that are absent from the training set. The distinct series of T72 tanks leads to differences in their SAR image features in detail. The author treats the five series of T72 in the test set as belonging to the same classification (T72), and the four preset classifications of the test set are precisely identical to those of the training set. Therefore, the test set has only one ground-truth classification (T72) among the four classifications. Overall, the EOC-2 subset faces a more severe covariate shift issue than the above two subsets.

The proposed EMC\textsuperscript{2}A-Net model achieved an acceptable accuracy of 95.3\% with the EOC-2 subset. Fig. 8 displays the predicted labels, where the majority of samples are classified as “T72,” while misclassified samples are mainly categorized as “BMP-2” and “BRDM-2.” As shown in Table VIII, all models tested with EOC-2 exhibit significantly lower accuracy than previous measurements, possibly due to the severe covariate shift issue in the EOC-2 subset, where some models may not effectively distinguish similar sample characteristics from specific angles. In contrast, EMC\textsuperscript{2}A-Net and SKNet achieved relatively high accuracy (first and second), with EMC\textsuperscript{2}A-Net being more efficient than SKNet. Interestingly, both models are multibranch networks with a channel attention mechanism, which is a multiscale feature selection mechanism. This mechanism strengthens useful features, thus alleviating the covariate shift problem of the dataset, and weakens negative features, thereby reducing their negative impact on the covariate shift effect. The test with the EOC-2 subset demonstrated that this mechanism has a positive role in combating the dataset’s covariate shift problem.

### D. Other Indicators

In addition to test accuracy, precision, recall, and F1-score [52] are commonly used evaluation indicators for dichotomous classification problems. They are defined as follows:

\[
\text{Precision} = \frac{TP}{TP + FP} \quad \text{(6)}
\]

\[
\text{Recall} = \frac{TP}{TP + FN} \quad \text{(7)}
\]

\[
\text{F1-Score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \quad \text{(8)}
\]

In the above equation, TPs, FPs, and FNs represent true positives, false positives, and false negatives, respectively. In the case of multiclass classification problems such as the SOC and EOC-1 subsets. The macroaverage of the above indicators is utilized to evaluate the performance of the models. These indicators are defined as follows:

\[
\text{macro}_R = \frac{1}{N} \sum_{j=1}^{\text{N}} \text{Recall}_j \quad \text{(9)}
\]

\[
\text{macro}_P = \frac{1}{N} \sum_{j=1}^{\text{N}} \text{Precision}_j \quad \text{(10)}
\]

\[
\text{macro}_F1 = \frac{2 \cdot \text{macro}_P \cdot \text{macro}_R}{\text{macro}_P + \text{macro}_R} \quad \text{(11)}
\]

In (9) and (10), \( N \) represents the total number of classifications, and subscript \( j \) indicates classification \( j \). The precision, recall, and F1-score of each model are as described below.

Table IX demonstrates that the precision, recall, and F1-score metrics of EMC\textsuperscript{2}A-Net surpass those of the other models. It is noteworthy that the test set of EOC-2 comprised four preset classifications, whereas in reality, there was only one classification. Thus, only the recall and F1-scores were computed in this scenario.

The receiver operating characteristic (ROC) curves are a critical evaluation metric for classification performance, depicting the correlation between the TP rate (TPR) and the FP ratio (FPR) as the threshold value utilized in classification is continuously altered. TPR and FPR are defined as follows:

\[
\text{TPR} = \frac{TP}{TP + FN} \quad \text{(12)}
\]

\[
\text{FPR} = \frac{FP}{FP + TN} \quad \text{(13)}
\]

To evaluate the noise robustness performance of the aforementioned network, we introduced 5\% noise into the test set of the EOC-1 subset using the method described in [27]. The model trained on the EOC-1 subset training set (without noise) was utilized for testing, and the results were presented using the ROC curves, as depicted in Fig. 9. The area under the ROC curve (AUC) was adopted as an indicator to evaluate the performance of the models. According to the definition of the ROC curve, models with larger AUCs
TABLE IX
PRECISION, RECALL, AND F1-SCORE RESULTS FOR THE MSTAR DATASET

| Networks       | SOC macro_P | SOC macro_R | SOC macro_F1 | EOC-1 macro_P | EOC-1 macro_R | EOC-1 macro_F1 | EOC-2 Recall | EOC-2 F1-Score |
|---------------|-------------|-------------|--------------|---------------|---------------|---------------|--------------|---------------|
| ResNet-18     | 99.3        | 99.2        | 99.2         | 99.2          | 99.2          | 99.2          | 87.1         | 93.1          |
| ResNext-50     | 98.9        | 98.9        | 98.9         | 98.7          | 98.7          | 98.7          | 82.2         | 90.2          |
| DenseNet-121   | 99.3        | 99.3        | 99.3         | 93.5          | 91.8          | 91.6          | 78.5         | 87.9          |
| VGG-11        | 95.1        | 94.4        | 94.2         | 98.9          | 98.9          | 98.9          | 91.6         | 95.6          |
| SENet         | 92.0        | 92.2        | 92.1         | 95.8          | 95.8          | 95.8          | 83.1         | 90.8          |
| SKNet         | 96.3        | 96.2        | 96.2         | 98.9          | 98.9          | 98.9          | 94.4         | 97.1          |
| A-CONV Net    | 98.7        | 98.7        | 98.7         | 96.2          | 96.2          | 96.2          | 93.8         | 96.8          |
| EMC2A-Net     | 99.7        | 99.7        | 99.7         | 99.5          | 99.5          | 99.5          | 95.3         | 97.6          |

TABLE X
TEST RESULTS OF THE PROPOSED MULTIBRANCH STRUCTURE WITH DIFFERENT COMBINATIONS OF fn and fs IN EMC2-NET

| Stage 1 | Stage 2 | Stage 3 | Stage 4 | #Parameters (M) | Accuracy |
|---------|---------|---------|---------|----------------|----------|
| f1n_1  | f1n_2  | f1n_3  | f1n_4  | f2n_1          | f2n_2    | f2n_3    | f2n_4    | f3n_1          | f3n_2    | f3n_3    | f3n_4    | f4n_1          | f4n_2    | f4n_3    | f4n_4    | Accuracy |
| 1       | 1       | 1       | 1       | 1              | 1        | 1        | 1        | 0.638386       | 97.3     |
| 1       | 1       | 2       | 1.2     | 1              | 1        | 1        | 1        | 0.669746       | 98.9     |
| 1       | 1       | 2       | 1.3     | 1              | 1        | 1        | 1        | 0.669746       | 98.8     |
| 2       | 1.3     | 2       | 1.3     | 2              | 1.2      | 1        | 1        | 0.873846       | 99.1     |
| 2       | 1.2     | 2       | 1.2     | 2              | 1.2      | 1        | 1        | 0.873846       | 99.1     |
| 3       | 1,2,3   | 3       | 1,2,3   | 2              | 1,2      | 1        | 1        | 0.964100       | 99.2     |
| 3       | 1,2,4   | 3       | 1,2,3   | 2              | 1,2      | 1        | 1        | 0.964100       | 99.1     |

Fig. 9. ROC curves and the corresponding AUC values for different classification methods in the EOC-1 antinoise experiment. The test set of the EOC-1 subset was subjected to a 5% noise introduction.

E. Ablation Experiment

To provide an objective and comprehensive explanation of the advantages of the two main innovations proposed in this article, namely, EMC2-A-Net and the EMC2-A module, the authors designed an ablation experiment using the SOC subset. The experiment was divided into two parts.

In the first part, the EMC2-A module was removed from EMC2-A-Net, resulting in a network called EMC2-Net. Based on EMC2-Net, experiments were conducted with single-scale RF and multiscale RF adaptation in four stages. Except for the dilation rate of the convolution kernel (fs_i) and the number of RFs (fn_i) in stage i, the other structural parameters of the network were identical. The results presented in Table X indicate that the parallel multiscale RFs, based on the multibranch structure in EMC2-Net, contribute to improve the recognition accuracy of SAR targets. However, it can also be observed that different combinations of RF scales contribute differently to accuracy improvements.

In the second part, the author added the EMC2-A module to the model with the best accuracy in Table IX, and the test results are presented in Table XI. The incorporation of the EMC2-A module can assist EMC2-Net in further enhancing the recognition accuracy of SAR targets while keeping the generally exhibit better performance. Compared with other models, the proposed EMC2-A-Net achieved a higher AUC and superior performance.

In the first part, the EMC2-A module was removed from EMC2-A-Net, resulting in a network called EMC2-Net. Based on EMC2-Net, experiments were conducted with single-scale RF and multiscale RF adaptation in four stages. Except for the dilation rate of the convolution kernel (fs_i) and the number of RFs (fn_i) in stage i, the other structural parameters of the network were identical. The results presented in Table X indicate that the parallel multiscale RFs, based on the multibranch structure in EMC2-Net, contribute to improve the recognition accuracy of SAR targets. However, it can also be observed that different combinations of RF scales contribute differently to accuracy improvements.

In the second part, the author added the EMC2-A module to the model with the best accuracy in Table IX, and the test results are presented in Table XI. The incorporation of the EMC2-A module can assist EMC2-Net in further enhancing the recognition accuracy of SAR targets while keeping the generally exhibit better performance. Compared with other models, the proposed EMC2-A-Net achieved a higher AUC and superior performance.

E. Ablation Experiment

To provide an objective and comprehensive explanation of the advantages of the two main innovations proposed in this article, namely, EMC2-A-Net and the EMC2-A module, the authors designed an ablation experiment using the SOC subset. The experiment was divided into two parts.
number of added parameters relatively small. In addition, as detailed in Section III-A3, the author examined the performance of different values of $R_ncks$ with the EMC$^2$A module. Interestingly, the results indicate that for a specific number of channels, a smaller kernel size has a more significant impact on improving the EMC$^2$A module’s performance. As a result, the author set $R_ncks$ to 32, and eventually appended the EMC$^2$A module to EMC$^2$-Net to generate EMC$^2$A-Net.

F. Classification Experiments With Training Dataset of Small Size

To validate the effectiveness of the models in situations with small sample sizes, a series of comparative experiments were conducted involving EMC$^2$A-Net and the other aforementioned models. The training set samples for each category were selected randomly from the SOC training set, with the number of selected samples ranging from 70 to 220 per category and incrementing by 30 (approximately 10% of the number of each category). The SOC test set was used as the testing data. The outcomes of these experiments are presented in Fig. 10.

As shown in Fig. 10, when the number of training samples per category is limited to 70, EMC$^2$A-Net, ResNet-18, Resnext-50, and DenseNet-121 demonstrate higher classification accuracies than the other networks, achieving accuracies of 95.31%, 95.84%, 95.88%, and 94.16%, respectively. In addition, the results reveal that when there are more than 100 training samples available per category, EMC$^2$A-Net outperforms other models, achieving the highest accuracy of 99.46% (with 220 samples). These results demonstrate that due to its efficient structure, EMC$^2$A-Net achieves superior classification performance and mitigates the risk of overfitting, even when the training sample size is limited.

To further verify the antinoise performance of the four top-performing models in the above experiments under small training sample conditions, we added noise only to the test set samples instead of the training set. The noise simulation method referred to the literature [27], which randomly selects a certain proportion of pixels and replaces their values with samples from a uniform distribution. Fig. 11 shows the target images under different noise levels.

First, we added 5% noise to the samples of the test set, and the classification accuracies of the four top-performing models are presented in Fig. 12. It can be observed that EMC$^2$A-Net exhibits significant advantages in terms of classification accuracy under different numbers of training samples. In addition, the classification accuracies of the four models decrease as the number of training samples diminishes. However, even under the condition of only 70 samples per category, EMC$^2$A-Net still maintains a classification accuracy of over 80% (80.09%).

Second, we fixed the number of training samples to 100 for each category and evaluated the classification accuracies of the four models under different noise levels. Fig. 13 shows that EMC$^2$A-Net demonstrates remarkable superiority in classification accuracy under different noise levels, particularly achieving a classification accuracy of 67.69% at the 15% noise level, while the accuracies of the other networks all decreased to below 40%. The results also indicate that increasing noise levels have a considerable impact on the recognition accuracy.
In the aforementioned two sets of comparative experiments, except for EMC\textsuperscript{2}A-Net, Resnext-50 performed relatively better. We believe that the multiscale RFs of convolution kernels play an important role in feature extraction and noise reduction. It is noteworthy that our proposed EMC\textsuperscript{2}A-Net not only features convolution kernels with multiscale RFs but also has an even more efficient and lightweight structure, along with a channel attention module, all of which play critical roles in achieving outstanding performance.

V. CONCLUSION

In this article, aiming to solve the problem of SAR target recognition, we proposed two residual blocks with multiscale RFs based on a multibranch structure and further designed a four-stage isotopic architecture DCNN, EMC\textsuperscript{2}A-Net, based on the aforementioned blocks. EMC\textsuperscript{2}A-Net has an efficient and lightweight structure. In addition, the authors proposed a multiscale feature cross-channel attention module, the EMC\textsuperscript{2}A module, which is capable of adaptively generating channel weights by fusing the information of multiscale features from channels. The introduction of this module provided a boost to the general performance of the model without resorting to the data augmentation method. The performance evaluation results based on the MSTAR dataset show that compared with other classical and same-type target recognition models, EMC\textsuperscript{2}A-Net exhibits better classification performance and outstanding generalization ability.

Follow-up work will aim to enhance the performance of EMC\textsuperscript{2}A-Net, as presented in this article. Specifically, in terms of the basic blocks of EMC\textsuperscript{2}A-Net, it has been observed that the GAP operation compresses the feature maps with equal weights on a pixel-wise basis, which may result in substantial information loss. Consequently, the pooling operation should be improved in future work. In terms of the model design method, the EMC\textsuperscript{2}A-Net design rule was established based on probability and mathematical statistics theories, driven by the available datasets. However, the current publicly available SAR target recognition datasets are limited, hindering the discovery of more general design rules. As a result, we plan to collaborate with our partners to build new SAR datasets with the objective of advancing this work.

REFERENCES

[1] Z. J. Liu, L. K. Zhuang, Y. F. Cao, and M. Ding, “Target recognition of SAR images using principal component analysis and sparse representation,” Syst. Eng. Electron., vol. 35, no. 2, pp. 282–286, Oct. 2013, doi: 10.3969/j.issn.1001-506X.2013.02.08.
[2] R. H. Huan and R. L. Yang, “Synthetic aperture radar images target recognition based on wavelet domain NMF feature extraction,” J. Electron. Inf. Technol., vol. 31, no. 3, pp. 588–591, Mar. 2009, doi: 10.1016/j.ajep.2007.10.019.
[3] Z. G. He, J. Lu, and G. Y. Kuang, “A survey on feature extraction and selection of SAR images,” Signal Process., vol. 24, no. 5, pp. 813–823, Oct. 2008.
[4] J. Lou, T. Jin, Q. Song, and Z.-M. Zhou, “Feature extraction of scattering centers in high-resolution SAR image,” J. Electron. Inf. Technol., vol. 33, no. 7, pp. 1661–1666, Aug. 2011, doi: 10.3724/spj.1146.2010.00960.
[5] T. D. Ross, S. W. Worrell, V. J. Velten, J. C. Mosing, and M. Lee Bryant, “Standard SAR ATR evaluation experiments using the MSTAR public release data set,” Proc. SPIE, vol. 3370, pp. 566–573, Sep. 1998.
[6] Q. Zhao and J. C. Principe, “Support vector machines for SAR automatic target recognition,” IEEE Trans. Aerosp. Electron. Syst., vol. 37, no. 2, pp. 643–654, Apr. 2001, doi: 10.1109/7.937475.
[7] Y. Sun, Z. Liu, S. Todorovic, and J. Li, “Adaptive boosting for SAR automatic target recognition,” IEEE Trans. Aerosp. Electron. Syst., vol. 43, no. 1, pp. 112–125, Jan. 2007, doi: 10.1109/taes.2007.357120.
[8] G. Dong, G. Kuang, N. Wang, L. Zhao, and J. Lu, “SAR target recognition via joint sparse representation of monogenic signal,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 8, no. 7, pp. 3316–3328, Jul. 2015, doi: 10.1109/jstars.2015.2436694.
[9] M. Amrani, K. Yang, D. Zhao, X. Fan, and F. Jiang, “An efficient feature selection for SAR target classification,” in Advances in Multimedia Information Processing—PCM, B. Zeng, Q. Huang, A. El Saddik, H. Li, S. Jiang, and X. Fan, Eds. Cham, Switzerland: Springer, 2018, pp. 68–78.
[10] M. Amrani, F. Jiang, Y. Xu, S. Liu, and S. Zhang, “SAR-oriented visual saliency model and directed acyclic graph support vector metric based target classification,” IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens., vol. 11, no. 10, pp. 3794–3810, Oct. 2018, doi: 10.1109/jstars.2018.2866684.
[11] M. Amrani, S. Chaib, I. Omarra, and F. Jiang, “Bag-of-visual-words based feature extraction for SAR target classification,” in Proc. 9th Int. Conf. Digit. Image (ICDIP), 2017, pp. 1–6, doi: 10.1117/12.2281707.
[12] M. Amrani and F. Jiang, “Deep feature extraction and combination for synthetic aperture radar target classification,” J. Appl. Remote Sens., vol. 11, no. 4, p. 1, Oct. 2017, doi: 10.1117/1.jrs.11.042616.
[13] M. Amrani, A. Bey, and A. Amamra, “New SAR target recognition based on YOLO and very deep multi-canonical correlation analysis,” Int. J. Remote Sens., vol. 43, nos. 15–16, pp. 5800–5819, Aug. 2021, doi: 10.1080/01431161.2021.1953719.
[14] Z. Feng, Y. Xu, B.-N. Wang, X. Yu, D.-Y. Zhu, and G. Zhang, “Target recognition in SAR images by deep learning with training data augmentation,” Sensors, vol. 23, no. 2, p. 941, Jan. 2023, doi: 10.3390/s23020941.
[15] Y. Gu and Y. Xu, “Architecture design of deep convolutional neural network for SAR target recognition,” J. Image Graph., vol. 23, no. 6, pp. 928–936, 2018, doi: 10.11184/j.170473.
[16] D. H. Hubel and T. N. Wiesel, “Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex,” J. Physiol., vol. 160, no. 1, pp. 106–154, Jan. 1962, doi: 10.1113/jphysiol.1962.sp006837.
[17] C. Szegedy et al., “Going deeper with convolutions,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Boston, MA, USA, Jun. 2015, pp. 1–9.
[18] S. Yang, G. Lin, Q. Jiang, and W. Lin, “A diluted inception network for visual saliency prediction,” IEEE Trans. Multimedia, vol. 22, no. 8, pp. 2163–2176, Aug. 2020, doi: 10.1109/tmm.2019.2947352.
[19] X. Li, W. Wang, X. Hu, and J. Yang, “Selective kernel networks,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Long Beach, CA, USA, Jun. 2019, pp. 510–519.
[20] J. Ai, Y. Mao, Q. Luo, L. Jia, and M. Xing, “SAR target classification using the multikernel-size feature fusion-based convolutional network,” IEEE Trans. Geosci. Remote Sens., vol. 60, Sep. 2022, Art. no. 5214313, doi: 10.1109/TGRS.2021.3106915.
[47] S. Santurkar, D. Tsipras, A. Ilyas, and A. Madry, “How does batch normalization help optimization?” in Proc. 32nd Conf. Neural Inf. Process. Syst. (NIPS), Montréal, CA, USA, 2018, pp. 1–11.

[48] J. Bourvie, “Notes on convolutional neural networks,” Neural Netw., vol. 19, no. 2, pp. 137–144, 2006.

[49] E. R. Keydel, S. W. Lee, and J. T. Moore, “MSTAR extended operating conditions: A tutorial,” in Proc. 3rd SPIE Conf. Algorithms SAR Imagery, Orlando, FL, USA, 1996, pp. 228–242.

[50] The Air Force Moving and Stationary Target Recognition Database. Accessed: Apr. 2, 2016. [Online]. Available: https://www.sdms.afrl.af.mil/datasets/mstar

[51] G. Dharani, Y. N. G. Nair, P. Satpathy, and J. Christopher, “Covariate shift: A review and analysis on classifiers,” in Proc. Global Conf. Advancement Technol. (GCAT), Bengaluru, India, Oct. 2019, pp. 1–6.

[52] F. Yang, Q. Xu, and B. Li, “Ship detection from optical satellite images based on saliency segmentation and structure-LBP feature,” IEEE Geosci. Remote Sens. Lett., vol. 14, no. 5, pp. 602–606, May 2017, doi: 10.1109/LGRS.2016.2551720.
Zhe Geng (Member, IEEE) received the Ph.D. degree in electrical engineering from Florida International University, Miami, FL, USA, in 2018. From 2018 to 2019, she was a Research Scientist with Wright State University, Dayton, OH, USA. In 2019, she joined the College of Electronic and Information Engineering, Nanjing University of Aeronautics and Astronautics, Nanjing, China, where she is currently a Research Associate. Her research interests include deep learning, image processing, and automatic target recognition.

Xiaohua Huang (Senior Member, IEEE) received the B.S. degree in communication engineering from Huaqiao University, Quanzhou, China, in 2006, and the Ph.D. degree in computer science and engineering from the University of Oulu, Oulu, Finland, in 2014. He was a Senior Researcher with the Center for Machine Vision and Signal Analysis, University of Oulu, from 2015 to 2019. He was also a Research Associate with the University of Cambridge, Cambridge, U.K. He has been a Research Assistant with Southeast University, Nanjing, China, since 2006. He is currently a Professor with the Nanjing Institute of Technology, Nanjing, and a Distinguished Professor of Jiansu Province. He has authored or coauthored more than 40 articles in journals and conferences. His research interests include emotion recognition, object detection, handwritten recognition, and texture classification. Dr. Huang has served as a reviewer for respective journals and conferences.

Qinglu Wang was born in Fujian, China, in 2001. He is currently pursuing the bachelor’s degree with the Nanjing Institute of Technology, Nanjing, China. His research interests include remote sensing image processing and artificial intelligence.

Daiyin Zhu (Member, IEEE) was born in Wuxi, China, in 1974. He received the B.S. degree in electronic engineering from Southeast University, Nanjing, China, in 1996, and the M.S. and Ph.D. degrees in electronics from the Nanjing University of Aeronautics and Astronautics (NUAA), Nanjing, in 1998 and 2002, respectively. From 1998 to 1999, he was a Guest Scientist with the Institute of Radio Frequency Technology, German Aerospace Center, Oberpfaffenhofen, Germany, where he worked in the field of synthetic aperture radar (SAR) interferometry. In 1998, he joined the Department of Electronic Engineering, NUAA, where he is currently a Professor. He has developed algorithms for several operational airborne SAR systems. His research interests include radar imaging algorithms, SAR/inverse SAR (ISAR) autofocus techniques, SAR/ground moving target indication (SAR/GMTI), and SAR interferometry.