SAR Image Classification: A Comprehensive Study and Analysis

ALICIA PASSAH[1], **SAMARENDRA NATH SUR**[2], **(Senior Member, IEEE),**
BABUSENA PAUL[3], **(Member, IEEE), AND DEBDATTA KANDAR**[1]

[1]Department of Information Technology, North-Eastern Hill University, Shillong, Meghalaya 793022, India
[2]Department of Electronics and Communication Engineering, Sikkim Manipal Institute of Technology, Sikkim Manipal University, Rangpo, Sikkim 737136, India
[3]Institute for Intelligent Systems, University of Johannesburg, Johannesburg 2006, South Africa

Corresponding author: Debdatta Kandar (kdebdatta@gmail.com)

ABSTRACT Deep learning has obtained wide attention in various fields enabling systems to derive essential information from digital inputs. Lately, the use of deep learning in remote sensing applications has also been motivated and applied, wherein considerable improvements in the results are witnessed. Synthetic aperture radar images have been used in various earth observation systems because of their all-day imaging capacity and self-illuminating nature. Various works concentrating on extracting meaningful information from SAR data for various other applications have been proposed in the literature. Classification of SAR images has been one of the utmost steps in numerous SAR applications. Therefore, this work focuses on studying several existing techniques that use deep learning for SAR image classification by examining the architectures involved. Based on the study, crucial observations are made, highlighting the merits and demerits of several approaches, allowing researchers to better understand how the methods can impact the performance of the deep learning models for SAR image classification in the future. Potential hybrid models for the classification of SAR images are also presented in this paper.

INDEX TERMS Deep learning, image classification, image processing, remote sensing, synthetic aperture radar.

I. INTRODUCTION

With enormous volumes of data produced from Synthetic Aperture Radar (SAR) and many other SAR carrying satellites, the processing and interpretation of SAR data have become immediate for a wide range of applications. The SAR data are worth processing because of the advantages it offers. SAR is a radar attached to air-crafts or any other moving platforms and is used to construct images from the returned signals of the electromagnetic waves that it transmits to the earth's surface. SAR can operate regardless of weather situations and with no time constraints. The images obtained from SAR have been applied to numerous applications that include military surveillance, disaster management, maritime vigilance, inspection of illegal mining activities [1], etc. With more SAR sensors mounted satellites emerging in the coming years, gigantic data needs to be processed, archived and analyzed. Accessibility of such SAR data is challenging as

they are difficult to interpret. One of the major processing steps of SAR images is classification since classification is profitable in categorizing the detected targets based on their classes and can be applied in real-world scenarios.

Recently, machine learning methods, notably deep learning, have adequately improved the performances of various SAR image processing methods wherein they outperform traditional methods [2]–[8], but the existence of speckle noise in SAR images makes the interpretation challenging and hence difficult for posterior processing such as classification. Numerous methods have been explored in the literature for performance improvement in SAR image interpretation [9]–[12]. Despite SAR having advantageous characteristics, various issues referring to the adoption of SAR images in several applications still needs attention. Issues include inaccurate classification of objects due to inappropriate recognition of features associated with a particular target. The grainy features present in SAR images are one of the major causes of issues such as misinterpretation [13], [14]. Enhancing SAR images by

---

The associate editor coordinating the review of this manuscript and approving it for publication was Sudipta Roy.

removal of such unwanted features also gave rise to another issue that is the unintentional removal of relevant features during the noise removal process resulting in a misclassification issue.

Several works on SAR image classification focusing on various issues have been proposed in the literature, especially using the prominent deep learning models. A study on various existing SAR image classification approaches concentrating on the techniques adopted, including architectures involved, along with the advantages and various disadvantages, would help researchers to get insights about capable approaches to be adopted in the future for SAR image classification. Several informative surveys have been done in the literature regarding the processing of SAR images, including classification giving constructive conclusions. The comparisons of the different surveys are shown in Table 1. El-Darymli *et al.* [15] have surveyed and introduced classification methodologies in SAR analysis. In a different survey, El-Darymli *et al.* [16] have assessed the different methods in SAR target interpretations. The work by Zhu *et al.* [17] analyses the different challenges faced when deep learning meets remote sensing. Another work by Yao *et al.* [18] focuses on the different classification methods in remote sensing along with their configurations by also highlighting the existing problems and future works. Wang *et al.* [19] reviewed the different SAR image classification algorithms from traditional methods to deep learning methods. It may be mentioned that all the aforementioned surveys have helped researchers in the field of SAR image analysis. However, it would be more helpful if the different deep learning architectures used in the literature for SAR image classification along with the future models were also highlighted. Therefore this has motivated us in conducting a study on various works related to SAR image interpretation by focusing mainly on one of the major processing steps of SAR image interpretation called classification by going in-depth the techniques and architectures involved with an aim to ease researchers and those who are new to the field of SAR imaging, enabling them to be at par with recent advances in this area. The main contributions of the paper are as follows:

1) State-of-the-art works related to SAR image classification are discussed.
2) The deep learning models adopted for SAR image classification, along with the different parameter settings, are also discussed.
3) The advantages and disadvantages of the different approaches are highlighted, along with different issues and challenges.
4) Future approaches obtained from the study for improving the performances in the future are also discussed.
5) Based on the study, potential hybrid models for future adoption are also highlighted.

This paper considers the performances of various SAR image classification approaches, and a comparison of the different methods has also been made, which will help in understanding and solving existing issues bagged with SAR image classification. The rest of the paper is organized as follows. Section II discusses the background study, comprising SAR image processing, deep learning and convolutional neural networks. The study on SAR image classification is discussed in Section III. Section IV discusses the research issues and challenges, followed by a discussion on future approaches in Section V and a conclusion in Section VI.

TABLE 1. Different survey works on SAR image analysis.

Paper	Survey Covered	Architecture Explained	Future Discussion	Potential Models
[15]	✓	✗	✓	✗
[16]	✓	✓	✓	✗
[17]	✓	✓	✓	✗
[18]	✓	✗	✓	✗
[19]	✓	✓	✗	✗
[20]	✓	✓	✗	✗
Our Work	✓	✓	✓	✓

II. BACKGROUND

This section presents the background study by introducing SAR image processing followed by explanations on deep learning in SAR processing and convolutional neural networks.

A. SAR IMAGE PROCESSING

SAR is a radar that is used for constructing two-dimensional images by transmitting electromagnetic waves to the surface of the earth with the help of a transmitter. These waves get reflected from the earth's surface in the form of echos to the receiver of the radar, and images are constructed with respect to the received signals [13]. Figure 1 shows the basic block diagram of a typical radar system [21].
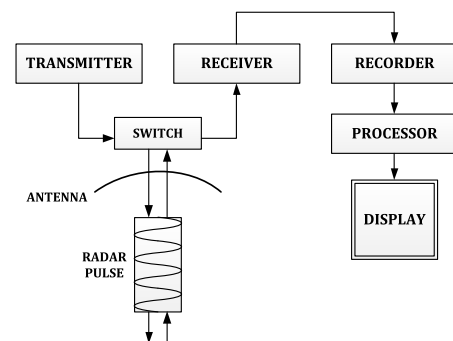

FIGURE 1. The basic block diagram of a radar [21].

The flight path of the platform in which SAR is mounted determines the large antenna. Because of this simulated antenna, it is possible to regenerate the signal which would have been attained by an antenna of length $v \times T$, where v is the platform speed, and T is the period traveled by the radar from one position to another [22]. Typically, with a

larger aperture and bandwidth, the resolution of the image becomes high. This enables SAR to generate images of higher resolution even with smaller physical antennas. However, images obtained from SAR are usually distorted and highly corrupted due to the ability to function in various conditions, making the system more susceptible to speckle noises. This is because the reflected signals may not be in phase due to the coarseness of the objects or targets and the pathway in which the signal navigates towards the receiver. The occurrence of speckle noise in SAR images makes understanding the images difficult and complicated even by humans. Therefore to make machines understand the complex SAR images is very challenging. Despite the projection of various algorithms in the literature [2], [23], [24], loss of information during the process is still a concern, and this leads to erroneous interpretation in subsequent stages such as SAR image classification.

B. DEEP LEARNING IN SAR PROCESSING

One of the prominent branches of machine learning, called deep learning, is discussed in this section. Deep learning, whose working is a mimicry of the human brain, is comprised of multi-layered neural networks to learn high-level features from raw data [25] successively. By multi-layered neural networks, it means that the lower layers extract low-level information or features such as edges, and higher layers extract high-level information such as faces, letters or digits [26]. Unlike primitive methods where features are extracted manually, feature extraction in deep learning appears to be automatic inside the network. Another major advantage of deep learning compared to other algorithms is that it can learn automatically from unlabeled data [27]. Deeper networks can train large volumes of data giving better accuracy than traditional methods. The work in [28] have shown that deep learning could substantially reduce the error rates in image classification [29], [30]. Therefore researchers have started applying deep learning in various other fields including SAR image applications [3], [4], [31], [32]. The adoption of deep learning approaches for various SAR image processing tasks has resulted in considerable improvement in the performances [5], [33], [34]. However, the presence of noise in SAR images weakens the ability of deep learning to make accurate predictions and hence reduces the applicability of SAR images in potential fields of remote sensing. Research on boosting the performance of SAR interpretations like image classification using deep learning for future applications is still being carried out. Existing works for various SAR image classifications will be discussed in the later sections. The most widely used deep learning network called convolutional neural network is briefly discussed in the next section.

C. CONVOLUTIONAL NEURAL NETWORKS

The Convolutional Neural Network, also known as CNN, is one of the most widely used deep learning algorithms that is being adopted for various computer vision tasks since the witness of their capabilities in the ImageNet Large Scale Visual Recognition Competition (ILSVRC) [35]. The CNN consists of an input layer, convolutional layers, sampling layers and fully connected layers [36], [37]. In the case of image applications, the CNN takes in input and allocates weights and biases to various patterns of the objects present in the input image, resulting in the capability to differentiate between the different objects in the image. With adequate training, the CNN automatically learns the relevant filter values without human-engineered filters. The configurations of CNN depend upon the application domain, and hence the performance differs from one application to another. Because of its efficiency, CNN has been adopted in numerous research areas showing commendable outcomes. It has also been adopted for various SAR image analysis such as denoising for noise suppression, detecting SAR targets, and even for SAR image classification.

Some of the well known CNN architectures from the ILSVRC event [35] that are being applied to various applications comprise of the LeNet [38], Alexnet [28], ZFNet [39], Inception [40], VggNet [41], ResNet [42]. LeNet architecture is made up of 7 layers and has an error rate below 1% on the MNIST dataset. Alexnet has a more profound architecture similar to LeNet and has 60 million parameters with an error rate 15.3%. On the other hand, the structure of ZFNet is identical to Alexnet but differs in the size of the filter and the convolutional stride of the first layer. ZFNet obtained an error rate of 14.8%. The Inception architecture consists of 22 layers and is an improvement of Alexnet as it reduces the parameters from 60 million to only 4 million. However, the architecture used in Inception was influenced by the LeNet model. Inception achieved 6.67% error rate. An orderly placed architecture consisting of 16 convolutional layers forms the VggNet. VggNet consists of 138 million parameters with 7.3% error rate. The residual network popularly known as ResNet is made up of 152 layers with skip connections. The complexity of ResNet is lower than that of VggNet, and it achieved an error rate of 3.57%. The adoption of this architecture in various applications using optical images has been of great help in improving the performances [43]–[46]. In SAR images analysis, the use of the aforementioned architectures is still limited, hence leaving research space in this area. While this has encouraged researchers to adopt CNN in SAR image analysis, it is observed that the classification and interpretation of SAR images is still challenging because of the complex and noisy characteristics of SAR. Therefore the main focus of this paper is to study recent existing works, especially the deep learning-based approaches in the field of SAR image classification addressing different issues and challenges along with their advantages and disadvantages, and are discussed in the next section.

III. SAR IMAGE CLASSIFICATION

This section discussed the classification of SAR images, followed by the literature survey on various works related to SAR image classification. Classification is a process of categorizing a given set of data into classes based on the features observed on the training set of data. Hence, SAR

image classification refers to the process of categorizing the objects or targets present in the image into their respective classes. For instance, categorizing the different marine targets present in a given SAR image based on the target features is called classification.

Classification of SAR targets is a challenging field of research. Even with the identity of deep learning to be one of the best target classifiers, as is proved by the ILSVRC competition [35], misclassification of targets in SAR is still a major issue that is being researched to date. The deteriorating quality of SAR images makes the classification algorithm hard to differentiate between relevant and non-relevant features in order to understand the target type for proper classification. On the other hand, similar-looking targets are also hard to distinguish because features that differentiate between the two might have been suppressed or ignored by a classifier during the process of noise suppression [47]. Therefore, classifying similar-looking SAR targets is also an open field of research at present. Various works on SAR image classification considering several issues have been proposed in the literature, each having its own advantages and disadvantages, whose details are discussed in the following subsections.

A. DNN-DAE-CONV

Even with the increase in SAR satellites resulting in the need to process more remote sensing data, there is insufficient labeled SAR data to enable the use of automated models like deep learning for various applications such as the classification of oceanographic objects from SAR images. The aim of the work in [48] is to apply deep neural networks for oceanographic object classification with less labeled data. The authors in [48] have incorporated two models wherein the first model comprised of the Deep Neural Networks with denoising auto-encoder (DNN-DAE) and the other model is called the classification model wherein it uses DNN with convolutional layers (DNN-Conv). We, therefore, refer to this work as DNN-DAE-Conv (DNN with Denoising Auto Encoder and Convolution). The reason for incorporating denoising auto-encoder is to enable the learning of higher-level representation of features. The flowchart of the DNN-DAE-Conv [48] is shown in Figure 2. The noisy SAR input image is first passed through a constant false alarm rate (CFAR) for object detection. The detected regions are extracted, normalized, and then used as input to the unsupervised block of the DNN-DAE. The output of the unsupervised block then enters the supervised block wherein training was performed using labeled data from a database containing targets that are identified manually and labels generated from Automated Identification System (AIS). The DNN-DAE model could then produce labeled data with the help of only relevant features and hence use the labeled data for classification in DNN-Conv. In conclusion, the DNN-DAE-Conv [48] was able to learn only the higher-level representation of features because of the stacked auto-encoder layers in the unsupervised block [49]–[51]. As future work, denoising the input image before detecting targets may

be helpful. Incorporating the DNN-DAE with deeper CNNs clubbed together, forming a hybridized model can be worked upon experimentally and may improve the network performance as a whole.
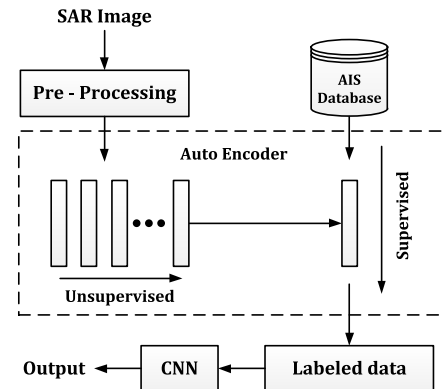

FIGURE 2. Flowchart depicting the technique used in DNN-DAE-Conv [48].

B. A-CONVNET

When convolutional neural networks are applied to SAR-Automatic Target Recognition (SAR-ATR), it results in severe overfitting due to the availability of limited SAR datasets. Also, the classification performance on the Moving and Stationary Target Acquisition and Recognition (MSTAR) targets usually degrades, especially in Extended Operating Conditions (EOC), because of the change in position of the target features like the turret or fuel drums, leading to the differences in the target configurations from those in the database. With the advances of deep learning, the application of networks such as CNN to SAR-ATR tasks resulted in severe overfitting because SAR data are not large enough. The work in [31] proposes an all convolutional neural network called A-Convnet that focuses on mitigating the overfitting issue as well as limited dataset issue while applying CNN to MSTAR target classification. The novelty of the A-ConvNet architecture is the non-inclusion of the fully connected layers allowing fewer degrees of freedom in the model, and the architecture along with its hyperparameter settings are shown in Figure 3 and Table 2 respectively. It may be noted that data augmentation was also performed to enlarge the dataset for training the A-ConvNet. Classification results of experiments tested under EOC and Standard Operating Conditions (SOC) showed improvements but resulted in few misclassifications. Though A-ConvNet outperforms several existing methods [52]–[57], its performance dropped approximately by 7% when used for images with even just 1% of noise [31]. This shows that the antinoise performance of A-ConvNet is low. An end-to-end based experiment was also conducted, where detection of MSTAR targets from a cluttered environment was also considered before classification. Two stages of A-ConvNet were used in which the first stage consists of a binary classifier in order to classify target and clutter, followed by the second stage, which is the A-ConvNet itself. Results show 98% accuracy with few false alarms and

false recognition in images containing no noise. A-ConvNet can be trained further for noisy images, or a preprocessing stage may be added to its model to improve performance and make it adaptable to noisy images.
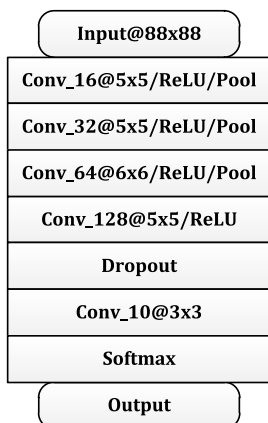

FIGURE 3. Architecture used in A-ConvNet [31].

TABLE 2. Hyperparameters used in A-ConvNet [31].

Hyperparameters	Values
Mini batch size	100
Momentum	0.9
Weight decay	0.004
Learning rate	0.001

C. CNN-MR

Classification of maritime targets has become an apple of the eye for many researchers in SAR imaging. Therefore numerous works have been done to classify maritime targets from challenging SAR images [31], [48], [58], [59]. However, there still lies concerns about misclassification issues faced by the existing works. Therefore to mitigate the misclassification issue, the authors in [47] have come up with the multiple resolutions convolutional neural network model named CNN-MR (Multiple Resolution CNN). As the name suggests, this model learns a mapping from inputs with multiple resolutions ($3m$, $12m$, $24m$) to the corresponding targets. The intention of involving multiple resolution inputs was to enable the model to learn more related features in order to result in better classification accuracy. The architecture of the CNN-MR model shown in Figure 4 was designed by collectively observing four different networks, namely DNN-DAE-Conv [48], all-in-one CNN [58], A-ConvNet [31] and CNN-A [59] proposed in the literature and comparing them with a baseline classifier which is the SVM-PCA (Support Vector Machine based on Principle Component Analysis) since it is an efficient classifier [60]. The CNN-MR model uses targets from the TerraSAR-X data detected using the Constant False Alarm Rate (CFAR). Experimentally it was observed that the all-in-one CNN [58] model performs similar

to SVM-PCA but yet suffers from misclassification. This is because the model was too complex, and larger datasets may improve the performance of such complex networks. Also, models DNN-DAE-Conv [48], A-ConvNet [31] and CNN-A [59] balanced well with the internal parameters. The activation ReLU (Rectified Linear Unit) [61] is preferable for internal layers, whereas, for dense layers, either ReLU or Softmax works. Based on these observations, the CNN-MR was modeled, resulting in further performance improvement due to multiple resolution inputs and well-tuned architecture. Table 3 shows the f1-score comparisons between networks DNN-DAE-Conv, all-in-one CNN, A-ConvNet, CNN-A and CNN-MR.
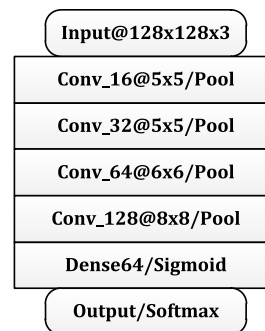

FIGURE 4. Architecture used in CNN-MR [47].

TABLE 3. Table showing average f1-scores of [31], [48], [58], [59] and [47].

NetworkModels	Avergae f1-score
DNN-DAE-Conv [48]	0.87
all-in-one CNN [58]	0.82
A-ConvNet [31]	0.87
CNN-A [59]	0.89
CNN-MR [47]	0.94

D. SI-CNN

SAR images are predominantly used in sea ice monitoring [62], [63]. Of the aforementioned techniques, the work in [32], hereafter known as SI-CNN, aims at classifying the different types of sea ice and sea from the SENTINEL-1 SCANSAR image with the help of a convolutional neural network. The SENTINEL-1 product is first pre-processed using the ESA SNAP tool in SI-CNN. Pre-processing includes radiometric correction, thermal noise removal and Lee filtering. The image is then classified manually into ice and sea-ice, resulting in a series of class layers. Chips with three spatial scales 32 × 32, 64 × 64, 128 × 128 are then extracted from the class layers for training and validation. This is achieved by checking each pixel, and if matched with the true class layer, chips with the checked pixel at the center are cropped. Before feeding and training the CNN, the chips with size 64 × 64 and 128 × 128 are resized to 32 × 32 since images fed to the CNN are required to be of uniform sizes. The CNN architecture used in SI-CNN is

basically the traditional convnet architecture and is shown in Figure 5. It consists of 3 convolutional layers, each followed by a pooling and normalization layer. It also consists of 2 fully connected layers, followed by a dropout layer. A softmax layer is appended at the end of the model to produce the final output.
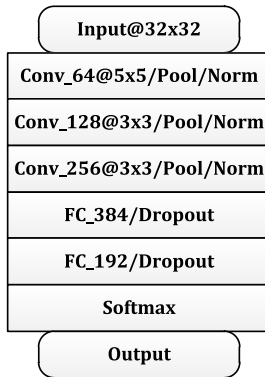

FIGURE 5. Architecture used in SI-CNN [32].

The hyper-parameters used are shown in Table 4.

TABLE 4. Hyperparameters used in SI-CNN [32].

Hyperparameters	Values
Moving average decay	0.9999
Epoch for LR decay	100
LR decay	0.6
Initial LR	0.05

The results show that the classification of grained ice, striped ice, rough sea and smooth sea is promising and achieved a precision of $0.945 - 0.922$. However, in some cases, massive ice is wrongly classified as the rough sea, and smooth ice is wrongly classified as the smooth sea. The reason for misclassification is due to the fact that fewer features are learned by the model in each scale of the wrongly classified classes. The performance of SI-CNN [32] can further be improved if various window sizes for finding the optimal feature scale used for feature description are explored. The next aim is to reduce the need to classify the class layers manually.

E. M-NET
Various works have shown considerable improvements with the advancement of deep learning and its application in various SAR image processing problems. But the non-availability of the SAR dataset is causing an issue, and therefore training deep learning models with a limited dataset usually results in over-fitting. To mitigate the issue of over-fitting caused by a limited dataset, the authors in [3] proposed a deep memory network called M-Net that aimed at classifying several MSTAR targets using a limited dataset. An information recorder whose format is shown in Figure 5, was designed along with a mapping matrix in order to save and remember

the spatial features of samples and use these features to predict the unseen samples with the help of spatial similarity measure. The CNN part of M-Net first extracts the features of an input SAR image, and the resulting vector gets multiplied with a mapping matrix resulting in another vector that serves as a query to the memory. The system will match the query with the feature records stored in the information recorder, and the features that match the most will select their corresponding label as the final classification output. The mapping matrix keeps the information recorder's size unchanged as the vector size changes, thereby reducing the dimensions when the feature vectors have large dimensions.

TABLE 5. Table showing information recorder adopted in M-Net [3].

Index	Key	Value	Time
$1....m$	$k_1....k_m$	$v_1....v_m$	$t_1....t_m$
*Index: Item indicator for searching and writing			
*Key: Item key			
*Value: True class label			
*Time: Time each record was last updated			

On the other hand, the convergence of the M-Net model turned out to be problematic because of the involvement of an information recorder and a matrix amidst the model, which results in unstable and slow convergence. However, considering this issue, the authors have used pre-tuned parameters trained on a CNN model same as the CNN used in M-Net except that it uses softmax as a classifier instead of matrix and information recorder and a cross-entropy loss [64] as a loss function instead of hinge loss [65]. The M-net is then trained using the pre-tuned parameters. The entire process is shown in Figure 6. The architecture of the CNN used in MNet is shown in Figure 7.
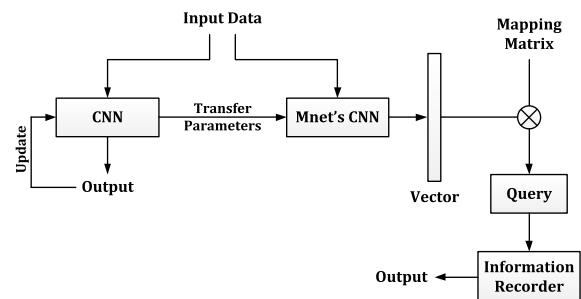

FIGURE 6. Flowchart summarizing the technique in M-Net [3].

Simulation results on the augmented MSTAR dataset show that M-Net outperforms other methods, namely SAEED (SAR Auto-Encoder based on Euclidean Distance) [66], A-Convnet [31] and SVM [67] but yet suffers few misclassifications, especially under extended operating conditions. However, MNet [3] worked better than the other comparing algorithms when there were variations in the number of training samples. Table 6 shows comparison results of
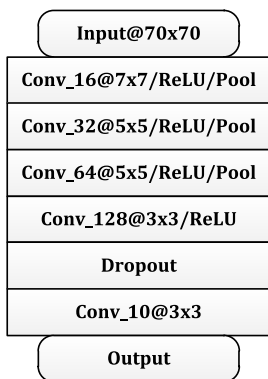
FIGURE 7. Architecture used in M-Net [3].

M-Net with other methods. The major advantage of MNet is that it can extract more relevant features from fewer training samples and are better separated than conventional CNN. Several improvements can be worked upon related to MNet to solve the misclassification issue, and the information recorder can be modified to enable it to output probabilistic values rather than deterministic ones.

TABLE 6. Comparison table showing average accuracy of the well-known algorithms [3].

Algorithms	Overall performance accuracy
SVM [67]	77.32%
SAEEED [66]	90.27%
A-ConvNet [31]	98.09%
M-Net [3]	98.90%

F. SM-CNN

With the increase in the application of SAR image scene matching technology in airplane navigation guidance, various works have been proposed in the literature [68]–[70]. The work in [4], hereafter referred to as SM-CNN (Scene Matching using CNN), is one of the most recent in this field and is the first to use CNN for classifying the suitability of areas for scene matching in SAR images into suitable or unsuitable. Considering the characteristics of SAR, scene matching is challenging because it may be affected by topographic variances [71]. Also, a single feature is not enough to analyze the similarity between two scenes. On the contrary, a combination of several indicators as a feature descriptor could not properly reflect the matching suitability and results in redundancies between features [70]. In SM-CNN, a reference image and a candidate image were taken over the same region at different angles were selected and are matched using cross-correlation coefficient and matching error with respect to certain thresholds, generating the training labels for a set of images. Two Digital Elevation Model (DEM) data of the images are then formed and appended on the SAR

image, forming a three-channel image to provide the elevation information that includes undulating terrain features, which helps determine the matching suitability. Input images of size 228 × 228 are extracted from the DEM included SAR image and fed the CNN along with the respective generated labels. Figure 8 shows the overall flow of the SM-CNN [4] method and Table 8 shows the configuration of its CNN. The CNN used by SM-CNN [4] for classification is the same as the fully convolutional network model proposed in A-ConvNet [31]. However, the model was first pretrained on large networks like CaffeNet [72]. The SM-CNN has achieved an improved classification accuracy compared to SVM. The author in SM-CNN also verifies the performance of pretrained models such as VggNet [73] and CaffeNet, whose results are shown in Table 7. It was observed that A-ConvNet [31] has the lowest number of parameters and thus is preferable when time and storage are a concern. On the other hand, CaffeNet outperforms the other models in terms of accuracy. The SM-CNN [4] however, suffers the misclassification issue, especially in regions where there are high-rise buildings. This is because the DEM cannot differentiate between high-rise and low-rise buildings. The Digital Surface Model (DSM) provides the 3D representation of the terrain surface. Therefore, including DSM in the input data might reduce the misclassification results and improve classification accuracy.

TABLE 7. Table showing performances of pretrained models tested in SM-CNN [4].

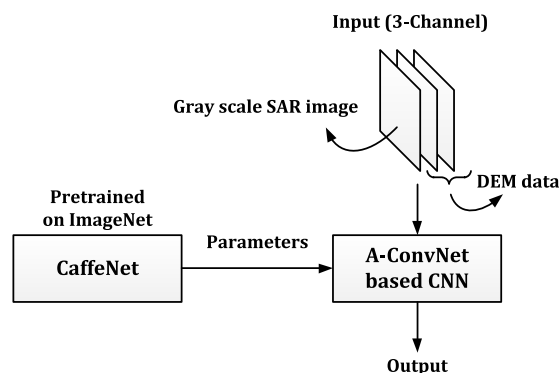| Metric	CaffeNet	VggNet	A-ConvNet
Accuracy	93.3%	92.5%	91.1%
No. of Parameters	227.5M	537.1M	4.8M

FIGURE 8. Flowchart summarizing the technique in SM-CNN [4].

G. AN-CNN

In the CNN paradigm, the neighboring pixels in an image are not given much consideration, and this may be one reason why image classification models face the misclassification issue. To address this issue, the work in [74], referred to as AN-CNN (Adaptive Neighbourhood based Convolutional

TABLE 8. Hyperparameters used in SM-CNN [4].

Hyperparameters	Values
Mini batch size	50
LR initial	0.001
LR final	0.0001
Weight decay	0.002
Momentum	0.3

Neural Network) introduced an adaptive CNN based on neighboring pixels for SAR image classification. AN-CNN uses the bilateral spatial and feature-based distance from the central pixel to adapt weights to the neighboring pixels. The feature-based weighting was done to improve the classification of boundary regions, whereas the spatial-based weighting was done to minimize the misclassification error. The architecture of AN-CNN is shown in Figure 9. The AN-CNN architecture is comprised of two convolutional layers, each followed by pooling layers. It also has one fully connected layer before the final classification layer. The adaptive neighborhood of the input image was first generated and behaved as input to the convolutional layers. The CNN was trained using a customized cost function. The AN-CNN model was tested on real SAR images, namely the San Francisco Bay dataset and the Flevoland dataset and achieved an overall accuracy of 83.90% and 87.13%, respectively. The AN-CNN achieved better performance in boundary as well as homogeneous regions due to the importance it gave to neighboring pixels by incorporating both spatial and feature distance-based pixel weighting mechanisms. However, when a certain percentage of labeled samples were used as training data, the AN-CNN became less effective on real SAR data when compared to traditional CNN. This is because a limited training sample makes the model generate less discriminating features.
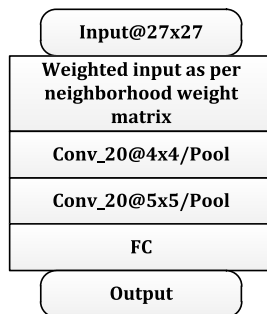
Input@27x27
Weighted input as per neighborhood weight matrix
Conv_20@4x4/Pool
Conv_20@5x5/Pool
FC
Output

FIGURE 9. Architecture used in AN-CNN [74].

H. SSR-TC

Since the application of deep learning in SAR images requires sufficient data and since processed SAR data are limited, the work in [75] hereafter known as SSR-TC (Sample Spectral Regularization based Target Classification) aims at classifying SAR targets by regularizing the singular values of each feature associated with the SAR image giving rise to better distinguishable features. The regularization method is done by lessening the variation between small and large singular values of features. This way, the performance improvement for classification using CNN with limited data is also achieved. It may be mentioned that SSR-TC uses transfer learning by first training the CNN on substantial simulated SAR data and fine-tuning the network on real SAR data. SSR-TC outperforms other recent works [76]–[79].

I. RCC-MRF

The RCC-MFR (Region Category Confidence degree-based Markov Random Field) [80] is another recent approach that uses deep learning for the classification of SAR images. The RCC-MFR was proposed by emphasizing the spatial constraints between the super-pixel regions in SAR images. It claimed that the super-pixel regions might improve SAR image classification performance when considered. Therefore, the RCC-MRF uses the detailed features and the necessary constraints between super-pixels in their algorithm. The role of CNN comes into play during the production of region labels. The CNN used for region label generation consists of only two convolutional layers and a fully connected layer. Since misclassification is a common drawback faced by most classification models, the RCC-MRF uses an energy function that clubbed together unary and binary energy to rectify the misclassification of regions and enable region predictions based on their categories. The unary energy function uses the RCC term based on probability distributions over entire pixels, which helps the RCC-MRF make improved predictions of regions concerning various categories. The RCC-MRF achieved an overall accuracy of 88.85% on the Radarsat-2 San Francisco Bay and 89.56% on the Radarsat-2 Flevoland image. However, for regions that the CNN misclassifies, the RCC-MRF loses its effect, especially for regions that are adjacent to each other.

J. MFFN-CPMN

The Deep Feature Fusion and Covariance Pooling Manifold Network (MFFN-CPMN) [81] was recently proposed in the literature for classification of SAR images by combining the merits of both local and global features. The MFFN-CPMN also takes advantage of multiple feature fusion in their network. The MFFN was first designed using Gabor filtering to retrieve crucial spatial information and relevant deep features. The MFFN is comprised of CNN, whereby the weight optimization is done using the unsupervised dual-sparse encoder. Next, a CPMN was designed to retrieve the global statistical information using the fused features, which are finally used to distinguish between various classes associated with a SAR scene using a softmax classifier. The flowchart of MFFN-CPMN is shown in Figure 10. The CNN used in MFFN was trained in a greedy unsupervised learning manner by excavating the hidden spatial information and high-level global information with the help of different filters. Therefore MFFN becomes an effective model for retrieving the most

relevant features with respect to high-resolution SAR data without using deeper networks. The MFFN-CPMN could attain an overall accuracy of 89.33% on TerraSAR-X SAR image, 90.03% on GF-3 SAR image, 88.37% on Airborne SAR image and 96.61% on F-SAR image.
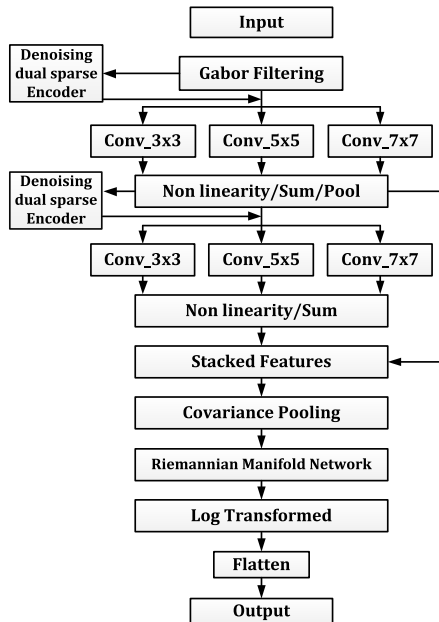

FIGURE 10. Flowchart of MFFN-CPMN [81].

K. CNNE-ML

The work in [82], hereafter referred to as CNNE-ML incorporated two stages in their network for classification of ships and military trucks using the OpenSARShip data and MSTAR data, respectively. In the first stage, a standard CNN was trained for the classification of SAR targets. Features are then extracted and flattened from the trained CNN. The flattened feature vector is then fed to a non-linear classifier containing three fully connected layers followed by a softmax classifier. In the second stage, a metric network is designed mainly for clustering of the feature vectors in the respective feature space. This is achieved with the help of prototypes calculated based on feature vector moderation in every class. The distance between the sample query and support sample is calculated in order to predict the class it belongs to. Classification results show that the incorporation of metric learning in the target classification approach helps improve the performance by attaining accuracy of 99.79% on the MSTAR dataset and 83.67% on the OpenSARShip dataset.

L. OSL-HSN

The OSL-HSN (One-Shot Learning-based classification using Hybrid Siamese Network) [83] uses an innovative approach for classification of SAR targets by exploiting deep learning and aims at determining information from only a limited amount of training samples. The OSL-HSN uses

a Siamese network whose training is based on the triplet approach, in the sense that it is trained using three types of images: random image (anchor), image from a common class (positive), and an image from a different class (negative). The Siamese network produces a single vector of a certain length known as embeddings as the output. The embeddings of all three types of images are then compared using a distance measure in order to check their similarity. The network is trained so that it possesses minimum distance between the anchor and positive and maximum distance between the anchor and negative. While training the network, each iteration undergoes two stages. The triplet selection is made in the first stage, followed by network training using the selected triplets in the second stage. It may be mentioned that the image used in this work was pre-processed such that it forms a 3-channeled image with the original image as the first channel, Lee filtered image as the second channel and inverted filtered image as the third channel. The architecture of the OSL-HSN is shown in Figure 11. The OSL-HSN also incorporated feature fusion in their model. The overall architecture is lightweight and attained a classification accuracy superior to other classifiers [73], [84]–[86] on the OpenSAR Ship dataset.
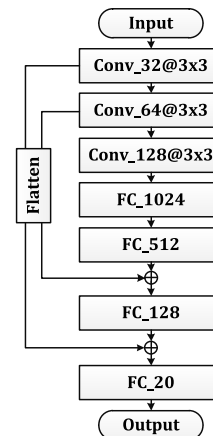

FIGURE 11. Architecture used in OSL-HSN [83].

M. OBSERVATIONS

This section briefly discusses the observations that can be concluded from various SAR image classification works. Firstly, as observed from SI-CNN [32], few convolutional layers caused the model to learn fewer features, especially on targets with similar characteristics, thereby causing misclassification. On the other hand, as seen from DNN-DAE-Conv [48], higher-level features can be learned with the help of stack auto-encoder and from less-labeled data since labeled SAR data is not readily available. The idea of collaborating two architectures in DNN-DAE-Conv has also helped produce high-level feature-based labeled data for training CNN. Similar to DNN-DAE-Conv, the MNet [3] could also learn more features but from less data, meaning

that only seed data that are labeled are used. But the MNet requires an information recorder that records every data the model learns for subsequent comparisons. Also, the model suffers from a few misclassifications that may be improved when an optimized approach is adopted for comparing the data with that from the information recorder to produce the output. While MNet works poorly under EOC conditions of the MSTAR targets, the A-ConvNet [31] showed improvement in the performance under EOC conditions. The A-ConvNet has omitted the fully connected layers in its model and simultaneously mitigates the over-fitting problem. However, A-ConvNet causes few false alarms, and the anti-noise performance is also low. Influenced by the model of A-ConvNet, the SM-CNN [4] have also omitted the fully connected layers in its model for scene matching recognition. The exclusion of fully connected layers has also reduced the number of parameters in the model. Since the SM-CNN model is meant to classify matching and non-matching regions, the DEM data clubbing in the inputs has helped classify scene matching regions except for scenes with high-rise buildings. This issue can be solved by clubbing DSM data along with DEM as inputs. Apart from scene-matching, the classification of most confused targets in SAR images has also improved when inputs are channeled with different image resolutions as observed from CNN-MR [47], although few misclassifications still exist. It is also observed that most of the recent works like AN-CNN, RCC-MRF and MFFN-CPMN have concentrated on extracting spatial information in their approach. This shows that emphasizing spatial features is a recent trend as this seems to benefit the classification performances in SAR images. Also, the blending of multiple feature types plays a role in bringing up the performance. This is because a model that learns multiple feature information has better representation capability. It may also be noted that different filter sizes reduce the parameters of a model while simultaneously enabling the model to learn distinct features associated with a particular input. Despite all the aforementioned innovative approaches adopted by recent methods, the misclassification issue persists. Therefore, it is noticed that misclassification is still a major concern in most cases. Hence, incorporating the ideas from each method discussed above in order to form a hybridized model can be explored and tuned to improve the SAR image classification performance in the future. A comparison summary of the different classification methods discussed in this review paper is shown in Table 10. At the same time, Table 9 summarizes the average classification results of the different methods.

IV. RESEARCH ISSUES AND CHALLENGES
One of the major issues associated with the classification of SAR images, be it a scene or target classification, is the misclassification issue, which means a part or entire image is being predicted incorrectly by the model or the algorithm. Despite having several beneficial characteristics, such issues hesitate the application of SAR images in real-life applications. The misclassification is caused mainly due to the high

TABLE 9. Average classification results of different methods.

Method	Dataset	Average Results
CNN-MR [47]	TERRASAR-X	$Avg.F1-score : 0.94$
SM-CNN [4]	TERRASAR-X	$Avg.Accuracy : 93.3\%$
MFFN-CPMN [81]	TERRASAR-X	$Avg.Accuracy : 89.33\%$
All-in-one CNN [58]	MSTAR	$Avg.F1-score : 0.82$
A-ConvNet [31]	MSTAR	$Avg.F1-score : 0.87$
CNN-A [59]	MSTAR	$Avg.F1-score : 0.89$
M-Net [3]	MSTAR	$Avg.Accuracy : 98.9\%$
AN-CNN [74]	San Francisco Bay	$Avg.Accuracy : 83.90\%$
	Flevoland	$Avg.Accuracy : 87.13\%$
RCC-MRF [80]	San Francisco Bay	$Avg.Accuracy : 88.85\%$
	Flevoland	$Avg.Accuracy : 89.56\%$
CNNE-ML [82]	OpenSAR Ship	$Avg.Accuracy : 83.67\%$
OSL-HSN [83]	OpenSAR Ship	$Avg.F1-score : 97.9$
SI-CNN [32]	SENTINEL-1 SCANSAR	$Avg.Accuracy : 78.65\%$

complexity and coarseness of the images obtained from SAR, along with the presence of unwanted features called speckle noise. The features present in SAR images are very hard to distinguish in the sense that clutter and targets are hard to discriminate even by the human eye because of their resemblance. Therefore identifying which target or image scene falls under which class is complex and hence challenging. Even though substantial noise suppression techniques have been brought forward to remove noise from SAR images, another issue arises during the process. The noise removal techniques tend to remove even some of the important features that may have contributed to the proper prediction by the model. This again gives rise to improper classification of SAR images. The noise filtering technique also results in the blurriness of the images and the occurrence of artifacts. These issues lead to improper subsequent processing of SAR images such as detection, segmentation and even classification. The advancement of deep learning has caught the attention of researchers working in SAR image analysis, and another issue was faced in the application of deep learning to SAR analysis. The unavailability of sufficient labeled SAR data resulting in overfitting issues poses an immense challenge in using deep learning models for SAR image analysis [75], [87]. Data augmentation methods to precisely enlarge the SAR data to enable the smooth application of deep learning in SAR image classification are still under research. Classification of similar-looking targets is yet another challenge since targets with similar features are usually misinterpreted, especially in the case of SAR images with sophisticated backgrounds.

V. POTENTIAL MODELS AND FUTURE APPROACH
It is observed from this study that each SAR image classification method follows a unique approach and has advantages and disadvantages associated with them, which we have highlighted in this study in order to benefit the research

TABLE 10. Summarizing the concept of different SAR target classification networks.

| Paper	Type	Dataset	Preprocessing	Input Datatype	Architecture	Results	Advantage	Disadvantage	Future
DNN-DAE-Conv- [48]	Multiclass classification	Oceanographic SAR images	Target detection using CFAR	ROI chips of empirical size	2Conv +2Maxpool +3FC	-	Learns only higher level features due to stack auto encoder	Requires manually identified targets and labels from AIS	Denoising of input images before detecting targets + clubbing DNN-DAE with more CNN layers in a hybridized manner may improve network performance
A-ConvNet- [31]	Multiclass classification	MSTAR	Data augmentation + Target detection	88x88 input image	Conv16@5x5/ ReLU/ Maxpool +Conv32@5x5/ ReLU/ Maxpool +Conv64@6x6/ ReLU/ Maxpool +Dropout +Conv128@5x5/ ReLU +Conv10@3x3/ Softmax	98%	Outperforms [52], [54]–[57], [67] specially in EOC conditions	Few false alarms, antinoise performance is low	Preprocessing stage for denoising may improve performance and may be adaptable to noisy images
CNN-MR- [47]	Multiclass classification	TerraSAR-X	Augmentation + Target Detection using CFAR + binary masking + segmentation	128x128 ROI chips (3 channels each with resolutions 3m, 12m & 24m respectively)	Conv16@5x5/ Maxpool +Conv32@5x5/ Maxpool +Conv64@6x6/ Maxpool +Conv128@8x8/ Maxpool +Dense +sigmoid +softmax	F1-score obtained is 0.94 and outperforms SVM+PCA, DNN-DAE-Conv [48], A-ConvNet [31], CNN-A [59]	Most confused targets have achieved improved classification accuracy	Results in few Mis-classification	Use of target focusing and target super resolution in the pre-processing step can potentially improve classification result.
SI-CNN- [32]	Multiclass classification	SENTINEL-1 SCANSAR	Radiometric correction + Thermal Noise removal + Lee Filtering	32x32 3-channel image chips	Conv64@5x5/ Pool/ Norm +Conv128@3x3/ Pool/ norm +Conv256@3x3/ Pool/ norm+ FC+ Dropout+ FC+ Dropout +Softmax	Obtained precision 0.945	Few classes attained accuracy more than 96%	Classifying class layers for training is manual	Various window sizes for finding optimal feature scale for feature description can be explored
M-Net- [3]	Multiclass Classification	MSTAR	Pretraining	70x70 input size	Conv16@7x7/ ReLU/ Pool +Conv32@5x5/ ReLU/ Pool +Conv64@5x5/ ReLU/ pool +Conv128@3x3/ ReLU +Dropout +Conv10@3x3	Outperforms SVM [67], SAEED [66], A-ConvNet with average accuracy of 98.9%	Uses very limited SAR labels without suffering from overfitting, can extract more relevant features from less training sample	Model is unstable and has slow convergence, misclassification occurs	Information recorder can be worked upon to output probabilistic values.
SM-CNN- [4]	Binary classification	TERRASAR-X	DEM data of the images are formed and appended to become 3-channels	228X228 3-channel inputs	Conv16@5x5/ BN/ReLU/ Pool +Conv32@3x3/ BN/ReLU/ Pool +Conv64@6x6/ BN/ReLU/ Pool +Conv128@6x6/ BN/ReLU/ Pool +Conv256@5x5/ BN/ReLU/ Pool +Conv2@3x3 +Softmax	Accuracy better than SVM [67]	Less parameters since there is no fc layer	Suffers misclassification in regions with high rise buildings	Inclusion of Digital Surface Model (DSM) in the input data might reduce misclassification
SSR-TC- [75]	Multiclass classification	MSTAR	SAR images are simulated from CAD models	128X128	Conv16@5x5/ BN/ReLU/ Pool +Conv32@5x5/ BN/ReLU/ Pool +Conv64@3x3/ BN/ReLU/ Pool +Conv64@3x3/ BN/ReLU/ Pool +Fc +Softmax	-	Feasible method for limited samples like SAR	Few misclassification still exists	-
AN-CNN [74]	Multiclass classification	RadarSat-2 San Francisco Bay, RadarSat-2 Flevoland		27x27 image patch	Conv20@4x4/Pooling +Conv20@5x5/Pooling +Fc	Overall accuracy of 83.90% on San Francisco and 87.13% on Flevoland dataset	Achieved good performance in boundary as well as homogeneous regions	Less effective when less sample is used for training	The model can be optimized to make it effective with less training samples
RCC-MRF [80]	Multiclass classification	RadarSat-2 San Francisco Bay, RadarSat-2 Flevoland		27x27 image patch	Conv/Pooling +Conv/Pooling +Fc	Overall accuracy of 88.85% on San Francisco and 89.56% on Flevoland dataset	Classification performance improved due to emphasis on spatial constraints	Performance highly dependent on the region classification	
CNNE-ML [82]	Multiclass classification	OpenSARship dataset +MSTAR		100x100 input chips	Conv16@5x5/BN/ ReLU/MaxPool +Conv32@5x5/BN/ ReLU/MaxPool +Conv64@3x3/BN/ ReLU/MaxPool +Conv128@3x3/BN/ ReLU/MaxPool +Flatten+Fc1024/ReLU +Fc128/ReLU +Fc3/ReLU +Softmax	Accuracy of 99.76% on MSTAR data and 83.67% on OpenSARship data	Performs well even with less amount of training sample	Misclassification issue persists	-
MFFN-CPMN [81]	Multiclass classification	TerraSAR-X, GF-3, Airborne SAR, and F-SAR	-	-	-	Overall accuracy of 89.33% on TerraSAR-X, 90.03% on GF-3, 88.37% on Airborne SAR, and 96.61% on F-SAR	Capable to retrieve most relevant features without using deeper networks	Not end-to-end model and misclassification exists	Data generation techniques such as GANs can be applied and the model can be made end-to-end
OSL-HSN [83]	Multiclass classification	OpenSARship dataset	3-Channeled input image generation	40x40 image	Conv32@3x3 +Conv64@3x3 +Conv128@3x3 +Fc1024 +Fc512 +Fc128 +Fc20	F1-score of 97.9	Lightweight and uses limited training samples	-	-

*Conv = Convolution, Fc = Fully Connected
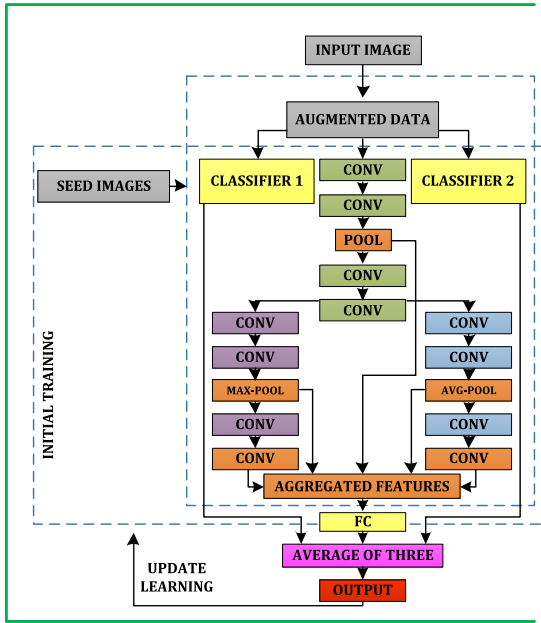
FIGURE 12. A potential future model for SAR image classification based on existing works.

FIGURE 13. A potential future model for SAR image classification using recent networks.

community who are experimenting in the field of SAR image analysis for use in the future. Efficient networks such as VggNet [73], GoogleNet, EfficientNet [88], DenseNet [89] are rarely being explored for the purpose of classification in SAR images. The aforementioned networks may be adopted in the future in the form of pretrained networks or parameter adoption and may result in better performances since these networks outperform the classical networks for optical images in terms of classification accuracy [35]. Based on the observations made from this study, a hybrid architecture can be implemented in the future for SAR image classification. Inspired by MNet [3], the hybrid model can be trained using a small number of labeled images as a startup and then eventually increased using update learning wherein an encoder-decoder based network can be used since encoder-decoder model gives better feature representations [34]. The recent trend on feature aggregation is encouraged in the hybrid model since aggregating features implies more information learned by the model [78]. Therefore the features from the encoder-decoder model can be clubbed with those from one or more complementary classifiers to obtain better feature learning which ultimately may improve the classification performance in SAR images. It may be mentioned that assistant classifiers also help in better decision-making by suggesting the most probable result [87]. Therefore assistant classifiers can also be added to the network. The pictorial representation of the hybrid model discussed can be seen in Figure 12. Apart from the potential hybrid model for classification shown in Figure 12 discussed previously, another classification architecture can be implemented firstly using multi-resolution inputs as inspired from CNN-MR [47] to enable detailed feature learning. Deeper models can be used adopting architectures such as EfficientNet [88], auto-encoder, DenseNet [89]
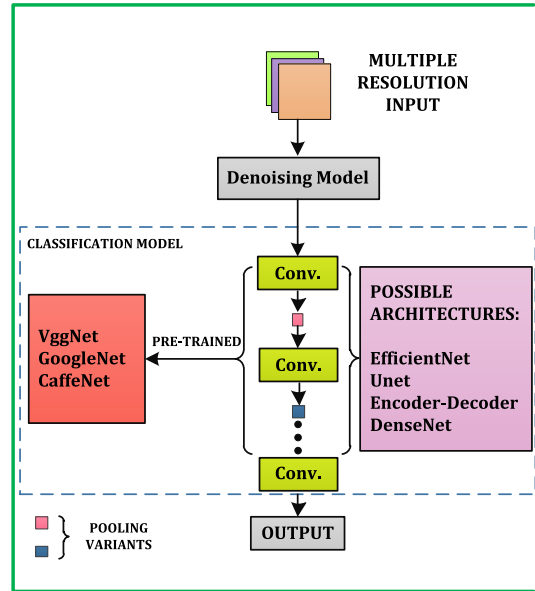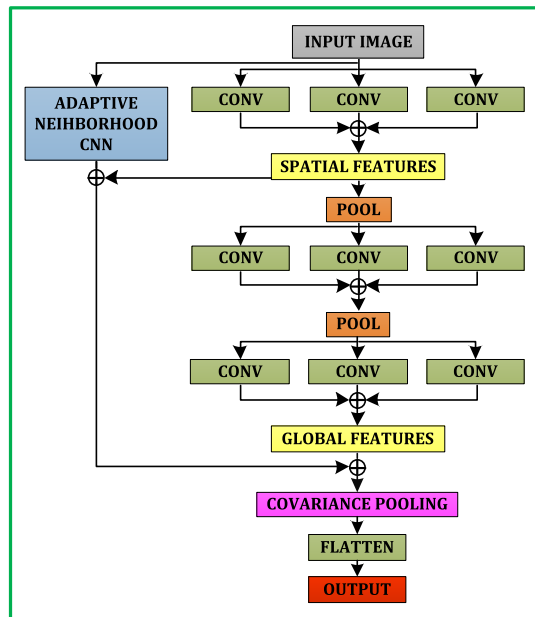

FIGURE 14. A potential future model for SAR image classification emphasizing spatial information.

or UNet [90]. It may be mentioned that the majority of the models used in the literature do not have a prior denoising model attached to them for noise reduction. Therefore a hybrid model can be proposed in the future that clubbed together denoising model followed by classification so that the network is adaptable to noisy SAR images. Multiple variants of the pooling layer, such as average and max-pooling, are also encouraged to be used simultaneously in the model as this may result in major feature learning by the model. The model is depicted in Figure 13. Emphasizing spatial information of the SAR images, another hybrid model as shown in Figure 14 can be experimented in the future
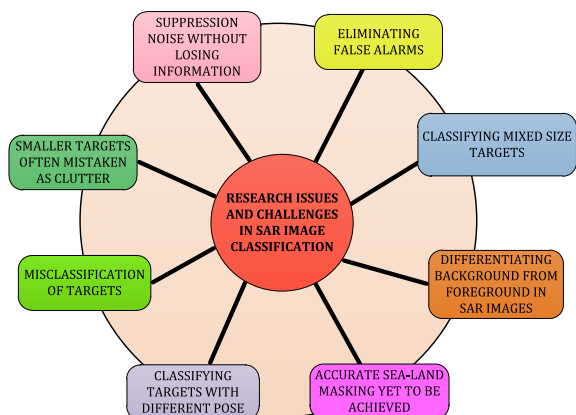
FIGURE 15. Existing research issues concerning SAR image interpretations.

for classification purposes. Inspired by MFFN-CPMN, the model blends together global and local features by considering spatial information. This will enable the model to learn the most discriminating features, which would help improve classification performance. It may be mentioned that the spatial information can be extracted using the first convolutional layer combined with adaptive neighborhood CNN from the AN-CNN method, as the adaptive neighborhood strategy gives good information for boundary and homogeneous regions. On the other hand, global features can be obtained by deeper layers convolved by multiple filter sizes. The combined features are then passed through covariance pooling to deeply mine the potential features from the combined features and finally get the desired output. The three hybrid models discussed here will be implemented as future works wherein results will be evaluated accordingly.

VI. CONCLUSION

Since classification is the ultimate processing step for SAR image interpretation in many real-world applications, and because deep learning has shown advancements in various computer vision efforts, we studied the various state-of-the-art methods using the deep learning approach for the classification of SAR images. This study also highlighted the architectures involved in each method and their configurations and parameter settings. It is observed from this study that there lies several issues in processing and interpreting SAR images, and these issues have also been pointed out in this work and are briefly represented in Figure 15. However, the major issue associated with SAR image classification is misclassification. This issue still needs to be addressed since misinterpretation of targets may lead to misinformation in real-world scenarios. Based on the study, the advantages and disadvantages of each work have also been discussed, followed by the future directions to ease researchers in adopting several characteristics of the existing approach for applications in several other processing fields related to SAR interpretations. Potential models for possible application in the future have also been developed based on the study.

REFERENCES

[1] S. K. Chaturvedi, "Study of synthetic aperture radar and automatic identification system for ship target detection," *J. Ocean Eng. Sci.*, vol. 4, no. 2, pp. 173–182, Jun. 2019.

[2] P. Wang, H. Zhang, and V. M. Patel, "SAR image despeckling using a convolutional neural network," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1763–1767, Dec. 2017.

[3] R. Shang, J. Wang, L. Jiao, R. Stolkin, B. Hou, and Y. Li, "SAR targets classification based on deep memory convolution neural networks and transfer parameters," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 8, pp. 2834–2846, Aug. 2018.

[4] C. Zhong, X. Mu, X. He, B. Zhan, and B. Niu, "Classification for SAR scene matching areas based on convolutional neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 9, pp. 1377–1381, Sep. 2018.

[5] J. Pei, Y. Huang, W. Huo, Y. Zhang, J. Yang, and T.-S. Yeo, "SAR automatic target recognition based on multiview deep learning framework," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 4, pp. 2196–2210, Apr. 2018.

[6] S. Liu, T. Liu, L. Gao, H. Li, Q. Hu, J. Zhao, and C. Wang, "Convolutional neural network and guided filtering for SAR image denoising," *Remote Sens.*, vol. 11, no. 6, p. 702, Mar. 2019.

[7] T. Zeng, H. K.-H. So, and E. Y. Lam, "Computational image speckle suppression using block matching and machine learning," *Appl. Opt.*, vol. 58, no. 7, p. B39, 2019.

[8] Y. Huang, G. Liao, Z. Zhang, Y. Xiang, J. Li, and A. Nehorai, "SAR automatic target recognition using joint low-rank and sparse multiview denoising," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1570–1574, Oct. 2018.

[9] Y. Xiang, F. Wang, and H. You, "OS-SIFT: A robust SIFT-like algorithm for high-resolution optical-to-SAR image registration in suburban areas," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 8, pp. 3078–3090, Jun. 2018.

[10] F. Biondi, "A polarimetric extension of low-rank plus sparse decomposition and radon transform for ship wake detection in synthetic aperture radar images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 1, pp. 75–79, Jan. 2019.

[11] G. Mercier and F. Girard-Ardhuin, "Oil slick detection by SAR imagery using support vector machines," in *Proc. Eur. Oceans*, Jun. 2005, pp. 90–95.

[12] A. H. S. Solberg, "Remote sensing of ocean oil-spill pollution," *Proc. IEEE*, vol. 100, no. 10, pp. 2931–2945, Oct. 2012.

[13] K. Tomiyasu, "Tutorial review of synthetic-aperture radar (SAR) with applications to imaging of the ocean surface," *Proc. IEEE*, vol. 66, no. 5, pp. 563–583, May 1978.

[14] F. J. Meyer, J. B. Nicoll, and A. P. Doulgeris, "Correction and characterization of radio frequency interference signatures in L-band synthetic aperture radar data," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 10, pp. 4961–4972, Apr. 2013.

[15] K. El-Darymli, P. McGuire, D. Power, and C. Moloneyb, "Target detection in synthetic aperture radar imagery: A state-of-the-art survey," *J. Appl. Remote Sens.*, vol. 7, no. 1, Mar. 2013, Art. no. 071598.

[16] K. El-Darymli, E. W. Gill, P. Mcguire, D. Power, and C. Moloney, "Automatic target recognition in synthetic aperture radar imagery: A state-of-the-art review," *IEEE Access*, vol. 4, pp. 6014–6058, 2016.

[17] X. X. Zhu, D. Tuia, L. Mou, G.-S. Xia, L. Zhang, F. Xu, and F. Fraundorfer, "Deep learning in remote sensing: A comprehensive review and list of resources," *IEEE Geosci. Remote Sens. Mag.*, vol. 5, no. 4, pp. 8–36, Dec. 2017.

[18] C. Yao, X. Luo, Y. Zhao, W. Zeng, and X. Chen, "A review on image classification of remote sensing using deep learning," in *Proc. 3rd IEEE Int. Conf. Comput. Commun. (ICCC)*, Dec. 2017, pp. 1947–1955.

[19] H. Wang, F. Xu, and Y.-Q. Jin, "A review of polsar image classification: From polarimetry to deep learning," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 3189–3192.

[20] H. Parikh, S. Patel, and V. Patel, "Classification of SAR and PolSAR images using deep learning: A review," *Int. J. Image Data Fusion*, vol. 11, no. 1, pp. 1–32, Jan. 2020.

[21] Y. K. Chan and V. C. Koo, "An introduction to synthetic aperture radar (SAR)," *Progr. Electromagn. Res. B*, vol. 2, pp. 27–60, 2008.

[22] C. Wolff. (2019). *Radar Tutorials*. Accessed: Aug. 25, 2019. [Online]. Available: https://www.radartutorial.eu/html/author.en.html

[23] A. Passah, K. Amitab, and D. Kandar, "SAR image despeckling using deep CNN," *IET Image Process.*, vol. 15, no. 6, pp. 1285–1297, May 2021.

[24] J. Wang, T. Zheng, P. Lei, and X. Bai, "Ground target classification in noisy SAR images using convolutional neural networks," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 11, pp. 4180–4192, Nov. 2018.

[25] J. Brownlee, "Deep learning and artificial neural networks," Tech. Rep., 2016. Accessed: Aug. 27, 2019. [Online]. Available: https://machinelearningmastery.com/what-is-deep-learning/

[26] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai, and T. Chen, "Recent advances in convolutional neural networks," *Pattern Recognit.*, vol. 77, pp. 354–377, May 2018.

[27] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, Feb. 2015.

[28] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[29] G. Li, J. Sun, Y. Song, J. Qu, Z. Zhu, and M. R. Khosravi, "Real-time classification of brain tumors in MRI images with a convolutional operator-based hidden Markov model," *J. Real-Time Image Process.*, vol. 18, no. 4, pp. 1207–1219, 2021.

[30] Ü. Atila, M. Uçar, K. Akyol, and E. Uçar, "Plant leaf disease classification using EfficientNet deep learning model," *Ecolog. Informat.*, vol. 61, Mar. 2021, Art. no. 101182.

[31] S. Chen, H. Wang, F. Xu, and Y. Q. Jin, "Target classification using the deep convolutional networks for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 8, pp. 4806–4817, Aug. 2016.

[32] C. Wang, H. Zhang, Y. Wang, and B. Zhang, "Sea ice classification with convolutional neural networks using Sentinel-L scansar images," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2018, pp. 7125–7128.

[33] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 779–788.

[34] H. Furukawa, "Deep learning for end-to-end automatic target recognition from synthetic aperture radar imagery," 2018, *arXiv:1801.08558*.

[35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, C. Alexander Berg, and L. Fei-Fei, "ImageNet large scale visual recognition challenge," *Int. J. Comput. Vis.*, vol. 115, no. 3, pp. 211–252, Dec. 2015.

[36] D. H. Hubel and T. N. Wiesel, "Receptive fields and functional architecture of monkey striate cortex," *J. Physiol.*, vol. 195, no. 1, pp. 215–243, 1968.

[37] K. Fukushima, "A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biol. Cybern.*, vol. 36, no. 4, pp. 193–202, Apr. 1980.

[38] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.

[39] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. Eur. Conf. Comput. Vis.*, Cham, Switzerland: Springer, 2014, pp. 818–833.

[40] C. Szegedy, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proc. IEEE Conf. CVPR*, Jun. 2015, pp. 1–9.

[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2015, *arXiv:1409.1556*.

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Feb. 2016, pp. 770–778.

[43] J. Yuan, H.-C. Xiong, Y. Xiao, W. Guan, M. Wang, R. Hong, and Z.-Y. Li, "Gated cnn: Integrating multi-scale feature layers for object detection," *Pattern Recognit.*, vol. 105, pp. 107–131, 2019.

[44] Y. Liu and L. Wu, "Geological disaster recognition on optical remote sensing images using deep learning," *Proc. Comput. Sci.*, vol. 91, pp. 566–575, Jan. 2016.

[45] Y. Xu, B. Zhao, Y. Zhai, Q. Chen, and Y. Zhou, "Maize diseases identification method based on multi-scale convolutional global pooling neural network," *IEEE Access*, vol. 9, pp. 27759–27970, 2021.

[46] S. Lawrence, C. L. Giles, A. C. Tsoi, and A. D. Back, "Face recognition: A convolutional neural-network approach," *IEEE Trans. Neural Netw.*, vol. 8, no. 1, pp. 98–113, Jan. 1997.

[47] C. Bentes, D. Velotto, and B. Tings, "Ship classification in TerraSAR-X images with convolutional neural networks," *IEEE J. Ocean. Eng.*, vol. 43, no. 1, pp. 258–266, Jan. 2018.

[48] C. Bentes, D. Velotto, and S. Lehner, "Target classification in oceanographic SAR images with deep neural networks: Architecture and initial results," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2015, pp. 3703–3706.

[49] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, and P.-A. Manzagol, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371—3408, 2010.

[50] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[51] M. Ranzato, C. Poultney, S. Chopra, and Y. LeCun, "Efficient learning of sparse representations with an energy-based model," in *Proc. 19th Int. Conf. Neural Inf. Process. Syst. (NIPS)*, Cambridge, MA, USA: MIT Press, 2006, p. 1137—1144.

[52] R. Singh and B. V. Kumar, "Performance of the extended maximum average correlation height (EMACH) filter and the polynomial distance classifier correlation filter (PDCCF) for multiclass SAR detection and classification," *Proc. SPIE*, vol. 4727, pp. 265–276, Aug. 2002.

[53] 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.

[54] J. O'Sullivan, M. DeVore, V. Kedia, and M. I. Miller, "SAR ATR performance using a conditionally Gaussian model," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 37, no. 1, pp. 91–108, Jan. 2001.

[55] 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.

[56] U. Srinivas, V. Monga, and R. G. Raj, "SAR automatic target recognition using discriminative graphical models," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 50, no. 1, pp. 591–606, Jan. 2014.

[57] 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.

[58] J. Ding, B. Chen, H. Liu, and M. Huang, "Convolutional neural network with data augmentation for SAR target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 3, pp. 364–368, Mar. 2016.

[59] M. Wilmanski, C. Kreucher, and J. Lauer, "Modern approaches in deep learning for SAR ATR," *Proc. SPIE*, vol. 9843, pp. 195–204, May 2016.

[60] Y. Chen, E. Blasch, H. Chen, T. Qian, and G. Chen, "Experimental feature-based SAR ATR performance evaluation under different operational conditions," *Proc. SPIE*, vol. 6968, pp. 176–187, Apr. 2008.

[61] V. Nair and E. G. Hinton, "Rectified linear units improve restricted Boltzmann machines," in *Proc. 27th Int. Conf. Int. Conf. Mach. Learn.*, 2010, pp. 807–814.

[62] W. Dierking, "Sea ice monitoring by synthetic aperture radar," *Oceanography*, vol. 26, no. 2, pp. 100–111, 2013.

[63] P. Maillard, D. A. Clausi, and H. Deng, "Operational map-guided classification of SAR sea ice imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 43, no. 12, pp. 2940–2951, Dec. 2005.

[64] C. Nilubol, Q. Pham, R. Mersereau, M. Smith, and M. Clements, "Hidden Markov modelling for SAR automatic target recognition," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, vol. 2, May 1998, pp. 1061–1064.

[65] G. Dong, N. Wang, G. Kuang, and H. Qiu, "Sparsity and low-rank dictionary learning for sparse representation of monogenic signal," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 11, no. 1, pp. 141–153, Jan. 2018.

[66] S. Deng, L. Du, C. Li, J. Ding, and H. Liu, "SAR automatic target recognition based on Euclidean distance restricted autoencoder," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 7, pp. 3323–3333, Jul. 2017.

[67] M. A. Hearst, S. T. Dumais, E. Osuna, J. Platt, and B. Scholkopf, "Support vector machines," *IEEE Intell. Syst.*, vol. 13, no. 4, pp. 18–28, Jul./Aug. 1998.

[68] S. L. Y. X. Liang, "Research on matching-area suitability for scene matching aided navigation," *Acta Aeronauticaet Astronautica Sinica*, vol. 31, no. 3, p. 553, 2010.

[69] Z. Yang, Y. Chen, X. Qian, M. Yuan, and E. Gao, "Predicting the suitability for scene matching using SVM," in *Proc. Int. Conf. Audio, Lang. Image Process.*, Jul. 2008, pp. 743–747.

[70] B. Li, J. Gong, and J. Tian, "Selection of matching area in SAR scene-matching-aided navigation based on manifold learning," *Proc. SPIE*, vol. 8002, pp. 306–313, Dec. 2011.

[71] T. Bayer, R. Winter, and G. Schreier, "Terrain influences in SAR backscatter and attempts to their correction," *IEEE Trans. Geosci. Remote Sens.*, vol. 29, no. 3, pp. 451–462, May 1991.

[72] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in *Proc. 22nd ACM Int. Conf. Multimedia*, Nov. 2014, pp. 675–678.

[73] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*.

[74] A. Zhang, X. Yang, L. Jia, J. Ai, and Z. Dong, "SAR image classification using adaptive neighborhood-based convolutional neural network," *Eur. J. Remote Sens.*, vol. 52, no. 1, pp. 178–193, Jan. 2019.

[75] W. Liang, T. Zhang, W. Diao, X. Sun, L. Zhao, K. Fu, and Y. Wu, "SAR target classification based on sample spectral regularization," *Remote Sens.*, vol. 12, no. 21, p. 3628, Nov. 2020.

[76] K. Wang, G. Zhang, and H. Leung, "SAR target recognition based on cross-domain and cross-task transfer learning," *IEEE Access*, vol. 7, pp. 153391–153399, 2019.

[77] K. Wang, G. Zhang, Y. Xu, and H. Leung, "SAR target recognition based on probabilistic meta-learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 18, no. 4, pp. 682–686, Apr. 2021.

[78] J. H. Cho and C. G. Park, "Multiple feature aggregation using convolutional neural networks for SAR image-based automatic target recognition," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 12, pp. 1882–1886, Dec. 2018.

[79] G. Dong and G. Kuang, "SAR target recognition via sparse representation of monogenic signal on Grassmann manifolds," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 3, pp. 1308–1319, Mar. 2016.

[80] A. Zhang, X. Yang, S. Fang, and J. Ai, "Region level SAR image classification using deep features and spatial constraints," *ISPRS J. Photogramm. Remote Sens.*, vol. 163, pp. 36–48, May 2020.

[81] W. Liang, Y. Wu, M. Li, Y. Cao, and X. Hu, "High-resolution SAR image classification using multi-scale deep feature fusion and covariance pooling manifold network," *Remote Sens.*, vol. 13, no. 2, p. 328, Jan. 2021.

[82] Y. Li, X. Li, Q. Sun, and Q. Dong, "SAR image classification using CNN embeddings and metric learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[83] J. A. Raj, S. M. Idicula, and B. Paul, "One-shot learning-based SAR ship classification using new hybrid Siamese network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[84] M. Ma, J. Chen, W. Liu, and W. Yang, "Ship classification and detection based on CNN using GF-3 SAR images," *Remote Sens.*, vol. 10, no. 12, pp. 20–43, 2018.

[85] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*.

[86] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Aug. 2016, pp. 21–37.

[87] Z. Cui, C. Tang, Z. Cao, and S. Dang, "SAR unlabeled target recognition based on updating CNN with assistant decision," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 10, pp. 1585–1589, Oct. 2018.

[88] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 6105–6114.

[89] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.

[90] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015*. Cham, Switzerland: Springer, 2015, pp. 234–241.

SAMARENDRA NATH SUR (Senior Member, IEEE) was born in Hooghly, West Bengal, India, in 1984. He received the B.Sc. degree (Hons.) in physics from the University of Burdwan, in 2007, the M.Sc. degree in electronics science from Jadavpur University, in 2007, the M.Tech. degree in digital electronics and advanced communication from Sikkim Manipal University, in 2012, and the Ph.D. degree from the National Institute of Technology (NIT), Durgapur, in 2019. Since 2008, he has been associated with the Sikkim Manipal Institute of Technology, India, where he is currently an Assistant Professor with the Department of Electronics and Communication Engineering. His current research interests include broadband wireless communication (MIMO and spread spectrum technology), advanced digital signal processing, remote sensing, and radar image/signal processing (soft computing). He has published more than 55 SCI/Scopus indexed international journals and conference papers. He is a member of the IEEE-IoT, IEEE Signal Processing Society, Institution of Engineers (India) (IEI), and International Association of Engineers (IAENG). He was a recipient of the University Medal and Dr. S. C. Mukherjee Memorial Gold centered Silver Medal from Jadavpur University, in 2007. He also serves as a Reviewer for the *International Journal of Electronics* (Taylor and Francis), *IET Communication*, *Ad Hoc Networks* (Elsevier), IEEE TRANSACTIONS ON SIGNAL PROCESSING, *SN Applied Sciences*, *Transactions on Emerging Telecommunications Technologies* (Wiley), and *Telecommunication Systems* (Springer). He has been serving as a Guest Editor in topical collection, *SN Applied Sciences* (Springer), since 2019.

BABUSENA PAUL (Member, IEEE) received the B.Tech. and M.Tech. degrees in radiophysics and electronics from the University of Calcutta, West Bengal, India, in 1999 and 2003, respectively, and the Ph.D. degree from the Department of Electronics and Communication Engineering, IIT Guwahati, India. He was with Philips India Ltd., from 1999 to 2000. From 2000 to 2002, he has worked as a Lecturer with the Electronics and Communication Engineering Department, SMIT, Sikkim, India. He has attended and published more than 60 research papers in international and national conferences, symposiums, and peer-reviewed journals. He has successfully supervised several postgraduate students and postdoctoral research fellows. He joined the University of Johannesburg, in 2010. He has worked as the Head of the Department of Electrical and Electronic Engineering Technology, University of Johannesburg, from April 2015 to March 2018. He has been working as an Associate Professor and the Director of the Institute for Intelligent Systems, University of Johannesburg. He is a Life Member of IETE. He was awarded the IETE Research Fellowship.

ALICIA PASSAH received the B.Sc. degree in computer science (C.S.) from North-Eastern Hill University, Meghalaya, India, and the M.Sc. degree in computer science (C.S.) from Pondicherry University, Puducherry, India. She is currently pursuing the Ph.D. degree in information technology with North-Eastern Hill University. Her research interests include artificial intelligence and SAR image processing and its applications.

DEBDATTA KANDAR was born in Deulia, Kolaghat, Purba Medinipur, West Bengal, India, in 1977. He received the Ph.D. (Engg.) degree from the Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata, West Bengal, India, in 2011, the Postdoctoral degree from the Department of Electrical and Electronics Engineering Technology, University of Johannesburg, South Africa, and the M.Tech. degree in 2006. He has more than 18 years of professional experience. He is currently working as an Associate Professor with the Department of Information Technology, North-Eastern Hill University, Meghalaya, India. His research interests include wireless mobile communication, artificial intelligence, soft computing, and radar operation. He has successfully guided several Ph.D., M.Tech., and B.Tech. students. He has national and international journals, conference proceedings, book chapters, and edited books. For his research work, he was a recipient of the Young Scientist Award from Union Radio Science International (URSI GA-2005) at Vigyan Bhaban, Delhi, India.

• • •