Complex-Valued End-to-End Deep Network With Coherency Preservation for Complex-Valued SAR Data Reconstruction and Classification

Reza Mohammadi Asiyabi, *Member, IEEE*, Mihai Datcu, *Fellow, IEEE*, Andrei Anghel, *Senior Member, IEEE*, and Holger Nies

***Abstract*— Deep learning models have achieved remarkable success in many different fields and attracted many interests. Several researchers attempted to apply deep learning models to synthetic aperture radar (SAR) data processing, but it did not have the same breakthrough as the other fields, including optical remote sensing. SAR data are in complex domain by nature and processing them with real-valued (RV) networks neglects the phase component which conveys important and distinctive information. A complex-valued (CV) end-to-end deep network is developed in this study for the reconstruction and classification of CV-SAR data. Azimuth subaperture decomposition is utilized to incorporate physics-aware attributes of the CV-SAR into the deep model. Moreover, the correlation coefficient amplitude (coherence) of the CV-SAR images depends on the SAR system characteristics and physical properties of the target. This coherency should be considered and preserved in the processing chain of the CV-SAR data. The coherency preservation of the CV deep networks for CV-SAR images, which is mostly neglected in the literature, is evaluated in this study. Furthermore, a large-scale CV-SAR annotated dataset for the evaluation of the CV deep networks is lacking. A semantically annotated CV-SAR dataset from Sentinel-1 single look complex stripmap mode data [S1SLC_CVDL (complex-valued deep learning) dataset] is developed and introduced in this study. The experimental analysis demonstrated the better performance of the developed CV deep network for CV-SAR data classification and reconstruction in comparison with the equivalent RV model and more complicated RV architectures, as well as its coherency preservation and physics-aware capability.**

***Index Terms*— Annotated benchmark dataset, classification, coherency preservation, complex-valued (CV) neural network, deep learning, physics-aware network, reconstruction, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) imagery has been used in many different applications, due to its unique characteristics, day/night and all-weather imagery. Especially in urban areas, SAR images convey rich information and several researchers utilized SAR images for land cover mapping in urban regions, using various algorithms [1], [2], [3].

Deep learning models have achieved admirable results in different fields and raised a lot of interests in the SAR community [4], [5], [6], [7], [8]. However, deep learning models did not have the same breakthrough in SAR as they have achieved in optical remote sensing, and the huge potential of deep learning models for SAR data processing is still unexploited [6]. A number of studies have applied deep learning models to SAR data processing [6], [7], [8], [9], [10], [11]. These networks are developed in the real domain, whereas SAR images are in complex domain by nature. Applying real-valued (RV) models to the complex-valued (CV) data neglects the phase component and only exploits the amplitude information [12].

A few articles have developed CV deep models to utilize both the amplitude and phase information of CV-SAR images. Hänsch and Hellwich [13] used CV multilayer perceptron (MLP) networks to classify polarimetric SAR (PolSAR) data. Different loss functions for CV-MLP are evaluated in this study. Wilmanski et al. [14] proposed the CV-SAR input convolutional neural network (CNN) for automatic target recognition (ATR). In their model, only the input and the first layer were in the complex domain and the rest of the model was ordinary RV-CNN. Nevertheless, they have reported a substantial classification improvement with a set of collected wide-angle SAR images. Later, Zhang et al. [12] proposed the first CV-CNN for CV-SAR image interpretation. All elements of the CV-CNN model are extended into the complex domain, and a CV backpropagation algorithm is derived based on the stochastic gradient descent (SGD) for training the CV-CNN. The performance of the model is tested for classifying benchmark PolSAR datasets and demonstrated a better performance than the RV-CNN with the same degree of freedom. Shang et al. [15] extended the encoding and decoding frameworks into the complex domain and proposed a CV convolutional autoencoder (CV-CAE). The proposed CV-CAE model is utilized for feature extraction and classification of PolSAR images, and spatial pixel-squares refinement method is used to enhance the classification results. Sunaga et al. [16] employed CV-CNN for classification and information discovery in interferometric SAR (InSAR) data.

Manuscript received 24 October 2022; revised 28 December 2022; accepted 8 March 2023. Date of publication 14 April 2023; date of current version 1 May 2023. This work was supported by the European Union's Horizon 2020 Research and Innovation Programme through the Marie Skłodowska-Curie Grant under Grant 860370. *(Corresponding authors: Reza Mohammadi Asiyabi; Mihai Datcu.)*

Reza Mohammadi Asiyabi, Mihai Datcu, and Andrei Anghel are with the Center for Spatial Information (CEOSpaceTech), University POLITEHNICA of Bucharest (UPB), 060042 Bucharest, Romania (e-mail: reza.mohammadi@upb.ro; mihai.datcu@rosa.ro; andrei.anghel@munde.pub.ro).

Holger Nies is with the Center for Sensor Systems (ZESS), University of Siegen, 57068 Siegen, Germany (e-mail: nies@zess.uni-siegen.de).

Digital Object Identifier 10.1109/TGRS.2023.3267185

The proposed model was successful in discovering unidentified lands similar to a prepared sample in InSAR data. Sun et al. [17] developed a CV generative adversarial network (GAN) for PolSAR data interpretation. The proposed CV-GAN outperformed the state-of-the-art models with a limited number of labeled training samples. Furthermore, Zhang et al. [18] proposed a novel CV-CNN, called SAR4LCZ-Net, for local climate zones (LCZ) classification and achieved higher classification accuracy than the RV-network with the same structure. Liu et al. [19] proposed an activation function with a hyperparameter to analyze the influence of the phase information in the complex neural networks. Moreover, a novel transfer learning technique, based on electromagnetic theory and network visualization, is proposed to initialize the weight kernels of the first convolutional layer in CV-CNN. More recently, Zeng et al. [20] proposed a multistream (MS) feature fusion based on the CV operation, called MS CV networks (MS-CVNets), for SAR target recognition. The moving and stationary target acquisition and recognition (MSTAR) dataset is used, and the results are compared with the state-of-the-art RV-based models. Mohammadi Asiyabi et al. [21] carried out a comprehensive comparison between the CV- and RV-CNN for PolSAR data classification and demonstrated the superiority of the CV-CNN. In another recent study, Mohammadi Asiyabi et al. [22] evaluated the significance of the coherency preservation for CV deep networks. Qian et al. [23] proposed the γ-Net to mimic the iterative optimization step in the sparse reconstruction of the CV learned iterative shrinkage thresholding algorithm (CV-LISTA) for super resolving SAR tomographic inversion.

Despite the recent interests and developments in CV deep models for SAR data processing, there are several restricting factors that limited the broad advancement of these models. In this study, some of these restricting factors are addressed. Scarcity of training dataset with semantic annotation is a limiting factor for many remote sensing applications. Deep models usually require a large number of training samples, and most of the available SAR datasets are in real domain. It is difficult to find a semantically annotated CV-SAR dataset, which is large enough to train a CV deep model. As a result, most of the studies in this field have used coherency or covariance matrices of PolSAR images, consisting of three CV and three RV feature channels (i.e., not the original amplitude and phase components of SAR data) as the input of the network. In order to tackle this problem, a CV dataset from Sentinel-1 (S1) single look complex (SLC) stripmap (SM) mode products is generated and semantically annotated in this study [S1SLC_CVDL (complex-valued deep learning) dataset]. More detailed explanation of the S1SLC_CVDL dataset is provided in Section III.

Another limiting factor for the development of CV networks is the implementation challenges. Most of the well-known machine learning frameworks for neural network implementation (e.g., PyTorch and TensorFlow libraries in Python environment) are designed for RV networks, and there are only a few operators that support CV data. As a result, different operators which are necessary for deep models are converted into the complex domain. Comprehensively, detailed explanations of the CV operators are provided in Section II to make the results of the study reproduceable. The CV operators later utilized to define the deep models in this study.

CV-SAR data convey physical properties of the backscattered target, and there is a particular structure and correlation between the amplitude and phase components of the CV-SAR data. Processing them separately will not preserve this structure, and well-defined CV operators are required to preserve the complex correlation coefficient (coherency) of the CV-SAR data during the processing steps. The coherency of CV-SAR data depends on the SAR system characteristics, as well as the physical properties of the target [24]. Most of the previous studies in this field have neglected the coherency preservation attribute of the CV networks. The coherency preservation of the CV-SAR data is also assessed in this study.

Furthermore, the synthetic aperture technology provides a higher azimuth resolution through the integration of the backscattered signal from different angles along the flight path of the platform. Azimuth subaperture decomposition technique provides the opportunity to obtain images of the targets observed by different azimuth angles in the SAR footprint [25]. Subaperture decomposition helps to extract physical properties of the target, and using them in the network will incorporate the physics-aware attributes into the deep model [25].

In this study, various CV operators are defined for CV deep models' implementation. An end-to-end architecture, consisting of a CV-CAE and a CV-CNN, is designed. The generated S1SLC_CVDL annotated dataset is used as the input of the CV end-to-end deep model, and the SAR patches are reconstructed and classified into the semantic classes. The performance of the model is evaluated for classification accuracy and coherency preservation. The main contributions of this study are as follows.

1) The coherency preservation of the developed CV model is evaluated and demonstrated that the developed CV network learned and preserved the data model and the CV-SAR data coherency.
2) The physics-aware network with subaperture images learned the image formation and preserved the synthetic aperture data model.
3) The S1SLC_CVDL dataset is generated using the S1 SLC SM products, and the patches are semantically annotated for training the CV models.
4) The performance of the developed CV end-to-end deep network is compared with the equivalent RV network as well as the more complicated state-of-the-art architectures.

The rest of this article is organized as follows. The architecture of the CV end-to-end deep network and the mathematical background of the CV operators are explained in Section II. Section III summarizes the generated S1SLC_CVDL dataset details. The implementation details, experimental results

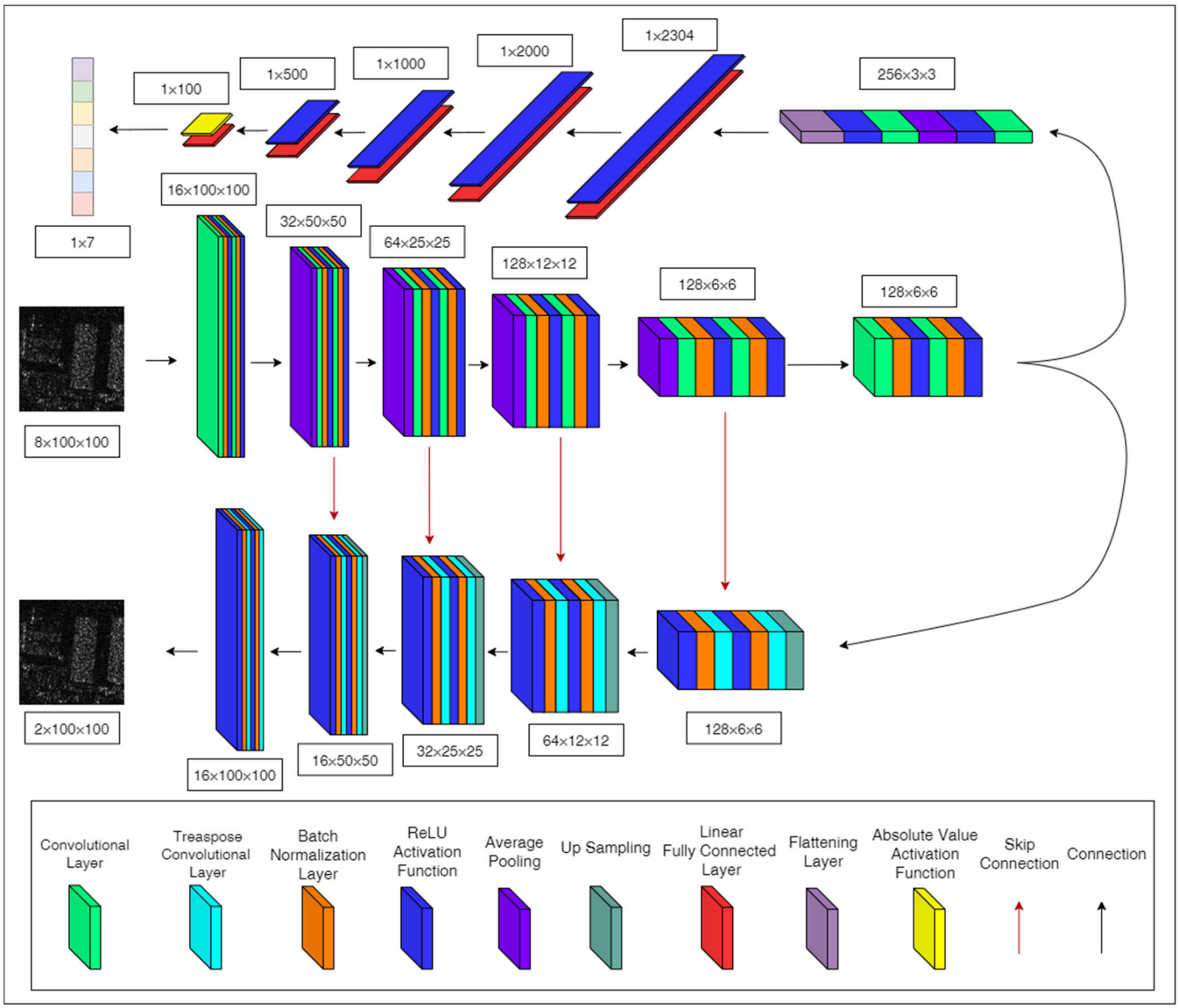


Fig. 1. Architecture of the proposed end-to-end deep network for CV and RV models.

for different case studies, and discussions are provided in Section IV, and finally, Section V concludes this study.

II. Mathematical Background

In this section, the architecture of the deep end-to-end network and the mathematical background for the transformation between the RV and CV operators are presented.

A. Architecture of the Deep End-to-End Network

The common practice to process SAR data is using the amplitude information in the RV networks and neglecting the phase information. In order to compare the conventional approach and the proposed CV approach, the same architecture is used for RV and CV networks in this study. Obviously, the input data, network parameters, and operators are in the complex domain for the CV network and in the real domain for the RV network. An end-to-end deep architecture, consisting of a CAE for feature extraction and a CNN for classification, is designed. Fig. 1 illustrates the architecture of the end-to-end deep network. A number of the input channels are different for each case study (Section IV), and eight input channels are shown as an example in Fig. 1 which includes the complex SAR images in two polarizations and the subaperture images (three from each polarization).

The architecture of the developed network consists of three main parts, encoding, decoding, and classification. In the encoding part, several convolutional layers with batch normalization, average pooling, and rectified linear unit (ReLU) activation functions are used to encode the input SAR image. The resulting bottleneck features from the encoding part of the network are used in the decoding part to reconstruct the SAR image. Several transpose convolutions with batch normalization, up-sampling, and ReLU activation functions are utilized in the decoding part of the architecture. Additionally, four skip connections are used in the autoencoder part of the

end-to-end network to help with the degradation problem of deep networks. Both polarization channels as the input SAR image are reconstructed in the decoding part. Furthermore, the bottleneck features (i.e., output of the encoder) are also used in the classification part to be classified in the defined semantic classes. The classification network comprises two convolutional layers with ReLU activation functions and one average pooling layer. Later, four linear fully connected layers are applied to the flattened vector, and finally, the output is produced using a linear fully connected layer with absolute-valued activation function.

For the convolutional and transpose convolution layers of the network, the kernel size is set to 3×3, stride is 1, zero padding is 1, and dilation is 1. For the pooling layers, the kernel size is 2×2, stride is the same as the kernel size, and dilation is 1 and without zero padding. And finally, for the up-sampling layers in the decoder, the bilinear up-sampling with the scale factor of 2 is used. Further explanations about each layer are provided in Section II-B.

B. CV Operators

When the CV data occur naturally in a system, such as SAR imagery systems, the real and imaginary components are statistically correlated. While extending the RV operators into the complex domain, the preservation of this correlation between the real and imaginary components is necessary to properly utilize the CV data and extract the correct physical information from the SAR data [26], [27]. Moreover, the coherency of the CV-SAR data conveys important information about the SAR system and the physical properties of the target [24]. The complex correlation coefficient of the CV-SAR data is the most important measurement for InSAR applications [24]. As a result, a fully CV network with coherency preservation attribute is necessary.

All of the elements of the CV network, including the weight and bias kernels, different layers (e.g., convolutional, fully connected, batch normalization, and pooling layers), and activation functions, are in the complex domain. However, loss function remains in the real domain to prevent empirical problems during the learning process (more explanation is provided in Section II-B6) [21], [28]. Comprehensively, detailed explanations and equations of the various CV operators are included in this article to provide the possibility of reproducing the network and repeating the experiments. The details of the CV elements of the CV network are provided in the following.

1) Convolutional Layer: Convolutional layer is the main building block of any convolutional network, and it is used to recognize the patterns existing in the input image and extract the features. The convolutional layers in the shallow layers of the network extract low-level features and in the deeper layers extract higher level features. In the CV convolutional layer, the CV output feature maps of the layer are computed through the convolution of the CV input feature maps with a set of CV weight kernels and adding the CV bias. Later, a nonlinear CV activation function $f()$ is also applied to the weighted sum of the input feature maps $V_i^{(l+1)}$. In another words [as shown in (1) and (2)], the CV input feature maps $O_k^{(l)} \in \mathbb{C}^{W_1 \times H_1 \times k}$ are convolved with the weight filters $\omega_{ik}^{(l+1)} \in \mathbb{C}^{F \times F \times k \times l}$ and a bias $b_i^{(l+1)} \in \mathbb{C}^l$ is added and passed through a CV activation function $f()$ to obtain the CV output feature maps $O_i^{(l+1)} \in \mathbb{C}^{W_2 \times H_2 \times l}$

$$O_i^{(l+1)} = f\left(V_i^{(l+1)}\right) \tag{1}$$

$$\begin{aligned} V_i^{(l+1)} &= \sum_{k=1}^{K} \omega_{ik}^{(l+1)} * O_k^{(l)} + b_i^{(l+1)} \\ &= \sum_{k=1}^{K} \left(\Re\left(\omega_{ik}^{(l+1)}\right) \cdot \Re\left(O_k^l\right) - \Im\left(\omega_{ik}^{(l+1)}\right) \cdot \Im\left(O_k^l\right)\right) \\ &\quad + j \sum_{k=1}^{K} \left(\Re\left(\omega_{ik}^{(l+1)}\right) \cdot \Im\left(O_k^l\right) - \Im\left(\omega_{ik}^{(l+1)}\right) \cdot \Re\left(O_{ik}^l\right)\right) \\ &\quad + b_i^{(l+1)}. \end{aligned} \tag{2}$$

In these equations, $\mathbb{C}$ denotes the complex domain, $j = \sqrt{-1}$ is the imaginary unit of the complex numbers, $\Re$ and $\Im$ represent the real and imaginary components of the complex number, respectively, and the character $*$ denotes the convolution operation.

In the convolutional layer, if the input consists of k feature maps with size $W_1 \times H_1$, the output will have l feature maps with size $W_2 \times H_2$, according to the following equations:

$$W_2 = \frac{W_1 - F + 2P}{S} + 1 \tag{3}$$

$$H_2 = \frac{H_1 - F + 2P}{S} + 1 \tag{4}$$

where F, S, and P are the filter size, stride, and zero-padding size, respectively. As a result, $F \times F \times K \times l$ weights and l biases have to be learned in each convolutional layer during the training stage of the network.

2) Pooling Layer: Pooling layer computes a statistical summary over a local patch in the feature map to reduce the dimension of the output feature maps. Moreover, pooling layer makes the model more invariant to the small changes in the position and orientation of the patterns and instead puts more emphasis on the feature itself.

Max pooling and average pooling which consider the maximum and average of the local patch, respectively, are the most commonly used pooling functions in deep learning. Maximum operation is not straightforward in the complex domain, and as a result, max pooling is not readily available for CV networks. However, the magnitude or norm of the complex number is used in some studies for max pooling [16], [19], [29].

The extension of average pooling into the complex domain is more straightforward and can be defined as follows:

$$O_i^{(l+1)}(x, y) = \operatorname*{avg}_{u,v=0,\ldots,g-1} O_i^{(l)}(x \cdot s + u, y \cdot s + v). \tag{5}$$

3) Batch Normalization Layer: Batch normalization is an intermediate layer in neural networks for normalizing the feature maps across a minibatch. Several studies have shown that the batch normalization can provide faster convergence for the model and alleviate the vanishing and exploding gradient problem [30], [31], [32], [33]. In the real domain, in order

to standardize the feature map, they are scaled such that their mean is 0 and variance is 1. However, this cannot be applied in complex domain as it will not guarantee the equal variance for real and imaginary components and the resulting distribution might not be circular [31].

A suitable extension of batch normalization into the complex domain is proposed in [31], as indicated in the following equation:

$$\mathrm{BN}(\tilde{x}) = \gamma\tilde{x} + \beta. \tag{6}$$

In this equation, $\tilde{x}$ is the scaled feature map which is calculated through multiplying the zero-centered data by the inverse square root of the covariance matrix (see the following equation):

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

where the covariance matrix is

$$V = \begin{pmatrix} V_{rr} & V_{ri} \\ V_{ir} & V_{ii} \end{pmatrix} = \begin{pmatrix} \mathrm{COV}(\Re\{x\}, \Re\{x\}) & \mathrm{COV}(\Re\{x\}, \Im\{x\}) \\ \mathrm{COV}(\Im\{x\}, \Re\{x\}) & \mathrm{COV}(\Im\{x\}, \Im\{x\}) \end{pmatrix}. \tag{8}$$

The scaled feature map, $\tilde{x}$, has normal complex distribution with mean $\mu = 0$, covariance $\Gamma = 1$, and pseudo-covariance $C = 0$, as can be calculated with the following equations:

$$\mu = E[\tilde{x}] \tag{9}$$

$$\Gamma = E\left[(\tilde{x} - \mu)(\tilde{x} - \mu)^*\right] = V_{rr} + V_{ii} + j(V_{ir} - V_{ri}) \tag{10}$$

$$C = E\left[(\tilde{x} - \mu)(\tilde{x} - \mu)\right] = V_{rr} - V_{ii} + j(V_{ir} + V_{ri}). \tag{11}$$

Moreover, β and γ in (6) are the shift and scaling parameters of the batch normalization, respectively. The shift parameter β is a learnable complex number. The scaling parameter γ is a learnable 2×2 positive semidefinite matrix with three degrees of freedom ($\gamma_{ri} = \gamma_{ir}$), as shown in the following equation:

$$\gamma = \begin{pmatrix} \gamma_{rr} & \gamma_{ri} \\ \gamma_{ir} & \gamma_{ii} \end{pmatrix}. \tag{12}$$

4) Fully Connected Layer: In the fully connected layers, each neuron in the layer is connected to all of the neurons in the next layer. The CV output feature maps $O_l^{(l+1)}$ are computed by the multiplication of the CV input feature maps $O_k^{(l)}$ by the CV weight kernels $\omega_{ik}^{(l+1)}$ and adding the CV biases $b_i^{(l+1)}$ and applying a CV activation function $f()$, as shown in the following equations:

$$O_l^{(l+1)} = f\left(V_i^{(l+1)}\right) \tag{13}$$

$$V_i^{(l+1)} = \sum_{k=1}^{K} \omega_{ik}^{(l+1)} \cdot O_k^{(l)} + b_i^{(l+1)}. \tag{14}$$

5) Activation Function: Activation functions are usually used to add nonlinearity to the network and to enable more complex mappings between the inputs and outputs of the network. It used to be believed that the nonlinearity of the activation functions has serious problems in the complex domain for the differentiability of the functions in the network. Many researches restricted themselves to use only the complex differentiable or holomorphic activation functions (linear activation function) [34]. Later, it has been proved that if the function is differentiable with respect to the real and imaginary components of each parameter, it can also be used in the CV neural networks [31], [34].

In general, two main approaches have been followed in the literature for extending the activation functions into the complex domain. The first approach is the real-imaginary-type activation functions, which applies the activation function on the real and imaginary components, separately [shown in (15)]. The second approach is the amplitude–phase-type activation functions which saturates the amplitude while preserving the phase [see (16)] [26], [35], [36], [37], [38], [39]

$$f(z) = f(\Re(z)) + jf(\Im(z)) \tag{15}$$

$$f(z) = f(|z|)e^{j\theta_z} \tag{16}$$

where $f()$ is the activation function, and θ_z is the phase component of the complex number z.

Numerous activation functions have been proposed and utilized in deep networks, including Sigmoid, Tanh, ReLU, and Softmax functions. Real-imaginary-type CV extension is the most frequently used approach in the literature. In this section, the CV ReLU and Sigmoid will be derived and presented.

ReLU in real domain is computed as shown in the following equation. ReLU activation function is computationally efficient and accelerates the convergence of the gradient descent toward the global minimum

$$\mathrm{ReLU}(x) = \max(0, x). \tag{17}$$

A number of CV ReLU have been proposed for CV networks. The most frequently used complex extension of ReLU is the Complex ReLU ($\mathbb{C}$ReLU) which applies the RV ReLU on the real and imaginary components, separately, as shown in the following equation:

$$\mathbb{C}\mathrm{ReLU}(z) = \mathrm{ReLU}(\Re(z)) + j\mathrm{ReLU}(\Im(z)). \tag{18}$$

Another CV extension of ReLU is zReLU which is proposed in [40] and is calculated as follows:

$$\mathrm{zReLU}(z) = \begin{cases} z, & \text{if } \theta_z \in \left[0, \dfrac{\pi}{2}\right] \\ 0, & \text{Otherwise} \end{cases} \tag{19}$$

where θ_z is the phase of the complex number z.

An amplitude–phase CV extension of the ReLU function is modReLU which is proposed in [41] and is defined as follows:

$$\begin{aligned} \mathrm{modReLU}(z) &= \mathrm{ReLU}(|z| + b)e^{i\theta_z} \\ &= \begin{cases} (|z| + b)\dfrac{z}{|z|}, & \text{if } |z| + b \geq 0 \\ 0, & \text{Otherwise.} \end{cases} \end{aligned} \tag{20}$$

In this equation, z is a complex number with the phase θ_z and b is a learnable real parameter. The Complex ReLU ($\mathbb{C}$ReLU) is proved to be the most suitable extension of the ReLU in complex domain and is used in this study [18], [28], [31], [33], [42].

Sigmoid as a special form of the logistic function is another frequently used activation function in deep learning. Sigmoid function is an S-shaped curve with a range between 0 and 1, and as a result, it is used for the networks with probability

output (see the following equation where $\sigma()$ is the sigmoid function and x is a real number):

$$\sigma(x) = \frac{1}{1+e^{-x}}. \tag{21}$$

A straightforward extension of the sigmoid function in the complex domain is the real-imaginary-type activations function [32] which is shown in the following equation:

$$\sigma(z) = \sigma(\Re(z)) + j\sigma(\Im(z)). \tag{22}$$

Moreover, an amplitude–phase-type extension of the sigmoid function is called "Siglog" [32] which is shown in the following equation:

$$\text{Siglog}(z) = \sigma(\log|z|)e^{i\theta_z}. \tag{23}$$

Absolute-value activation function is used to transform the output of the layer to the real domain. Absolute-value activation function simply computes the absolute value of the output as shown in the following equation:

$$\text{abs}(z) = \sqrt{\Re(z)^2 + \Im(z)^2}. \tag{24}$$

Similarly, other activation functions such as tanh and Softmax can be extended to the complex domain.

6) Loss Function: In neural networks, loss function (also called cost function and criteria in the literature) is a measure that evaluates how well the model has predicted the output in comparison to the desired target. Several different loss functions are introduced and utilized in various deep learning models for different applications, including mean squared error (MSE or L2), mean absolute error (MAE or L1), Huber loss, Log–Cosh loss, Quantile loss, Cross-Entropy (Log) loss, and Hinge loss.

In the complex domain, maximum and minimum are not clearly defined. However, one can use the magnitude of the complex number for comparison, but only the magnitude is not a sufficient comparison factor for complex numbers. As a result, the minimization of the complex numbers and, consequently, complex functions are not clearly defined [28], [32]. In order to prevent empirical risks during the learning process, the loss function of the CV model will remain in the real domain. The input and the output of the network are in complex domain, and using the magnitude value to compute the network error is not sufficient. As a result, the RV loss is computed for the real and imaginary components separately, and the final loss is the average value between them [12], [15], [19].

As described in Section II-A, the architecture of the model is composed of two main parts, reconstruction and classification parts. And the model has two outputs, the reconstructed CV SAR patch and the classification semantic label. The loss function has to evaluate both of these outputs, by which the model can learn to improve both of them.

For the reconstruction part, the MSE loss function is used. As proved in several articles, MSE loss is a suitable evaluation method for autoencoder networks and reconstruction tasks [15], [32], [33]. MSE loss $\mathcal{L}_{\text{reconstruction}}$ is computed as shown in the following equation where n is the number of the samples, and X_i and Y_i are the output of the model and the desired target, respectively:

$$\mathcal{L}_{\text{reconstruction}} = \frac{1}{2}\left(\left(\frac{1}{n}\sum_{i=1}^{n}(\Re(X_i) - \Re(Y_i))^2\right) + \left(\frac{1}{n}\sum_{i=1}^{n}(\Im(X_i) - \Im(Y_i))^2\right)\right). \tag{25}$$

For the classification part, Cross-Entropy loss is used. Cross-Entropy loss $\mathcal{L}_{\text{binary CrossEntopy}}$ is the sum of the negative logarithm of the predicted output probabilities of the model and can be calculated as shown in (26) for binary classification problems. In (26), λ_i and X_i are the label (1 for positive and 0 for negative) and the predicted probability of the sample being in the same class, respectively. However, for multiclass classification, a Softmax algorithm is used to normalize the scores for the given classes. Accordingly, the Cross-Entropy loss $\mathcal{L}_{\text{classification}}$ for the multiclass problem can be calculated as (27), where X_{λ_n} is the probability output for the target semantic label and X_{C_i} is the probability output for each semantic label C_i. Cross-Entropy is the most frequently used loss function for deep learning classification problems [18], [28], [33], [43]

$$\mathcal{L}_{\text{binary CrossEntropy}} = -\sum_{i=1}^{n}\lambda_i \log(X_i) \tag{26}$$

$$\mathcal{L}_{\text{classification}} = -\frac{1}{n}\log\frac{e^{X_{\lambda_n}}}{\sum_{i=1}^{C} X_{C_i}}. \tag{27}$$

Finally, the total loss $\mathcal{L}$ of the model is calculated through the weighted sum of the reconstruction and classification losses, as shown in the following equation, where α is the regularization term and is set to 0.5 in this study:

$$\mathcal{L} = \mathcal{L}_{\text{classification}} + (1-\alpha)\mathcal{L}_{\text{reconstruction}}. \tag{28}$$

7) Backpropagation: Backpropagation is an SGD-based method that is widely used for the training deep learning models. The backpropagation algorithm minimizes the total error (loss) of the training samples in the training stage of the network by adjusting the weight and bias parameters of the network. For RV networks, the adjusted weight ω and bias b of the lth layer in the tth iteration with $\mathcal{L}[t]$ loss can be calculated according to the following equations, respectively, where η is the learning rate:

$$\begin{aligned}\omega_{ik}^{(l)}[t+1] &= \omega_{ik}^{(l)}[t] - \eta\Delta\omega_{ik}^{(l)}[t] \\ &= \omega_{ik}^{(l)}[t] - \eta\partial\mathcal{L}[t]/\partial\omega_{ik}^{(l)}[t]\end{aligned} \tag{29}$$

$$\begin{aligned}b_{i}^{(l)}[t+1] &= b_{i}^{(l)}[t] - \eta\Delta b_{i}^{(l)}[t] \\ &= b_{i}^{(l)}[t] - \eta\partial\mathcal{L}[t]/\partial b_{i}^{(l)}[t].\end{aligned} \tag{30}$$

According to Liouville's theorem in complex analysis, a bounded holomorphic function must be constant on the entire complex distribution [28]. In other words, the loss and the activation functions of the CV model must be constant or unbounded. However, Wirtinger calculus [44] extended the complex derivation to nonholomorphic functions. Several articles utilized Wirtinger calculus to derive the CV backpropagation algorithm [12], [13], [19], [26], [28], [32], [39], [42], [45], [46]. A simplified version of the

CV backpropagation based on the Wirtinger calculus, tailored for the case study of this article, is provided in the following.

According to the Wirtinger calculus, if z is a complex variable, $z = x + jy \in C, (x, y) \in R^2$, the partial derivatives of a complex function $f(z)$ with respect to z and $\bar{z}$ are shown in the following equation:

$$\frac{\partial f}{\partial z}\frac{1}{2}\left(\frac{\partial f}{\partial x} - j\frac{\partial f}{\partial y}\right), \quad \frac{\partial f}{\partial \bar{z}}\frac{1}{2}\left(\frac{\partial f}{\partial x} + j\frac{\partial f}{\partial y}\right). \tag{31}$$

Later, the complex gradient can be defined as follows:

$$\nabla_z f = 2\frac{\partial f}{\partial \bar{z}}. \tag{32}$$

Accordingly, the correction term for the weights of the lth layer $\Delta\omega_{ik}^{(l)}[t]$ can be computed as follows:

$$\Delta\omega_{ik}^{(l)} = \frac{\partial \mathcal{L}}{\partial \Re\left(\omega_{ik}^{(l)}\right)} + j\frac{\partial \mathcal{L}}{\partial}\Im\left(\omega_{ik}^{(l)}\right). \tag{33}$$

Loss $\mathcal{L}$ is not directly related to the weights $\omega_{ik}^{(l)}$ of the model and as a result, in order to compute the correction term for the weights in each layer, complex chain rule should be applied [13], [47]

$$\begin{aligned}\Delta\omega_{ik}^{(l)} = &\left(\frac{\partial \mathcal{L}}{\partial \Re\left(V_i^{(l)}\right)}\frac{\partial \Re\left(V_i^{(l)}\right)}{\partial \Re\left(\omega_{ik}^{(l)}\right)} + \frac{\partial \mathcal{L}}{\partial \Im\left(V_i^{(l)}\right)}\frac{\partial \Im\left(V_i^{(l)}\right)}{\partial \Re\left(\omega_{ik}^{(l)}\right)}\right) \\ &+ j\left(\frac{\partial \mathcal{L}}{\partial \Re\left(V_i^{(l)}\right)}\frac{\partial \Re\left(V_i^{(l)}\right)}{\partial \Im\left(\omega_{ik}^{(l)}\right)} + \frac{\partial \mathcal{L}}{\partial \Im\left(V_i^{(l)}\right)}\frac{\partial \Im\left(V_i^{(l)}\right)}{\partial \Im\left(\omega_{ik}^{(l)}\right)}\right).\end{aligned} \tag{34}$$

If the complex error term $\delta_i^{(l)}$ is defined as (35), (34) can be simplified as (36)

$$\delta_i^{(l)} = -\frac{\partial \mathcal{L}}{\partial \Re\left(V_i^{(l)}\right)} - j\frac{\partial \mathcal{L}}{\partial \Im\left(V_i^{(l)}\right)} \tag{35}$$

$$\Delta\omega_{ik}^{(l)} = -\delta_i^{(l)}\overline{O_i^{(l-1)}}. \tag{36}$$

The error term $\delta_i^{(l)}$ has to be calculated in each layer to propagate the error and adjust the parameters of the model to obtain the desired outputs. As shown in the architecture of the model, three main layers are used in this model, including the convolutional, fully connected, and pooling layers. The complex backpropagation algorithm backpropagates the error till the parameters (i.e., weights and biases) of the first layer of the end-to-end deep network and adjusts these parameters concurrently according to the defined loss function.

III. S1SLC_CVDL Annotated Dataset

Scarcity of the annotated remote sensing dataset and the difficulty of creating a high-quality large-scale annotated Earth observation (EO) dataset is an important and restricting factor in the development of various machine learning algorithms. Numerous datasets with diverse characteristics have been developed in the remote sensing literature for different applications, consisting of optical and SAR EO images, including UC-Merced land use [48], EuroSAT [49], OpenSARUrban [1], and MSTAR [50] datasets [51]. However, according to the authors' best knowledge, there is no large-scale high-quality CV-SAR dataset. Although the MSTAR dataset provides CV-SAR data, it is very specific (i.e., military vehicle) and cannot provide a comprehensive evaluation of the classification performance of the deep learning models. Most of the previous articles on SAR data classification with CV networks used the elements of the coherency or covariance matrices of the PolSAR images [12], [13], [15], [17], [33], [52]. The diagonal elements of these matrices are RV and the rest are CV and conjugated at the symmetric position of the main diagonal [15]. Six elements, consisting of three RV and three CV elements, of the upper triangle of these matrices are usually used as the input of the CV networks (i.e., not the original amplitude and phase components of SAR data). Additionally, because of the small size of the available SAR scene for these datasets, a very limited number of small patches can be extracted, which is not suitable for training deep models. Consequently, most of the CV models developed for CV-SAR data classification are shallow models and are not capable of classifying larger scale CV-SAR datasets with larger patch sizes (i.e., low generalizability).

In order to tackle this problem in this study, a large-scale annotated CV-SAR dataset is developed using the SLC SM S1 SAR data, called S1SLC_CVDL dataset. S1 is an SAR imagery mission as a part of the Copernicus program. S1 measures the Earth surface with C-band frequency and up to 5 × 5 m spatial resolution, based on the operation mode [53]. In addition, S1 SAR data are accessible, free of charge, through the Copernicus datahub (https://scihub.copernicus.eu/), and provide a valuable data source for various remote sensing applications.

Three S1 scenes over Chicago and Houston in the United States (US) and Sao Paulo in Brazil are selected to include diverse urban areas as well as vegetation (VG) cover and water (WT) bodies. Table I summarizes the characteristics of the SAR scenes.

The abovementioned SAR scenes are divided into 289 760 nonoverlapping patches of 100 × 100 pixels. The patch size is selected to maintain the semantic integrity and meaningfulness of the patches, based on the semantic classes, as well as being in agreement with the other well-known S1 SAR datasets, such as [1]. No further preprocessing (including multilooking or slant range to ground range conversion) is applied to the SAR patches, and the intensity values of the SAR patches in two polarization channels are used directly to develop the dataset. Only the intensity information is used for dataset development as the existing well-known algorithms are in real domain. Later, the original CV-SAR images are provided in the developed dataset.

Since the semantic annotation of the SAR patches are not available, 1274 patches are annotated manually, using the visual inspection via Google Earth Pro, in three broad semantic classes including VG, constructed (CS), and WT regions. Later, another level of manual semantic annotation is applied to each of the broad classes and seven semantic classes, including agriculture (AG), forest (woodland) (FR),

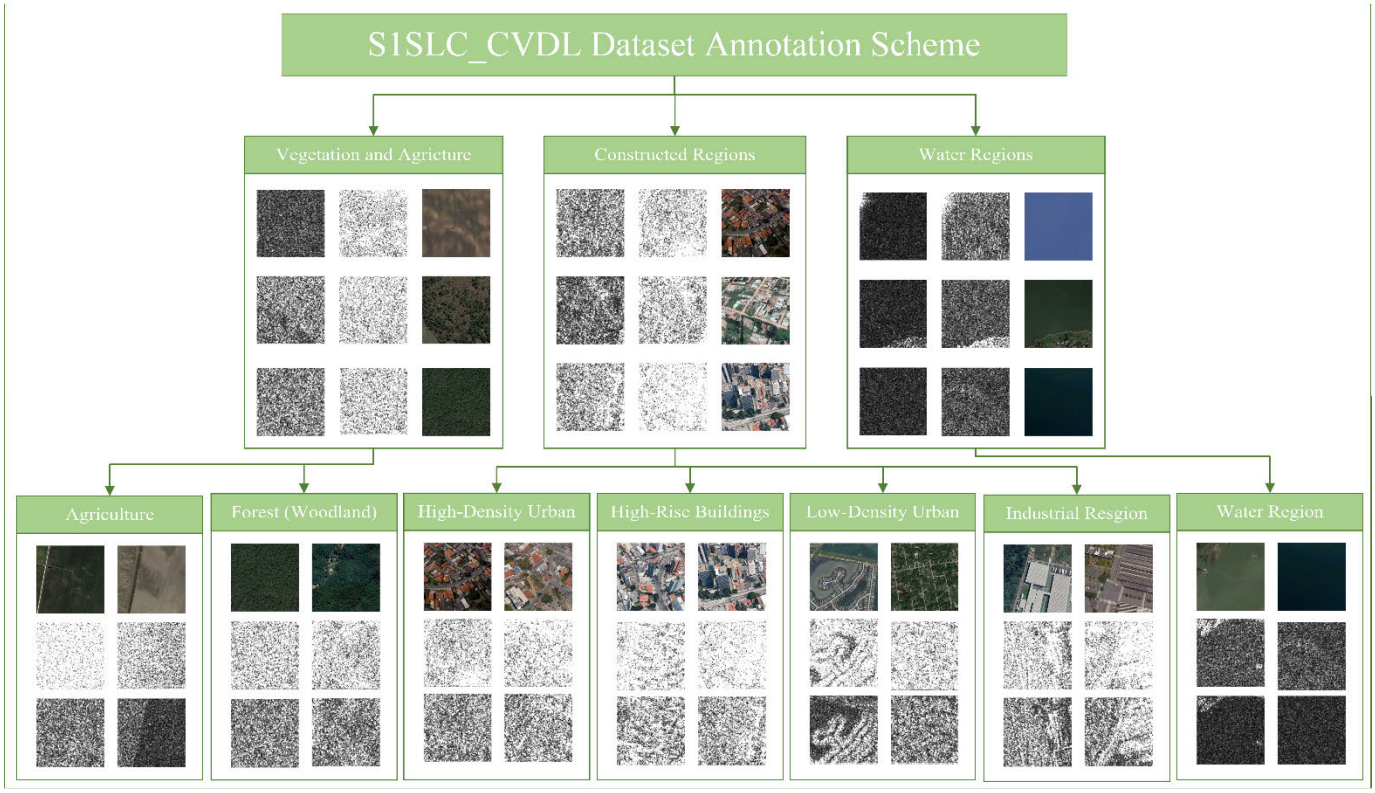


Fig. 2. Example patches of the Google Earth view and the HH- and HV-polarization SAR images for the first- and second-level semantic labels of the developed S1SLC_CVDL dataset.

TABLE I
SAR DATA SPECIFICATIONS IN S1SLC_CVDL DATASET

Feature	Chicago Scene	Houston Scene	Sao Paulo Scene
Product ID	S1A_S1_SLC__1SDH_20210518T121201_037942_047A61_DE5A	S1B_S3_SLC__1SDH_20210505T001823_026762_033263_16AD	S1A_S3_SLC__1SDH_20210516T213615_037918_0479A0_E1C4
Acquisition Date	2021/05/18	2021/05/05	2021/05/16
Orbit Pass	Descending	Ascending	Ascending
Beam	S1	S3	S3
Mission	Sentinel-1A	Sentinel-1B	Sentinel-1A
Acquisition Mode	SM	SM	SM
Product	Level-1 SLC	Level-1 SLC	Level-1 SLC
Polarization	HH, HV	HH, HV	HH, HV
Subset size (pixels)	54900 ×21000	46600×18800	46700×18600

SM = StripMap, SLC = Single Look Complex

high-density urban areas (HD), high-rise buildings (HR), low-density urban areas (LD), industrial regions (IR), and water regions (WR), are produced. Fig. 2 represents the first- and second-level semantic labels with example SAR patches and the corresponding Google Earth RGB image.

After the manual semantic annotation, the mean and standard deviation of the Gabor features in six directions ($1/6\pi$, $1/3\pi$, $1/2\pi$, $2/3\pi$, $5/6\pi$, and $1/\pi$) and four filter sizes (3×3, 7×7, 11×11, and 19×19) are computed for each polarization channel of the SAR patches and used as the low-level features (i.e., 96 low-level features for each patch) [54]. Well-known support vector machine (SVM) classifier with histogram intersection kernel is used to classify the SAR patches. Fig. 3 illustrates the HH polarization and the initially classified SAR scenes.

Obvious misclassifications are apparent in the classified maps. Many patches consist of mixed classes and have ambiguous semantic label, especially on the edges. In order to remove the misclassified patches and improve the quality of the annotation in the dataset, a latent Dirichlet allocation (LDA)-based semantic data mining technique is applied to the initially classified maps. For more details on the LDA-based semantic data mining, refer to [55].

As a result of the semantic data mining, the number of the patches reduced and the size of the dataset decreased by about 5%. However, this procedure enhances the dataset in terms of less incorrectly classified patches, as well as less mixed patches consisting of multiple semantic classes. Fig. 4 represents the classified patches. The patches, which are removed as a result of the semantic data mining, are shown in black. Apparently, the removed patches mostly consist of the mixed and ambiguous patches such as WT patches with ships, shallow WT areas with the effect of the sediment or above WT surface VG, beach and harbor, edge patches, roads, and other mixed patches with ambiguous semantic label, as well as the misclassified patches as a result of the confusion between WT surface and flat agricultural fields, and between various

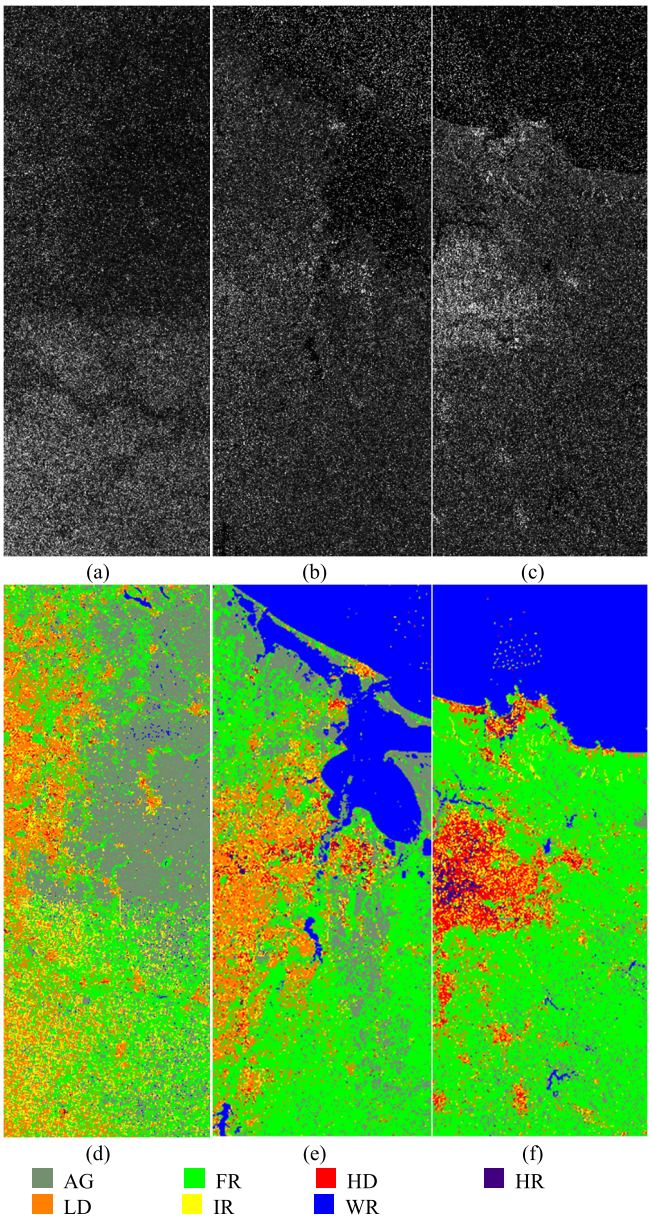


Fig. 3. HH-polarization SAR images and initially classified maps with SVM classifier and Gabor features for: (a) and (d) Chicago; (b) and (e) Houston; and (c) and (f) Sao Paulo scenes.

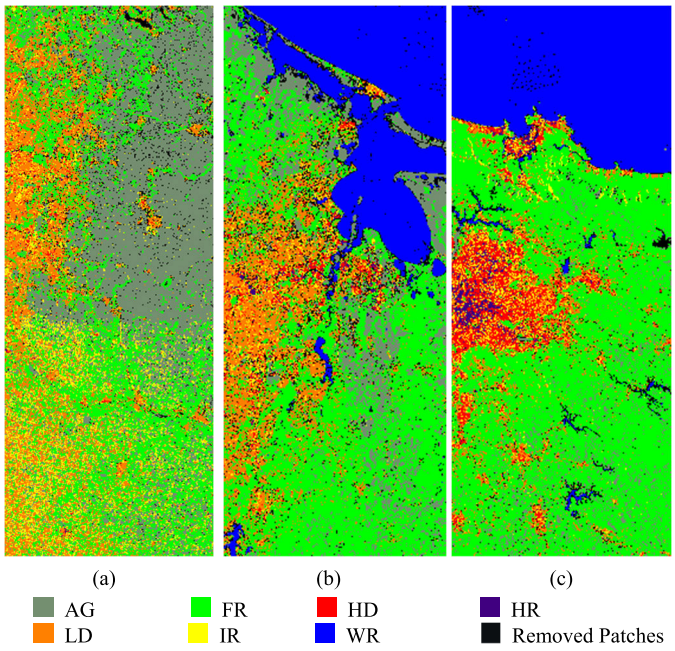


Fig. 4. Classified maps with SVM classifier and Gabor features after the semantic analysis for: (a) Chicago; (b) Houston; and (c) Sao Paulo scenes.

TABLE II
NUMBER OF THE CLASSIFIED PATCHES IN EACH SEMANTIC CLASS OF THE S1SLC_CVDL DATASET

Semantic Class	Before Semantic Datamining	After Semantic Datamining
Agriculture (AG)	61,980	59,327
Forest (Woodland) (FR)	112,192	108,981
High Density Urban Area (HD)	8,875	7,884
High Rise Buildings (HR)	1,745	1,680
Low Density Urban Area (LD)	42,049	40,747
Industrial Regions (IR)	20,713	16,627
Water Regions (WR)	42,206	41,325
Total Number of the Patches	**289,760**	**276,571**

CS areas. Fig. 5 shows an example of the removed patches for each class. Table II shows the number of patches in each semantic class, before and after clearing the dataset with the semantic data mining technique.

In the developed dataset, four CS semantic classes have very similar features, which makes the classification of those classes very challenging. The similarities are also apparent in the LDA-based data mining results in [55]. Many of the CS patches consist of VG cover but the majority of the covered area is considered as the semantic label of the patch. Moreover, despite the interclass similarities, each of the semantic classes in the CS areas has a broad intraclass diversity. For instance, low-density urban regions in Chicago mostly consist of big villas with pool and dense VG cover, while in Houston, LD class mostly represents smaller villas with less dense VG around them and in Sao Paulo is mostly sparse and small CS fields at the middle of the VG areas. The intraclass diversity complicates the classification of the CS classes.

The main purpose of the semantic data mining is to remove the misclassification and decrease the false positive outputs of the classification. The precision measure is increased about 0.02 after the semantic data mining, which indicates the decrease in the false positive in the dataset (i.e., higher quality annotation).

The developed S1SLC_CVDL annotated dataset is available at IEEE DataPort (http://ieee-dataport.org/11016) for research purposes [56]. Higher quality annotation of the dataset will be produced and published in the future studies to obtain more accurate evaluations of the CV deep models. A subset of the developed S1SLC_CVDL dataset is used in Section IV to evaluate the performance of the developed CV end-to-end deep model.

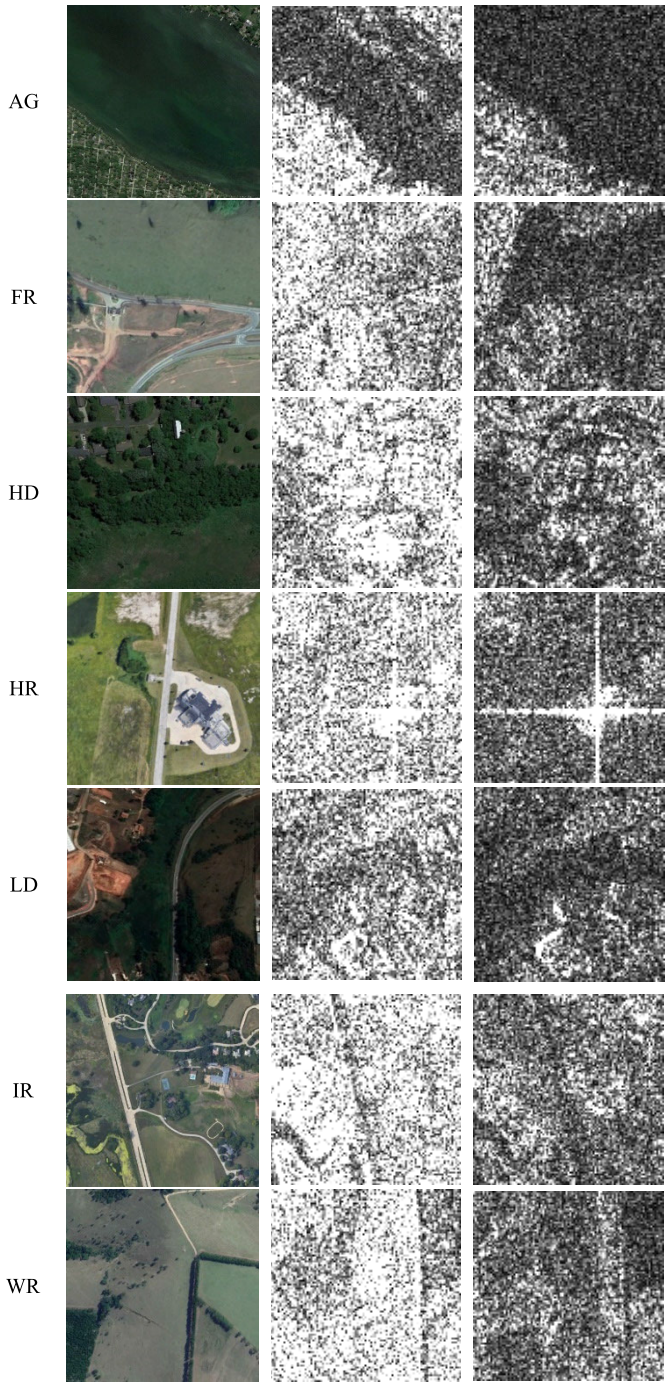


Fig. 5. One example with the Google Earth view, HH polarization, and HV polarization of the SAR patch for the removed patches for each semantic class.

IV. Experimental Results

In this section, the developed CV end-to-end deep model is utilized to classify the annotated S1SLC_CVDL dataset. For this purpose, a randomly selected subset of the S1SLC_CVDL dataset consisting of 30 000 CV-SAR patches is selected. About 66% of the selected subset is used for training the model, and the rest is used for evaluation. The performance of the proposed deep model and the annotated S1SLC_CVDL dataset are evaluated in different case studies.

In the first case study, the CV-SAR patches in two polarization channels are directly fed into the CV end-to-end deep model and the classification accuracies are computed for the first-level (i.e., three semantic classes) and second-level (i.e., seven semantic classes) semantic annotations. In the next case study, three azimuth subaperture images for each polarization channel of the CV-SAR images are computed and used alongside the CV-SAR patches in the CV deep model to evaluate the effect of the physical features extracted by the azimuth subaperture decomposition. The subaperture images contain physical information about the surface and incorporate physics-aware attributes to the deep network. Later, only the amplitude of the SAR and the azimuth subaperture images are used in the equivalent RV deep model to compare the classification performance of the CV and RV deep models. The classification performance of the CV end-to-end deep network is compared with the state-of-the-art classification deep architectures, high-resolution network (HRNet) [57]. Moreover, the reconstructed CV-SAR images are compared with the input CV-SAR images to evaluate the reconstruction quality and coherency preservation ability of the CV deep model. For this purpose, the coherency between the two CV-SAR images is computed. Furthermore, the azimuth subaperture decomposition is applied to the reconstructed CV-SAR images to compare with the azimuth subaperture of the input images. The high coherency between the SAR images and the subaperture images demonstrates the capability of the developed CV end-to-end deep model to learn the data model and preserve the coherency of the CV-SAR data. The abovementioned case studies and the obtained results are discussed in more detail in Sections IV-A–IV-E.

In these experiments, the input SAR and azimuth subaperture images are normalized, using the mean and standard deviation values as suggested in [15] to reduce the risk of the vanishing or exploding gradients. Moreover, with the normalized data, the reconstruction $\mathcal{L}_{\text{reconstruction}}$ and classification $\mathcal{L}_{\text{classification}}$ losses will be in a similar range and ensure the balanced training of both networks, simultaneously. In all of the case studies, the learning rate is 0.001, batch size is 10, and the models are trained for ten epochs.

A. Case Study 1, CV-SAR Classification

In this case study, the input of the network consists of two CV channels. The CV-SAR images in two polarization channels (i.e., HH and HV) are used in the proposed CV deep model for classification with the first- and second-level semantic annotation. Orange color charts in Figs. 6 and 7 demonstrate the overall and class-based classification accuracies obtained from the proposed CV model for this case study. The CV end-to-end deep network obtained about 90.7% overall classification accuracy. The lowest accuracies are obtained in the CS classes where the LD class is classified with only 43.85% accuracy. However, in the three-class scenario, the CS classes are separated with high accuracy and the most important confusion is between different types of the CS semantic classes. Table III illustrates the confusion matrix

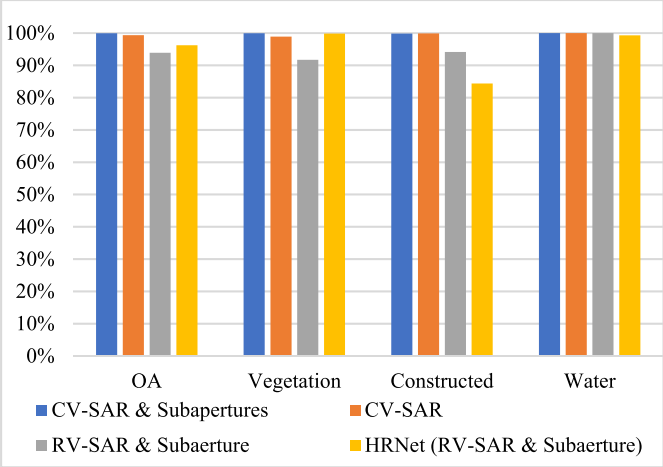


Fig. 6. Classification accuracies for the three-class semantic annotations in different case studies.

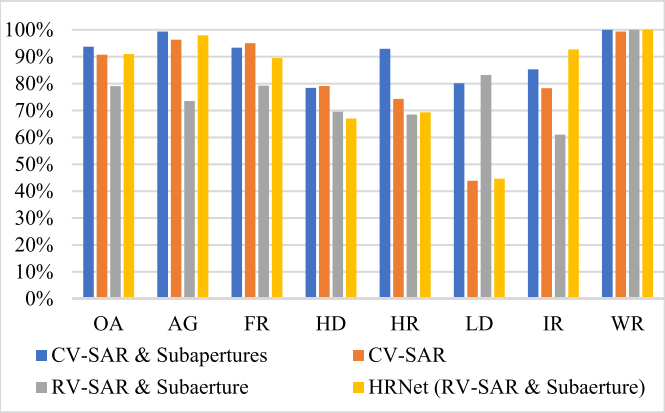


Fig. 7. Classification accuracies for the seven-class semantic annotations in different case studies.

TABLE III

CONFUSION MATRIX FOR SEVEN-CLASS CLASSIFICATION OF THE CV-SAR IMAGES WITH THE PROPOSED METHOD

	AG	FR	HD	HR	LD	IR	WR
AG	0.9631	0.0344	0.0000	0.0000	0.0016	0.0010	0.0000
FR	0.0493	0.9502	0.0001	0.0000	0.0000	0.0004	0.0000
HD	0.0000	0.0000	0.7910	0.0186	0.1045	0.0859	0.0000
HR	0.0207	0.0000	0.1950	0.7427	0.0415	0.0000	0.0000
LD	0.0845	0.1140	0.1343	0.0000	0.4385	0.2287	0.0000
IR	0.0034	0.2029	0.0034	0.0000	0.0073	0.7828	0.0000
WR	0.0068	0.0000	0.0000	0.0000	0.0000	0.0000	0.9932

for the seven-class classification of the case study. The main reason for the very low classification accuracy of the LD class is the confusion between this class and the other CS classes, especially industrial fields. The confusion between the LD and the VG classes, especially FR class, is because of the frequent VG covers in the LD patches (e.g., VG cover around the villas or between the houses).

The other three CS classes, HD, HR, and IR semantic classes, are classified with almost similar classification accuracies, around 78%. However, as demonstrated in Table III, the most frequent error is the confusion between different CS classes.

In conclusion, despite the confusion between different CS classes, the proposed CV end-to-end deep network successfully classified the challenging S1SLC_CVDL dataset with reasonable accuracy and utilized the amplitude and phase information of the CV-SAR images.

B. Case Study 2, CV-SAR and Azimuth Subaperture Classification

In the SAR technology, all the backscattered signals from a particular target from different angles are integrated to achieve a higher resolution in the azimuth direction. Azimuth subaperture decomposition can create subaperture images from the target which illuminates the target from different azimuth angles [25]. Subaperture images have a lower resolution in the azimuth direction but are proved to be advantageous for the detection of the artificial targets in SAR images [25]. Moreover, subaperture images contain physical properties of the target and using them in the deep model will incorporate the physics-aware attributes into the deep model.

In this case study, three overlapping (20%) subaperture images are extracted from each polarization channel of the CV-SAR patches (i.e., six CV subaperture images for each patch). Azimuth subaperture decomposition is applied using the Fourier transform technique, and the Hamming window coefficients are the same as the coefficients reported on the S1 product documentation [58]. Fourier transform converts the original CV-SAR images into the corresponding spectrum in the Fourier frequency domain. The hamming window is used to extract three overlapping subapertures, covering different portions of the spectrum in the azimuth direction. Later, the inverse Fourier transform is used to obtain the CV subaperture images [22]. For further explanation on the azimuth subaperture technique, refer to [22].

The input of the network in this case study consists of eight CV channels. The CV subaperture images (six channels, three from each polarization channel) along with the CV-SAR images (two channels) are fed into the proposed CV deep model, and the classification results are reported with the blue color charts in Figs. 6 and 7. The confusion matrix is also shown in Table IV. The overall accuracy increased about 3% in comparison to the previous case study and reached 93.72%. The most apparent improvement is in the LD semantic class. LD class is classified with more than 80% accuracy, which is almost two times more than the classification accuracy of this class in the previous case study. However, the confusion between different CS semantic classes is still a challenge. The most problematic confusion in this case study is between the HD and HR semantic classes.

In conclusion, adding the CV azimuth subaperture images has enabled the proposed CV deep model to better classify different CS semantic classes. The physics-aware attribute of the deep model improved the classification results significantly, especially in the CS areas. As shown in Section III, the proposed CV dataset is very challenging and contains very similar targets, especially in CS classes. However, the proposed CV end-to-end deep model is successful to classify the challenging dataset and exploit the valuable information in both amplitude and phase components of the CV-SAR

TABLE IV
CONFUSION MATRIX FOR SEVEN-CLASS CLASSIFICATION OF THE CV-SAR AND CV SUBAPERTURE IMAGES WITH THE PROPOSED METHOD

	AG	FR	HD	HR	LD	IR	WR
AG	0.9934	0.0019	0.0000	0.0000	0.0044	0.0003	0.0000
FR	0.0239	0.9334	0.0000	0.0000	0.0100	0.0327	0.0000
HD	0.0000	0.0000	0.7839	0.1497	0.0647	0.0018	0.0000
HR	0.0000	0.0000	0.0664	0.9295	0.0041	0.0000	0.0000
LD	0.0257	0.0000	0.1275	0.0423	0.8008	0.0038	0.0000
IR	0.0002	0.0000	0.0133	0.0005	0.1335	0.8526	0.0000
WR	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	1.0000

TABLE V
CONFUSION MATRIX FOR SEVEN-CLASS CLASSIFICATION OF THE RV-SAR AND RV SUBAPERTURE IMAGES WITH THE EQUIVALENT RV MODEL

	AG	FR	HD	HR	LD	IR	WR
AG	0.7353	0.1418	0.0000	0.0000	0.0790	0.0014	0.0425
FR	0.0772	0.7922	0.0001	0.0000	0.0957	0.0349	0.0000
HD	0.0000	0.0000	0.6953	0.0257	0.2710	0.0080	0.0000
HR	0.0000	0.0000	0.3112	0.6846	0.0041	0.0000	0.0000
LD	0.0015	0.0000	0.1426	0.0053	0.8317	0.0189	0.0000
IR	0.0000	0.0190	0.0044	0.0000	0.3669	0.6097	0.0000
WR	0.0003	0.0000	0.0000	0.0000	0.0000	0.0000	0.9997

images, as well as the physical properties in the CV azimuth subaperture images.

C. Case Study 3, CV Versus RV Model

In order to evaluate the effect of the CV network on the classification results, an equivalent RV end-to-end deep network is designed. Similar to the previous case study, eight input channels are used in this case study. The amplitude of the SAR and the subaperture images used in the RV network and the classification results are reported with the gray color charts in Figs. 6 and 7. The confusion matrix is also shown in Table V. The RV equivalent network with RV-SAR and RV-subaperture images obtained 79.01% overall classification accuracy.

The CV network achieved higher classification accuracy for all of the classes, when the same input images (SAR and azimuth subaperture images) are used. The better performance of the CV network demonstrates the superiority of this network. The CV model can exploit more valuable and distinctive information from both the amplitude and phase components of the SAR data.

In comparison to the CV network with only CV-SAR images (i.e., without the subaperture images), the RV network with SAR and subaperture images obtained higher classification accuracy in the LD and IR semantic classes. However, the CV network obtained more than 10% and 5% higher overall classification accuracies in seven-class and three-class classifications, respectively. The better performance of the RV model (with subaperture images) demonstrates the importance of the physics-aware attributes for the classification of the CS area in SAR images.

In conclusion, the CV deep network achieved about 14% higher overall classification accuracy with the same input data.

TABLE VI
CONFUSION MATRIX FOR SEVEN-CLASS CLASSIFICATION OF THE RV-SAR AND RV SUBAPERTURE IMAGES WITH THE RV HRNET MODEL

	AG	FR	HD	HR	LD	IR	WR
AG	0.9793	0.0148	0.0000	0.0000	0.0020	0.0034	0.0004
FR	0.0994	0.8948	0.0000	0.0000	0.0015	0.0043	0.0000
HD	0.0000	0.0000	0.6696	0.0168	0.0983	0.2152	0.0000
HR	0.0000	0.0000	0.3029	0.6929	0.0041	0.0000	0.0000
LD	0.0951	0.0249	0.1041	0.0075	0.4460	0.3223	0.0000
IR	0.0057	0.0488	0.0041	0.0000	0.0147	0.9266	0.0000
WR	0.0007	0.0000	0.0000	0.0000	0.0000	0.0000	0.9993

Only in the LD class, the RV network achieved slightly higher classification accuracy. However, the very high false-positive rate of this class in the RV model (see Table V) explains the higher accuracy and demonstrates the poor performance of the RV network despite the higher accuracy for this class.

D. Case Study 4, CV Versus RV HRNet Model

The architecture used in this study (shown in Fig. 1) is a simple and basic convolutional network, and it can be improved to achieve better results. Many different architectures have been proposed in the literature to improve the performance of the deep networks. HRNet is initially proposed for human pose estimation [59], but later extended to other applications, including classification [57]. An HRNet consists of parallel convolutions with different resolutions and repeatedly conduct multiscale fusions across the parallel convolutions [57].

In the third case study, we have compared the CV and RV networks with the same architecture and concluded that the CV network has superior performance. However, the state-of the-art architectures can improve the performance of the RV network. In this case study, CV network with the same architecture, as shown in Fig. 1, is compared with the RV HRNet. Similar to the previous case study, eight input channels, consisting of two SAR polarization channels and six subaperture images, are used (i.e., in complex domain for the CV network and in real domain for the RV HRNet).

The classification results are illustrated with yellow color in Figs. 6 and 7. The HRNet obtained 90.97% overall accuracy, which is almost 3% lower than the CV network. The CV network achieved higher classification accuracy for all of the classes, except for the IR class. Despite the higher classification accuracy for IR class, the false positive rate is very high, because of the misclassification of HD and LD patches in this class (see the confusion matrix in Table VI).

However, the RV HRNet obtained remarkably better performance than the RV network with the architecture in Fig. 1. The input for both of these networks is the same (eight RV channels), but the HRNet obtained almost 12% higher overall classification accuracy.

The results of this case study demonstrated that the HRNet has a better architecture than the network shown in Fig. 1, but the CV operators achieved higher classification accuracy even with a shallower and weaker architecture. It should also be noted that HRNet has more trainable parameters than the

CV network (HRNet has 21 301 884 parameters and the CV network has 18 466 186 parameters).

E. Case Study 5, Coherency Preservation Capability

In SAR imagery, the complex correlation coefficient (coherency) contains important information about the SAR system and the physical properties of the target [24]. For InSAR applications, the coherency between different CV-SAR observations over an area is the most important measurement and is utilized to derive the digital elevation model (DEM) of the Earth's surface [60]. The complex correlation coefficient is used in several studies to characterize the properties and the dynamics of the Earth's surface, including forestry applications, snow and glacier monitoring, and volcanic studies [24], [61], [62], [63]. As a result, the coherency of CV-SAR data must be preserved in the processing chain and a fully CV network with a coherency preservation feature is necessary.

Processing the amplitude and phase components (or the real and imaginary parts) of the CV-SAR data separately will interrupt the coherency of CV-SAR data. Although CV deep networks can be beneficial to preserve this coherency and exploit the phase information [28], the coherency preservation capability of the developed CV networks should be evaluated and monitored to ensure the coherency preservation. Moreover, the coherency preservation is a reliable measure to evaluate the performance of the CV deep networks to adequately process and preserve the phase information of the SAR data.

Most of the previous attempts for developing the CV deep networks for CV-SAR data processing and classification neglected the coherency preservation of the network and focused solely on the classification results. Although the high classification accuracy is the main objective of these CV models, the coherency preservation can enhance the stability and generalizability of the CV deep models as well as ensuring the reliability of the outputs. Moreover, coherency preservation is necessary for InSAR applications. As a result, the coherency preservation capabilities of the developed CV end-to-end deep network are evaluated in this case study.

The similar CV network as the network in case study 2 is used in this case study, with the same input channels (eight complex-valued input channels consisting of two CV-SAR and six CV subaperture images). For this purpose, in the first step, the correlation between the reconstructed CV-SAR image by the CV-CAE and the input image is assessed. The correlation is computed as shown in the following equation, where $C()$ is the coherence between any pair of images Z, and $E()$ and $*$ are the expected value over an arbitrary window and the complex conjugation operator, respectively. An 11×11 window with equal weights, ones, is applied in this study [24], [64]

$$C(Z_1, Z_2) = \frac{E\left(Z_1 Z_2^*\right)}{\sqrt{E\left(Z_1 Z_1^*\right) E\left(Z_2 Z_2^*\right)}}. \tag{37}$$

The average coherency between the input and the reconstructed CV-SAR patches is 0.9415. The high coherency demonstrates the preservation of the complex correlation coefficient of the CV-SAR patches in the CV deep model. Fig. 8 illustrates some examples of the input and the reconstructed SAR patches with their coherency patch. The coherency values range between 0 and 1; however, the coherency images are shown in the range of 0.9–1 for better visualization.

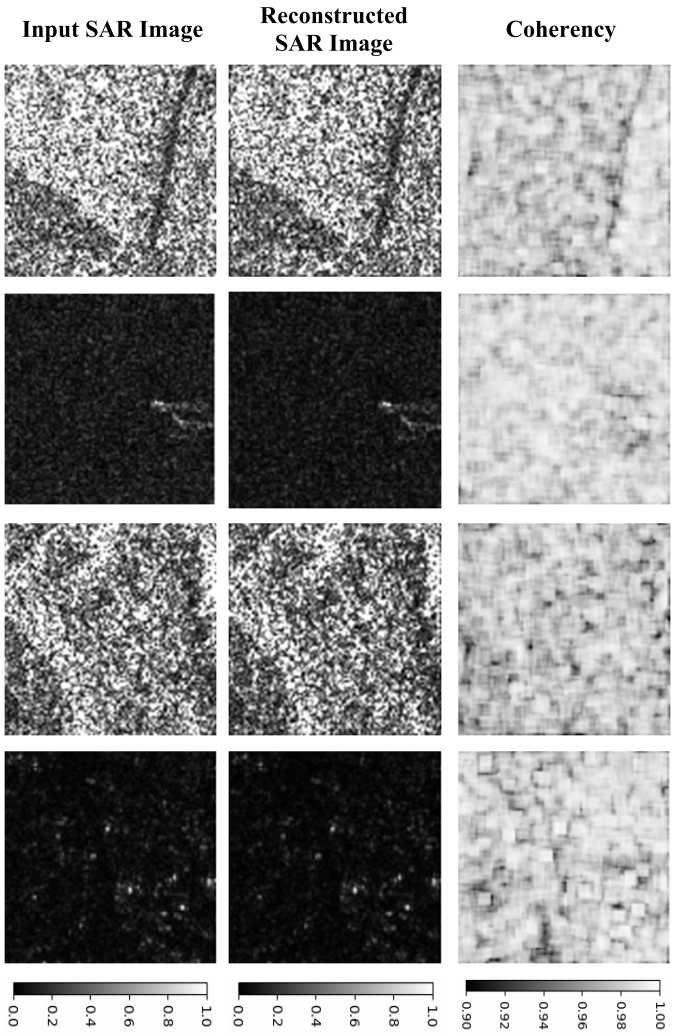


Fig. 8. Example input and the reconstructed SAR patches with their coherency. The coherency image is shown in the range of 0.90–1 for better visualization.

Furthermore, in the SAR technology, all the backscattered signals from a particular target from different angles are integrated to achieve a higher resolution in the azimuth direction. Azimuth subaperture decomposition can create subaperture images from the target which illuminates the target from different azimuth angles [25]. The processing chain of the CV-SAR data should preserve the data model. In other words, the developed CV deep network should learn the CV data model in a way that different apertures in the reconstructed CV-SAR image still represent different antenna angles.

In order to evaluate the developed CV end-to-end deep model for learning and preserving the CV data model, the subaperture decomposition is also applied to the reconstructed CV-SAR patches. The coherencies between the subaperture pairs of the input and reconstructed images are computed. The average coherency of 0.9214 between the subaperture images demonstrates the capability of the developed CV deep model to learn and preserve the data model. Fig. 9 illustrates some

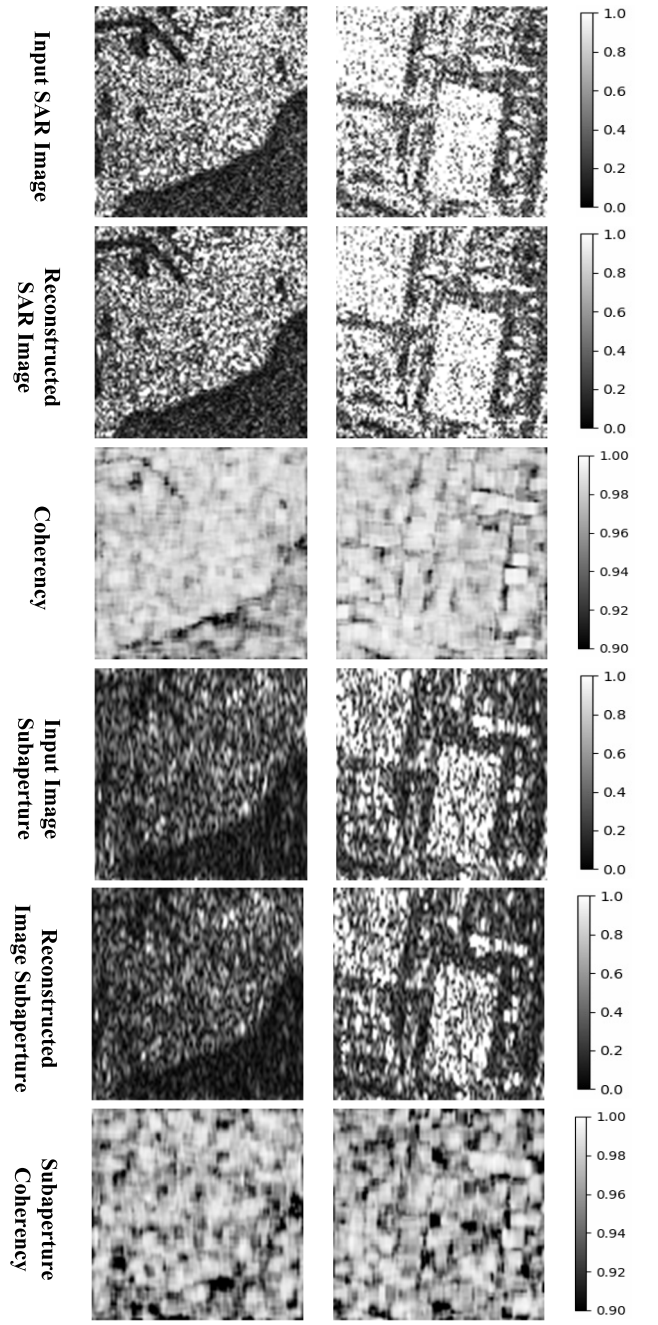


Fig. 9. Example input and the reconstructed SAR patches and their subaperture images, with their coherency. The coherency images are shown in the range of 0.90–1 for better visualization.

example patches with their subaperture and the coherency images. (Similar to Fig. 8, the coherency images are shown in the range of 0.9–1 for better visualization.)

V. Conclusion

Following the remarkable results, deep neural networks have attracted many interests in different fields. However, applying RV deep networks to CV-SAR data neglects the phase component and only uses the amplitude information, whereas the phase component of the CV-SAR data conveys important information about the characteristics and dynamics of the Earth's surface. CV deep networks have been developed in a few studies to address this issue and utilize the phase and amplitude of the CV-SAR data. In this study, a CV end-to-end deep network is developed for CV-SAR data reconstruction and classification and addressed a number of shortcomings of the previous CV models for SAR data processing.

All of the necessary CV operators for the CV deep model's development are derived from the RV operators. Comprehensive detailed explanations of the CV operators are provided. In addition, the lack of a large-scale semantically annotated CV-SAR dataset is one of the main restricting factors in the development of the CV deep networks for SAR data processing. As a result, S1SLC_CVDL annotated dataset, a large-scale CV-SAR dataset consisting of 100×100 pixels patches, is created, using the S1 SLC SM products. The well-known classification algorithms (i.e., Gabor features and SVM classifier) are used for semantic annotation and an LDA-based semantic data mining is applied to remove the misclassified and ambiguous patches. The developed CV end-to-end deep model is applied to the developed S1SLC_CVDL dataset and achieved the high classification accuracy of 90.7%, despite the challenging semantic classes of the dataset (especially the CS semantic classes).

Moreover, the azimuth subaperture decomposition is applied to the CV-SAR data and used as the input of the network, alongside the CV-SAR data. The subaperture images incorporated the physics-aware attributes to the deep network and improved the classification accuracy to 93.72%. The classification improvement is more apparent in the challenging CS semantic classes. The RV equivalent network with amplitude SAR data is used to compare the results of the CV and RV networks. The CV model outperformed the RV equivalent network in all of the classes and achieved the high overall classification accuracy of 93.72%, in comparison with the 79.01% overall accuracy of the RV network.

Later, the state-of-the-art classification architecture, HRNet, is also used for comparison. The HRNet is used to classify the RV-SAR data and obtained the 90.97% overall classification accuracy. HRNet achieved much higher classification accuracy of the RV network (79.01%), used in this study. However, the classification accuracy of the CV network surpassed the RV HRNet with about 3%. This comparison demonstrated that the CV network can achieve higher classification accuracy, even with a shallower and less complicated (lower number of the parameters) architecture.

Furthermore, the complex correlation coefficient (coherency) of the CV-SAR data contains very important information about the characteristics of the SAR system, as well as the physical properties of the target. In many InSAR applications, coherency is the most important measurement. As a result, the CV model should preserve the coherency of the CV-SAR data. The coherency preservation has been neglected in the previous CV networks for SAR data processing. The experiments on the coherency of the reconstructed CV-SAR data demonstrated that the developed CV deep model preserves the coherency. Additionally, the high coherency between the subaperture images of the input and the reconstructed CV-SAR data illustrated that the CV deep network learned and preserved the synthetic aperture data model.

As the future prospect, more studies in this field are necessary to develop more CV deep networks. The HRNet has obtained admirable and state-of-the-art results in different computer vision applications, and the CV extension of this network should be developed in the future studies.

ACKNOWLEDGMENT

The authors would like to thank European Space Agency (ESA), Paris, France, for making the data, used in this study, publicly available. This study is implemented in Python environment and the code is available for research purposes at https://github.com/Reza-Asiyabi/Complex-Valued-End-to-End-Deep-Network-for-SAR. The data used in this article are Sentinel-1 data, an ESA mission, which are available free of charge through the Sentinel data hub at https://scihub.copernicus.eu/dhus/#/home.

REFERENCES

[1] J. Zhao, Z. Zhang, W. Yao, M. Datcu, H. Xiong, and W. Yu, "OpenSARUrban: A Sentinel-1 SAR image dataset for urban interpretation," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 187–203, 2020.

[2] H. Zhang, H. Lin, and Y. Wang, "A new scheme for urban impervious surface classification from SAR images," *ISPRS J. Photogramm. Remote Sens.*, vol. 139, pp. 103–118, May 2018.

[3] R. M. Asiyabi, M. R. Sahebi, and A. Ghorbanian, "Segment-based bag of visual words model for urban land cover mapping using polarimetric SAR data," *Adv. Space Res.*, vol. 70, no. 12, pp. 3784–3797, Dec. 2022, doi: 10.1016/j.asr.2021.10.042.

[4] Z. Huang, M. Datcu, Z. Pan, and B. Lei, "Deep SAR-Net: Learning objects from signals," *ISPRS J. Photogramm. Remote Sens.*, vol. 161, pp. 179–193, Mar. 2020.

[5] 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.

[6] X. Xiang Zhu et al., "Deep learning meets SAR," 2020, *arXiv:2006.10027*.

[7] T. Zhang et al., "HOG-ShipCLSNet: A novel deep learning network with hog feature fusion for SAR ship classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–22, 2021.

[8] F. Zhao et al., "Monthly mapping of forest harvesting using dense time series Sentinel-1 SAR imagery and deep learning," *Remote Sens. Environ.*, vol. 269, Feb. 2022, Art. no. 112822.

[9] Y. Wang, C. Wang, H. Zhang, Y. Dong, and S. Wei, "A SAR dataset of ship detection for deep learning under complex backgrounds," *Remote Sens.*, vol. 11, no. 7, p. 765, 2019.

[10] X. Ma, C. Wang, Z. Yin, and P. Wu, "SAR image despeckling by noisy reference-based deep learning method," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8807–8818, Dec. 2020.

[11] Z. Huang, X. Yao, Y. Liu, C. O. Dumitru, M. Datcu, and J. Han, "Physically explainable CNN for SAR image classification," 2021, *arXiv:2110.14144*.

[12] Z. Zhang, H. Wang, F. Xu, and Y.-Q. Jin, "Complex-valued convolutional neural network and its application in polarimetric SAR image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 12, pp. 7177–7188, Dec. 2017.

[13] R. Hänsch and O. Hellwich, "Classification of polarimetric SAR data by complex valued neural networks," in *Proc. ISPRS Workshop High-Resolution Earth Imag. Geospatial Inf.*, vol. 38, 2009, pp. 4–7.

[14] M. Wilmanski, C. Kreucher, and A. Hero, "Complex input convolutional neural networks for wide angle SAR ATR," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2016, pp. 1037–1041.

[15] R. Shang, G. Wang, M. A. Okoth, and L. Jiao, "Complex-valued convolutional autoencoder and spatial pixel-squares refinement for polarimetric SAR image classification," *Remote Sens.*, vol. 11, no. 5, p. 522, Feb. 2019.

[16] Y. Sunaga, R. Natsuaki, and A. Hirose, "Land form classification and similar land-shape discovery by using complex-valued convolutional neural networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7907–7917, Oct. 2019.

[17] Q. Sun, X. Li, L. Li, X. Liu, F. Liu, and L. Jiao, "Semi-supervised complex-valued GAN for polarimetric SAR image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 3245–3248.

[18] R. Zhang, Y. Wang, J. Hu, W. Yang, J. Chen, and X. Zhu, "SAR4LCZ-Net: A complex-valued convolutional neural network for local climate zones classification using Gaofen-3 quad-pol SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–16, 2021.

[19] J. Liu, M. Xing, H. Yu, and G. Sun, "EFTL: Complex convolutional networks with electromagnetic feature transfer learning for SAR target recognition," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–11, 2021.

[20] Z. Zeng, J. Sun, Z. Han, and W. Hong, "SAR automatic target recognition method based on multi-stream complex-valued networks," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5228618.

[21] R. Mohammadi Asiyabi, M. Datcu, H. Nies, and A. Anghel, "Complex-valued vs. real-valued convolutional neural network for PolSAR data classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2022, pp. 421–424.

[22] R. Mohammadi Asiyabi, M. Datcu, A. Anghel, and H. Nies, "Complex-valued autoencoders with coherence preservation for SAR," in *Proc. EuSAR*, 2022, pp. 1–6.

[23] K. Qian, Y. Wang, Y. Shi, and X. X. Zhu, "γ-Net: Superresolving SAR tomographic inversion via deep learning," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4706116.

[24] C. López-Martínez and E. Pottier, "Coherence estimation in synthetic aperture radar data based on speckle noise modeling," *Appl. Opt.*, vol. 46, no. 4, pp. 544–558, Feb. 2007.

[25] L. Ferro-Famil, A. Reigber, E. Pottier, and W. M. Boerner, "Scene characterization using subaperture polarimetric SAR data," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 10, pp. 2264–2276, Oct. 2003.

[26] J. Bassey, L. Qian, and X. Li, "A survey of complex-valued neural networks," 2021, *arXiv:2101.12249*.

[27] A. Hirose and S. Yoshida, "Comparison of complex- and real-valued feedforward neural networks in their generalization ability," in *Proc. Int. Conf. Neural Inf. Process.*, 2011, pp. 526–531.

[28] J. A. Barrachina, C. Ren, C. Morisseau, G. Vieillard, and J.-P. Ovarlez, "Complex-valued vs. real-valued neural networks for classification perspectives: An example on non-circular data," 2020, *arXiv:2009.08340*.

[29] Y. Sunaga, R. Natsuaki, and A. Hirose, "Similar land-form discovery: Complex absolute-value max pooling in complex-valued convolutional neural networks in interferometric synthetic aperture radar," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2020, pp. 1–7.

[30] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, Lille, France, vol. 2015, vol. 37, pp. 448–456.

[31] C. Trabelsi et al., "Deep complex networks," 2018, *arXiv:1705.09792*.

[32] P. M. Virtue, *Complex-Valued Deep Learning With Applications to Magnetic Resonance Image Synthesis*. Berkeley, CA, USA: Univ. California, 2019.

[33] Y. Cao, Y. Wu, P. Zhang, W. Liang, and M. Li, "Pixel-wise PolSAR image classification via a novel complex-valued deep fully convolutional network," *Remote Sens.*, vol. 11, no. 22, p. 2653, Nov. 2019.

[34] A. Hirose, "Complex-valued neural networks: The merits and their origins," in *Proc. Int. Joint Conf. Neural Netw.*, Jun. 2009, pp. 1237–1244.

[35] Y. Kuroe, M. Yoshid, and T. Mori, "On activation functions for complex-valued neural networks—Existence of energy functions," in *Artificial Neural Networks and Neural Information Processing—ICANN/ICONIP*. Cham, Switzerland: Springer, 2003, pp. 985–992.

[36] N. Benvenuto and F. Piazza, "On the complex backpropagation algorithm," *IEEE Trans. Signal Process.*, vol. 40, no. 4, pp. 967–969, Apr. 1992.

[37] G. M. Georgiou and C. Koutsougeras, "Complex domain backpropagation," *IEEE Trans. Circuits Syst. II, Analog Digit. Signal Process.*, vol. 39, no. 5, pp. 330–334, May 1992.

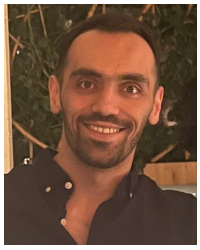
[38] A. Hirose, "Continuous complex-valued back-propagation learning," *Electron. Lett.*, vol. 28, no. 20, pp. 1854–1855, Sep. 1992.

[39] A. Hirose, *Complex-Valued Neural Networks*, vol. 400. Cham, Switzerland: Springer, 2012.

[40] N. Guberman, "On complex valued convolutional neural networks," 2016, *arXiv1602.09046*.

[41] M. Arjovsky, A. Shah, and Y. Bengio, "Unitary evolution recurrent neural networks," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1120–1128.

[42] H. Mu, Y. Zhang, Y. Jiang, and C. Ding, "CV-GMTINet: GMTI using a deep complex-valued convolutional neural network for multichannel SAR-GMTI system," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–15, 2021.
[43] J. Chen, X. Qiu, C. Ding, and Y. Wu, "CVCMFF Net: Complex-valued convolutional and multifeature fusion network for building semantic segmentation of InSAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–14, 2021.
[44] W. Wirtinger, "Zur formalen theorie der funktionen von mehr komplexen Veränderlichen," *Mathematische Annalen*, vol. 97, no. 1, pp. 357–375, Dec. 1927.
[45] T. Kim and T. Adali, "Fully complex backpropagation for constant envelope signal processing," in *Proc. Neural Netw. Signal Process. X. IEEE Signal Process. Soc. Workshop*, 2000, pp. 231–240.
[46] J. Zhang and Y. Wu, "A new method for automatic sleep stage classification," *IEEE Trans. Biomed. Circuits Syst.*, vol. 11, no. 5, pp. 1097–1110, Oct. 2017.
[47] P. J. Schreier and L. L. Scharf, *Statistical Signal Processing of Complex-Valued Data: The Theory of Improper and Noncircular Signals*. Cambridge, U.K.: Cambridge Univ. Press, 2010.
[48] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th SIGSPATIAL Int. Conf. Adv. Geographic Inf. Syst.*, Nov. 2010, pp. 270–279.
[49] P. Helber, B. Bischke, A. Dengel, and D. Borth, "EuroSAT: A novel dataset and deep learning benchmark for land use and land cover classification," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 12, no. 7, pp. 2217–2226, Jul. 2019.
[50] E. Keydel, S. Lee, and J. Moore, "MSTAR extended operating conditions: A tutorial," *Proc. SPIE*, vol. 2757, pp. 228–242, Jun. 1996.
[51] M. Schmitt, S. A. Ahmadi, and R. Hansch, "There is no data like more data–current status of machine learning datasets in remote sensing," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2021, pp. 1206–1209.
[52] W. Xie, G. Ma, W. Hua, and F. Zhao, "Complex-valued Wishart stacked auto-encoder network for PolSAR image classification," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 3193–3196.
[53] D. Geudtner, R. Torres, P. Snoeij, M. Davidson, and B. Rommen, "Sentinel-1 system capabilities and applications," in *Proc. IEEE Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 1457–1460.
[54] J.-K. Kamarainen, V. Kyrki, and H. Kalviainen, "Invariance properties of Gabor filter-based features—Overview and applications," *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1088–1099, May 2006.
[55] R. M. Asiyabi and M. Datcu, "Earth observation semantic data mining: Latent Dirichlet allocation-based approach," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 15, pp. 2607–2620, 2022.
[56] H. N. Reza Mohammadi Asiyabi, M. Datcu, and A. Anghel, "S1SLC_CVDL: A complex-valued annotated single look complex Sentinel-1 SAR dataset for complex-valued deep networks," *IEEE Dataport*, early access, Apr. 14, 2023, doi: 10.1109/TGRS.2023.3267185.
[57] K. Sun et al., "High-resolution representations for labeling pixels and regions," 2019, *arXiv:1904.04514*.
[58] M. Bourbigot, H. Johnsen, R. Piantanida, G. Hajduch, J. Poullaouec, and G. Hajduch. (2016). *Sentinel-1 Product Definition*. ESA Unclassified. Accessed: Feb. 18, 2021. [Online]. Available: https//sentinel.esa.int/documents/247904/1877131/Sentinel-1-Product-Definition
[59] K. Sun, B. Xiao, D. Liu, and J. Wang, "Deep high-resolution representation learning for human pose estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5693–5703.
[60] R. Bamler and P. Hartl, "Synthetic aperture radar interferometry," *Inverse Problems*, vol. 14, no. 4, pp. R1–R54, Aug. 1998.
[61] J. P. Matthews, H. Kamata, S. Okuyama, Y. Yusa, and H. Shimizu, "Surface height adjustments in pyroclastic-flow deposits observed at Unzen volcano by JERS-1 SAR interferometry," *J. Volcanol. Geothermal Res.*, vol. 125, nos. 3–4, pp. 247–270, Jul. 2003.
[62] M. E. Engdahl, J. T. Pulliainen, and M. T. Hallikainen, "Boreal forest coherence-based measures of interferometric pair suitability for operational stem volume retrieval," *IEEE Geosci. Remote Sens. Lett.*, vol. 1, no. 3, pp. 228–231, Jul. 2004.
[63] S. L. C. Benson, L. Shapiro, and K. Dean, "Aufeis in the Ivishak River, Alaska, mapped from satellite radar interferometry," *Remote Sens. Environ.*, vol. 60, no. 2, pp. 131–139, May 1997.
[64] J.-S. Lee, K. W. Hoppel, S. A. Mango, and A. R. Miller, "Intensity and phase statistics of multilook polarimetric and interferometric SAR imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 32, no. 5, pp. 1017–1028, Sep. 1994.



Reza Mohammadi Asiyabi (Member, IEEE) received the M.Sc. degree in geodesy and geomatics engineering—remote sensing from the K. N. Toosi University of Technology, Tehran, Iran, in 2018. His M.Sc. thesis focused on the enhancement of synthetic aperture radar (SAR) data classification using bag of visual words representation model. He is currently pursuing the Ph.D. degree with the University POLITEHNICA of Bucharest (UPB), Bucharest, Romania.

He is an Early Stage Researcher within the EU Marie Skłodowska-Curie Innovative Training Network (ITN), MENELAOS-NT, UPB, and a Ph.D. Researcher with the Research Center for Spatial Information (CEOSpaceTech), Faculty of Electronics Engineering, Telecommunications and Information Technology, UPB. His research is on semantic information discovery and classification of SAR data, using data mining techniques and complex-valued deep learning models with a focus on complex-valued and uncertainty aware deep networks for various SAR applications.

Mihai Datcu (Fellow, IEEE) received the M.S. and Ph.D. degrees in electronics and telecommunications from the University POLITEHNICA of Bucharest (UPB), Bucharest, Romania, in 1978 and 1986, respectively, and the Habilitation a Diriger des Recherches degree in computer science from the University Louis Pasteur, Strasbourg, France, in 1999.

Since 1981, he has been with the Department of Applied Electronics and Information Engineering, Faculty of Electronics, Telecommunications and Information Technology, UPB, where he is currently a Full Professor and the Director of the Research Center for Spatial Information (CEOSapceTech). Since 1993, he has been with the German Aerospace Center (DLR), Weßling, Germany, where he is also a Senior Scientist with the Remote Sensing Technology Institute (IMF). From 1992 to 2002, he had a longer Invited Professor Assignment with the Swiss Federal Institute of Technology (ETH Zurich), Zurich, Switzerland. Since 2001, he had been initiating and leading the Competence Center on Information Extraction and Image Understanding for Earth Observation, ParisTech, Paris Institute of Technology, Paris, France, a collaboration of DLR with the French Space Agency (CNES). From 2005 to 2013, he was a Professor with the DLR-CNES Chair, ParisTech, Paris Institute of Technology. From 2011 to 2018, he was leading the Immersive Visual Information Mining Research Laboratory, Munich Aerospace Faculty, Munich, Germany. From 2018 to 2020, he was the holder of the Blaise Pascal International Chair of Excellence at Conservatoire national des arts et métiers (CNAM), Paris. From 2020 to 2022, he was involved with the DLR-French Aerospace Laboratory (ONERA) Joint Virtual Center for AI in Aerospace, Paris, France. He was a Visiting Professor with the University of Oviedo, Oviedo, Spain; University Louis Pasteur and International Space University, Strasbourg; the University of Siegen, Siegen, Germany; the University of Innsbruck, Innsbruck, Austria; the University of Alcalá, Alcalá de Henares, Spain; University Tor Vergata, Rome, Italy; the Universidad Pontificia de Salamanca, Madrid, Spain; the University of Camerino, Camerino, Italy; the University of Trento, Trento, Italy, the Chinese Academy of Sciences, Shenyang, China; the Universidade Estadual de Campinas (UNICAMP), Campinas, Brazil; the University of Wuhan, Wuhan, China, and the Swiss Center for Scientific Computing, Manno, Switzerland. He has initiated and implemented the European frame of projects for Earth observation image information mining (IIM) and is involved in research programs for information extraction, data mining, big EO data knowledge discovery, and data understanding with the European Space Agency (ESA), NASA, Washington, DC, USA, and in a series of national and European projects. He is also a Visiting Professor with ESA's Φ-Laboratory, Frascati, Italy. He and his team have developed the operational IIM processor in the payload ground segment systems for the German mission TerraSAR-X, and data mining tools and systems for the Copernicus missions Sentinel-1 and Sentinel-2. He is developing algorithms for model-based information retrieval from high-complexity signals and methods for scene understanding from very-high-resolution synthetic aperture radar (SAR) and interferometric SAR data. His research interests include information theory, signal processing, explainable and physics-aware artificial intelligence, computational imaging, and quantum machine learning with applications in EO.

Dr. Datcu is a member of the ESA Working Group Big Data from Space. He was a recipient of the Romanian Academy Prize Traian Vuia for the development of the system for automatic digital image analysis (SAADI) system and his activity in image processing in 1987, the best paper award and the IEEE Geoscience and Remote Sensing Society Prize in 2006, the National Order of Merit with the rank of Knight, for outstanding international research results, awarded by the President of Romania in 2008, the Chaire d'excellence international Blaise Pascal 2017 for international recognition in the field of data science in EO, the 2018 Ad Astra Award for Excellence in Science, and the IEEE GRSS David Landgrebe Award in recognition of outstanding contributions to Earth observation analysis using innovative concepts for big data analysis, image mining, machine learning, smart sensors, and quantum resources, in 2022. He has served as a co-organizer for international conferences and workshops and as a guest editor and an associate editor for IEEE and other journals.

Andrei Anghel (Senior Member, IEEE) received the Engineering degree (as valedictorian) and the M.S. degree (Hons.) in electronic engineering and telecommunications from the University POLITEHNICA of Bucharest (UPB), Bucharest, Romania, in 2010 and 2012, respectively, the joint Ph.D. degree (summa cum laude) in signal, image, speech, and telecoms from the University of Grenoble Alpes, Grenoble, France, and in electronic engineering and telecommunications from UPB, in 2015, and the Habilitation degree in electronic engineering, telecommunications, and information technologies from UPB, in 2020.

From 2012 to 2015, he worked as a Doctoral Researcher with the Grenoble Image Speech Signal Automatics Laboratory (GIPSA-Lab), Grenoble. In 2012, he joined UPB as a Teaching Assistant, where he is currently a Professor with the Telecommunications Department, Faculty of Electronics, Telecommunications and Information Technology, and a Researcher with the Research Center for Spatial Information (CEOSpaceTech). He has authored more than 50 scientific publications, two textbooks, and a book about synthetic aperture radar (SAR) signal processing for infrastructure monitoring. His research interests include remote sensing, radar, microwaves, and signal processing.

Dr. Anghel regularly acts as a reviewer for several IEEE and IET journals. He was awarded the "Traian Vuia" Prize by the Romanian Academy for his work on bistatic radar systems with a fixed ground-based receiver and spaceborne transmitter of opportunity, in 2021. He was a recipient of two gold medals at the International Physics Olympiads, in 2005 and 2006.

Holger Nies received the Diploma degree in electrical engineering and the Dr.-Ing. degree from the University of Siegen, Siegen, Germany, in 1999 and 2006, respectively.

Since 1999, he has been a member of the Center for Sensor Systems (ZESS) and a Lecturer with the Department of Electrical Engineering, University of Siegen. Since 2010, he has been the Executive Director of the Doctoral Program of ZESS. He is a Team Leader of the synthetic aperture radar (SAR) group of the ZESS and responsible for the HITCHHIKER project. He was involved in different project works for Daimler AG, Stuttgart, Germany, in the field of engine modeling and parameter estimation with Kalman filters. He was working in the area of SAR interferometry for the German TerraSAR-X/TanDEM-X mission. His research interests include mono- and bistatic SAR analysis and mission design, trajectory estimation, SAR interferometry, and passive radar systems.