Self-Supervised Learning Based Anomaly Detection in Synthetic Aperture Radar Imaging

MAX MUZEAU [1,2] (Graduate Student Member, IEEE), **CHENGFANG REN** [2] (Member, IEEE),
SÉBASTIEN ANGELLIAUME [2], **MIHAI DATCU** [3] (Fellow, IEEE),
AND JEAN-PHILIPPE OVARLEZ [1,2] (Member, IEEE)

[1]SONDRA, CentraleSupélec, Université Paris-Saclay, 91190 Gif-sur-Yvette, France
[2]DEMR, CentraleSupélec, Université Paris-Saclay, 91190 Gif-sur-Yvette, France
[3]University Politehnica of Bucharest (UPB), Romania and German Aerospace Center (DLR), D 82234 Wessling, Germany

CORRESPONDING AUTHOR: Max Muzeau (max.muzeau@ onera.fr).

ABSTRACT In this paper, we proposed to investigate unsupervised anomaly detection in Synthetic Aperture Radar (SAR) images. Our approach considers anomalies as abnormal patterns that deviate from their surroundings without prior knowledge of their characteristics. This method deals with the crucial problems related to the presence of speckle, the spatial correlation structures in SAR images, and the lack of annotated data to train a detection algorithm. Our proposed method aims to address these issues through a self-supervised learning algorithm. First, we propose to mitigate the SAR speckle through the deep learning SAR2SAR algorithm. We then develop an Adversarial Autoencoder (AAE) to reconstruct anomaly-free SAR images from despeckled data taking into account potential spatial correlation structures. Finally, a change detection processing step is applied between the input and the output to detect anomalies. Experiments are performed to show the advantages of our method compared to the conventional Reed-Xiaoli algorithm, highlighting the importance of an efficient despeckling pre-processing step.

INDEX TERMS Adversarial autoencoder, anomaly detection, deep-learning, despeckling, SAR, self-supervised.

I. INTRODUCTION

Anomaly detection is one of the most critical issues in multidimensional imaging, mainly in hyperspectral [1], [2] and medical imaging. Even if we have no prior information about a target or background signature, anomalies generally differ from surrounding pixels due to their dissimilar signatures. Anomaly detection for radar and SAR imaging aims to discover abnormal patterns hidden in multidimensional radar signals and images. Such anomalies could be man-made changes in a specific location or natural processes affecting a particular area. These anomalies can characterize, for example, several potential applications: Oil slick detection [3], turbulent ship wake [4], levee anomaly [5] or archaeology [6]. They can also be related to any change detection in time-series SAR images. This research field is essential in data mining for quickly isolating irregular or suspicious segments in large amounts of the database. Many anomaly detection schemes have been proposed in the literature [7], [8], [9], [10], [11],

[12], [13], [14], [15], [16], [17]. Among them, the unsupervised methods are the most interesting since they are widely applicable and do not require labeling the data.

Deep learning techniques for anomaly detection are often based on an encoding-decoding network architecture to learn healthy data or images that do not contain anomalies, such as references [14], [15]. Another approach is based on the one-class SVM method in the latent space [17]. Anomaly detection in SAR using the difference between the input and the reconstructed image obtained through the autoencoder scheme is also proposed in [16]. Still, it could suffer from the SAR speckle noise.

In the literature, most model-based algorithms face three main issues. First, the speckle noise corrupts the image and potentially leads to numerous false detections. Second, statistical approaches may exhibit deficiencies in modeling spatial correlation in SAR images. Finally, neural networks based on supervised learning approaches are not recommended due to

the lack of annotated SAR data, notably for the class of abnormal patterns. Our proposed method aims to address these issues through a self-supervised algorithm. Our approach consists in mitigating this annoying speckle with [22]. The algorithm is adapted to our dataset and trained so that every polarization can be despeckled with the same weights. After, an Adversarial Autoencoder (AAE) is trained from these despeckled data. It does not need to separate abnormal SAR images from the training dataset.

Our method's last step is a change detection between the pairwise original and reconstructed images. This comparison is performed through a distance metric based on covariance matrices estimated locally. We finally compare our method with the conventional RX anomaly detector [7] by analyzing their global performance, and we highlight the importance of the despeckling process.

This paper is organized as follows. After the introduction in Section I, Section II presents the context of SAR imaging. The proposed methodology being based on despeckling of SAR images, Section III describes the filtering process. Section IV presents the self-supervised neural architecture that aims to reconstruct the final SAR image without anomalies. Section V defines the last phase of the processing based on the change detection step and discusses the quality of the anomaly detection procedure. The proposed method is then applied to experimental polarimetric SAR images. Final Section VI gives some conclusions and perspectives.

Notations: italic indicates a scalar quantity, lower (resp. upper) case boldface indicates a vector (resp. matrix) quantity. $.^T$ and $.^H$ represent respectively the transpose and the transpose conjugate operators. Finally, $|.|$ represents the cardinal operator, $\left\|.\right\|_p$ the L^p norm while $\left\|.\right\|_F$ represents the Frobenius norm.

II. SYNTHETIC APERTURE RADAR

Airborne and spaceborne SAR aims to provide images of Earth's surface at radar frequency bands. Contrary to optical imaging, they can work day and night using an active radar that transmits and receive pulses in the scanning area. They offer an opportunity to monitor changes and anomalies on the Earth's surface. Their principle is to combine multiple received signals coherently to simulate an antenna with a larger aperture. This procedure allows building a complex-valued image of the terrain with very high range and azimuth resolutions, independently of the distance between the radar and the imaged area. Electromagnetic waves can also be polarized at emission and reception. Horizontal and vertical polarizations are generally used, resulting in four coherent SAR channels HH, HV, VH, and VV where the first and second letters correspond respectively to the emission reception polarization. The information provided by polarimetry is crucial for most geoscience applications from SAR remote sensing data [25].

A. SPECKLE

The main issue in SAR imaging concerns the corruption of the backscattered signal within a resolution cell by a multiplicative noise called speckle. This is due to the coherent summation of multiple scatterers in one resolution cell, which may cause destructive or constructive inferences. This phenomenon disturbs the exploitation of radar data for detection and geoscience applications. Much work in the literature has proposed reducing its effect (e.g. [18], [19], [20], [21], [22], [23], [24]) while preserving the resolution as much as possible.

A model of the speckle has been defined by Goodman [26]. In the case of a single look complex (SLC) image, each pixel power or intensity I may be distributed according to an exponential distribution:

$$p(I|R) = R^{-1} \exp(-I/R), \qquad (1)$$

where R denotes the mean reflectivity level. A useful general statistical model representation is the so-called multiplicative noise model $I = RS$, where S characterizes the exponentially distributed speckle and R is a scalar positive random variable characterizing the texture.

To change the multiplicative noise model into an additive one, a log transformation is often used, which leads to a new distribution:

$$p(y|x) = \exp(y - x) \exp(-\exp(y - x)), \qquad (2)$$

where $y = \log(I)$ and $x = \log(R)$. One crucial detail is the spatial correlation between the pixels which is not always taken into account in the above model. This can be mainly explained by a possible apodization of the images or some oversampling lower than the spatial resolution during the image synthesis. We can directly work on full-resolution data thanks to recent developments in deep learning despeckling algorithms that know how to keep the spatial resolutions and preserve details such as lines and small bright targets (boats or vehicles, for example). They, therefore, render detection tasks and false alarm regulation easier on speckle-free data.

B. DATASET

In this paper, the analyzed dataset is composed of a time series of X-band SAR images (each of size $4800 \times 30000 \times 4$) acquired by SETHI, the airborne instrument developed by ONERA [27], [28]. The resolutions of these images are about 20 cm in both azimuth and range domains for the four polarization channels.

In the monostatic case, the channels VH and HV are often averaged because they contain the same information. The averaging decreases the speckle impact without degrading the resolution (so-called reciprocity principle [29]). The resulting three channels $HH, \frac{1}{2}(HV + VH), VV$ are then thresholded separately using a value $\lambda(I)$ chosen as:

$$\lambda(I) = \mu_I + 3\,\sigma_I, \qquad (3)$$

where μ_I and σ_I are respectively the estimated mean and the standard deviation of the intensity image I. The threshold $\lambda(I)$ is the same for all SAR images to have a consistent visualization, as shown in Fig. 1.
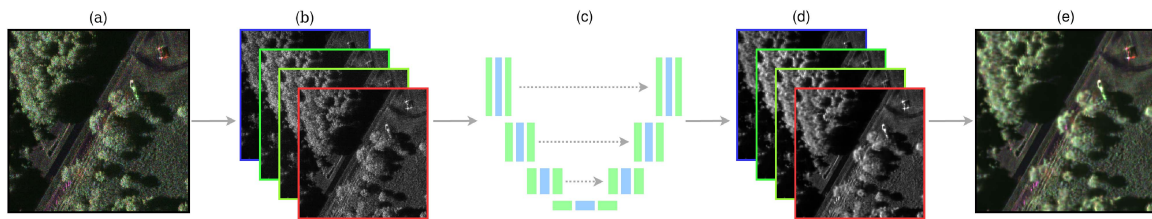
FIGURE 1. First, a polarimetric SAR image (a) is decomposed into four different images (b), discarding the phase. Each image is then denoised separately (c) with the network described in [22]. The four denoised images (d) are finally combined into a fully denoised polarimetric SAR image (e).

III. DESPECKLING

To reduce the speckle impact, the deep-learning despeckling algorithm SAR2SAR [22] has been used. As discussed previously, this method can estimate the spatial correlation between pixels and thus works with full-resolution images.

The training phase is based on a method called *Noise2Noise* first developed in [30]. It shows that there is no need for a ground truth to train a denoising deep neural network, but only for multiple images of the same area with different noise realizations. A common strategy for estimating the true unknown image is to find, for the quadratic L^2 loss function, an image $\mathbf{z}$ that has the smallest average deviation from the measurements:

$$\underset{\mathbf{z}}{\text{argmin}}\, \mathbb{E}_{\mathbf{y}}\left[\|\mathbf{z}-\mathbf{y}\|_2^2\right] = \mathbb{E}_{\mathbf{y}}[\mathbf{y}], \qquad (4)$$

where $\mathbf{y}$ represents a set of observations. If the sample noise is additive, centered, and independent in all the observations, estimating $\mathbb{E}_{\mathbf{y}}[\mathbf{y}]$ is the same as predicting noise-free observations.

The same principle has been applied and refined for SAR images in the algorithm [22]: let $\mathbf{y}_1$ and $\mathbf{y}_2$ be two independent realizations of identically distributed random variables and let a denoising network $f_{\boldsymbol{\theta}}(.)$ be parameterized by $\boldsymbol{\theta}$. A loss L_{speck} based on the negative log-likelihood of log-intensity images can be used according to the distribution defined in (2) and leads to:

$$L_{speck} = \sum_k f_{\boldsymbol{\theta}}([\mathbf{y}_1]_k) - [\mathbf{y}_2]_k + \exp\left([\mathbf{y}_2]_k - f_{\boldsymbol{\theta}}([\mathbf{y}_1]_k)\right) \quad (5)$$

The loss is a sum of all partial loss for each pixel represented with the index k. The use of (5), instead of a mean square error, allows a faster convergence.

The other difference between SAR2SAR and Noise2Noise is the training phase which is decomposed into three steps:

- *Phase A:* The training step is done with a speckle-free image $\mathbf{y}$. Two fake noisy image are created such that $\mathbf{y}_1 = \mathbf{y} + \mathbf{s}_1$ and $\mathbf{y}_2 = \mathbf{y} + \mathbf{s}_2$ where $\mathbf{s}_1$ and $\mathbf{s}_2$ are both distributed according the distribution from (2). The network is learning how to denoise data without any knowledge of spatial correlation.
- *Phase B:* The training step is performed with experimental SAR images and a change compensation pre-processing: two same areas acquired at a different time $\mathbf{y}_{t_n}$ and $\mathbf{y}_{t_m}$ are used. To compensate for changes that could have occurred between the acquisition at

two dates, an estimation of the reflectivity is done beforehand (on sub-sampled images to remove spatial correlation). This leads to $\mathbf{y}_1 = \mathbf{y}_{t_n}$ and $\mathbf{y}_2 = \mathbf{y}_{t_m} - f_{\boldsymbol{\theta}}(\mathbf{y}_{t_m}^{sub}) + f_{\boldsymbol{\theta}}(\mathbf{y}_{t_n}^{sub})$ where the *sub* exponent characterizes sub-sampled data. This finally allows $\mathbf{y}_2$ to have the speckle of $\mathbf{y}_{t_m}$ and the reflectivity of $\mathbf{y}_{t_n}$.

- *Phase C:* The same operation is done, but this time, the network has learned how to model spatial correlation, so the images are not sub-sampled to estimate a reflectivity beforehand. This leads to $\mathbf{y}_1 = \mathbf{y}_{t_n}$ and $\mathbf{y}_2 = \mathbf{y}_{t_m} - f_{\boldsymbol{\theta}}(\mathbf{y}_{t_m}) + f_{\boldsymbol{\theta}}(\mathbf{y}_{t_n})$.

IV. IMAGE RECONSTRUCTION

Once the pre-processing step has been performed, the goal is to highlight anomalies. This is obtained in an unsupervised manner, and the network has no information about what should be an anomaly and what shouldn't. This makes the task harder than, for example, with a supervised convolution network. But, if we can reach the state-of-the-art with an unsupervised algorithm, one of the most significant drawbacks of artificial intelligence, corresponding to the need for a high-quality labeled dataset will be overcome.

A. PROPOSED ARCHITECTURE

We use an AAE [31] with convolution layers to deal with this problem. To understand the architecture, it is first necessary to introduce what is a Generative Adversarial Network (GAN) [32]. We will use the notations $\mathcal{E}(.)$, $\mathcal{D}(.)$ and $\mathcal{D}_c(.)$ to define the encoder, the decoder, and the discriminator respectively. A description of the network is illustrated in Fig. 2.

1) GENERATIVE ADVERSARIAL NETWORKS

A GAN is a self-supervised deep learning algorithm that was originally used to generate synthesized images based on what it saw. A generator and a discriminator are trained jointly against each other. The generator's goal is to create the most realistic image possible, based on a source vector of low dimension. Usually, this vector is distributed according to a multivariate Normal distribution. The generated image should fool the discriminator into thinking there is no difference between real and fake images. In opposition, the discriminator's goal is to be able to distinguish fake generated data from real ones. In our architecture, the role of the GAN is to generate a latent vector that is distributed according to a
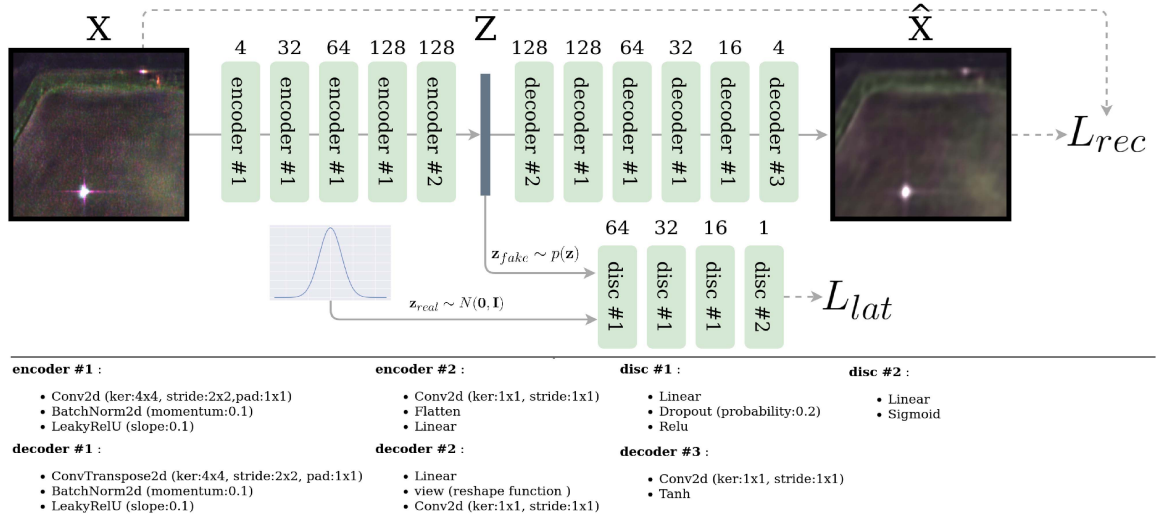
FIGURE 2. Architecture of the Adversarial Autoencoder. The numbers above blocks represent the number of input channels for the encoder and the number of output channels for the decoder and the discriminator (disc). The description of each block is done with Pytorch notations. The input **X** is a denoised patch of a SAR image, **Z** is the latent vector, and the output $\hat{\mathbf{X}}$ is the reconstructed patch. L_{rec} and L_{lat} are respectively (6) and (7).

Normal distribution representative of a patch of polarimetric SAR speckle-free image.

2) ADVERSARIAL AUTOENCODER

An AAE is composed of two things: (i) a generator, which is an encoder and a decoder, one after another, and (ii) a discriminator that is placed between the encoder and the decoder. His goal is to ensure that the latent space follows a normal distribution.

Our input is the logarithm of a denoised polarimetric SAR patch $\mathbf{X} \in \mathbb{R}^{h \times w \times c}$ of size $h \times w$ and depth c, where h and w are respectively the numbers of pixels in azimuth and range, and where c is the polarization channel. The logarithm operation is used to reduce the dynamic of the data (for example, there can be a difference of 10^3 between the amplitude of a strong scatterer and the amplitude of a pixel located in a vegetation area). After passing through the encoder, we get a latent vector $\mathbf{z} = \mathcal{E}(\mathbf{X})$ with $\mathbf{z} \sim p(\mathbf{z})$ and where $p(.)$ is the *a priori* distribution of our encoder. Based on this vector, the decoder will then make an estimation of our input patch $\hat{\mathbf{X}} = \mathcal{D}(\mathbf{z})$. In our architecture, we can observe a GAN present with the generator $\mathcal{E}(.)$ and the discriminator $\mathcal{D}_c(.)$ whose purpose is to differentiate $\mathbf{z}_{real} \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ from $\mathbf{z}_{fake} \sim p(\mathbf{z})$. Here, the latent space is distributed according to a reduced and centered Normal distribution, but the Uniform distribution on [0,1] could also have given similar results. One restriction is that it could not have been a long tail distribution because the information would not have been compacted in a small enough space. We then train the AAE in two successive phases:

1) *Reconstruction error:* this characterizes a loss that ensures to have a low pixel-per-pixel error. We minimize L_{rec} through L^1 norm:

$$L_{rec} = \frac{1}{h\,w\,c} \sum_{i,j,k} \left\| \mathbf{X}_{i,j,k} - \hat{\mathbf{X}}_{i,j,k} \right\|_1. \qquad (6)$$

2) *Regularization error:* such a loss allows us to control the distribution $p(\mathbf{z})$. It also gives a better reconstruction according to [17], [31]. To do so, the weights of the encoder $\mathcal{E}$ and the discriminator $\mathcal{D}_c$ are estimated according to:

$$\min_{\mathcal{E}} \max_{\mathcal{D}_c} L_{lat} \,\widehat{=}\, \mathbb{E}_{\mathbf{z}_{real} \sim \mathcal{N}(\mathbf{0},\mathbf{I})} \left[\log(\mathcal{D}_c(\mathbf{z}_{real})) \right]$$
$$+ \mathbb{E}_{\mathcal{E}(\mathbf{X}) \sim p(z)} \left[\log(1 - \mathcal{D}_c(\mathcal{E}(\mathbf{X}))) \right]. \qquad (7)$$

3) LINK BETWEEN ANOMALY DETECTION AND AAE

It is not apparent to see the link between this architecture and anomaly detection. The goal of an AAE for this application will be to make an accurate estimation for Normal distributed data and a bad estimation for abnormal data. This will be based on the assumption that the network is not powerful enough to reconstruct every aspect of the image. Only the recurrent patterns will be remembered. By definition, a spatial anomaly occurs rarely compared to the rest of the data. If all these assumptions are valid, only rare patterns will be seen as anomalies, which is exactly what we want.

V. CHANGE DETECTION METHOD

Once the network has delivered an estimation of the input image, the goal is to detect changes between $\mathbf{X}$ and $\hat{\mathbf{X}}$. This detection will be represented in an anomaly map $A^{\mathbf{X}} \in [0, 1]^{h \times w}$. The closer the value is to one, the more the pixel is likely to be an anomaly. The global flow of the detection procedure is shown in Fig. 3.

A. PROBLEM FORMULATION

There are multiple approaches to detect changes between two or more images [33]. As described in [33], pixel-level comparisons are widely used in SAR imaging community. The
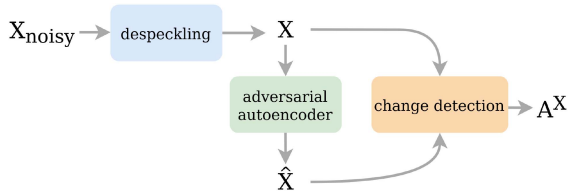
FIGURE 3. Overview of our proposed architecture for SAR anomaly detection. X_{noisy} corresponds to the amplitude of SAR image, X is the speckle-free version, and $\hat{X}$ is its reconstruction using an adversarial autoencoder. Finally, the anomaly map A^X is obtained through a change detection algorithm between X and $\hat{X}$.

FIGURE 4. Comparison of anomaly maps clipped at the highest 2% values: (left) Input image, (middle) L_1 loss between X and $\hat{X}$ and (right) A_E^X.

FIGURE 5. Comparison between (top left): *HH* SAR image, (top-right): its speckle-free version. (bottom-left): the intensity ratio of these images. (bottom-right): superposition of the histogram of the ratio image and the theoretical Exponential distribution with parameter $\lambda = 1$.

problem is formulated within the set of the following two hypotheses:

$$\begin{cases} H_0 : \boldsymbol{\theta}_1 = \boldsymbol{\theta}_2 \text{ (no anomaly)}, \\ H_1 : \boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2 \text{ (anomaly)}, \end{cases} \quad (8)$$

where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are vectors of the estimated parameters of the distribution used to model pixel values. Methods relying on the hypothesis test on the statistics of the image are difficult to exploit. The non-Gaussianity, the heterogeneity of the SAR images, or their complex-valued nature make this derivation very difficult.

When pixels have been transformed into log intensity for better contrast and when the speckle effect has been reduced beforehand, the final distributions of the proposed anomaly detection tests are generally unknown. Hence, to be able to compare all the proposed strategies, a visualization process is used to threshold each map according to the following clipping operation:
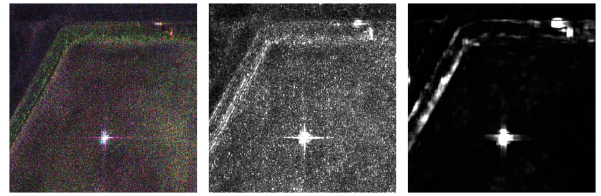
$$\tilde{x} = \min(x, t). \quad (9)$$

For a given percentage p, the threshold t is fixed such that $p\%$ of the pixels are above it. This ensures to have the same number of pixels of value t for each result. The dynamic is compressed in a way that allows us to see at the same time the anomalies and the background. The Probability of False Alarm (PFA) is here, for convenience, characterized by this value p in the sense that p and PFA are equal if all the pixels in the map correspond to H_0 hypothesis. They will have the same significance in the sequel.
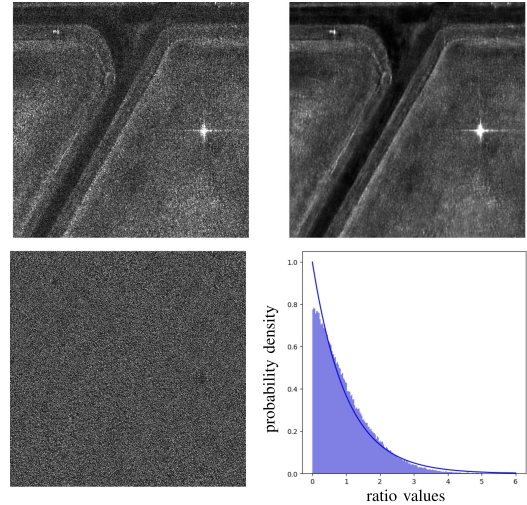
One way to detect changes between two SAR images consists of testing the equality between the two estimated covariance matrices of the corresponding pixel under test for each pixel. This can be made statistically if the knowledge of the data statistic is known [34] or through matrix distances [35].

B. COVARIANCE ESTIMATION
Covariance matrices are estimated locally around the pixel under test. This is useful to strongly reduce the noise in the anomaly map compared to a standard L^1 loss, as it is shown in Fig. 4. Indeed, to estimate a covariance matrix, we use a sliding window represented by a boxcar $\mathcal{B}_{k,l}$ where k, l is the coordinate of its center. For multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, the Maximum Likelihood Estimators of the mean vector $\boldsymbol{\mu}$ and the covariance matrix $\boldsymbol{\Sigma}$ lead to the well known Sample Mean Vector (SMV) and the Sample Covariance Matrix (SCM) which are defined as:

$$\hat{\boldsymbol{\mu}}_{k,l}^{\mathbf{X}} = \frac{1}{|\mathcal{B}_{k,l}|} \sum_{i,j \in \mathcal{B}_{k,l}} \mathbf{X}_{i,j}, \quad (10)$$

$$\hat{\boldsymbol{\Sigma}}_{k,l}^{\mathbf{X}} = \frac{1}{|\mathcal{B}_{k,l}|} \sum_{i,j \in \mathcal{B}_{k,l}} \left(\mathbf{X}_{i,j} - \hat{\boldsymbol{\mu}}_{k,l}^{\mathbf{X}} \right) \left(\mathbf{X}_{i,j} - \hat{\boldsymbol{\mu}}_{k,l}^{\mathbf{X}} \right)^T, \quad (11)$$

where $\hat{\boldsymbol{\mu}}^{\mathbf{X}}_{k,l}$ is the estimate of the mean vector $\boldsymbol{\mu}$, where $\hat{\boldsymbol{\Sigma}}^{\mathbf{X}}_{k,l}$ is the estimate of the covariance matrix $\boldsymbol{\Sigma}$ associated to the boxcar $\mathcal{B}_{k,l}$ in the image $\mathbf{X}$.

C. DISTANCE METRIC
There are multiple possibilities to compute a distance between two matrices [35], [36]. A common way to do this consists in computing the square of the Frobenius norm of the difference between the two matrices.

$$A_E^{\mathbf{X}}(k, l) = \left\| \hat{\boldsymbol{\Sigma}}_{k,l}^{\mathbf{X}} - \hat{\boldsymbol{\Sigma}}_{k,l}^{\hat{\mathbf{X}}} \right\|_F^2. \quad (12)$$

Other methods, based on the Frobenius norm, could also be applied on log-matrices or root matrices. In our detection case, a lot of importance is given to the intensity difference
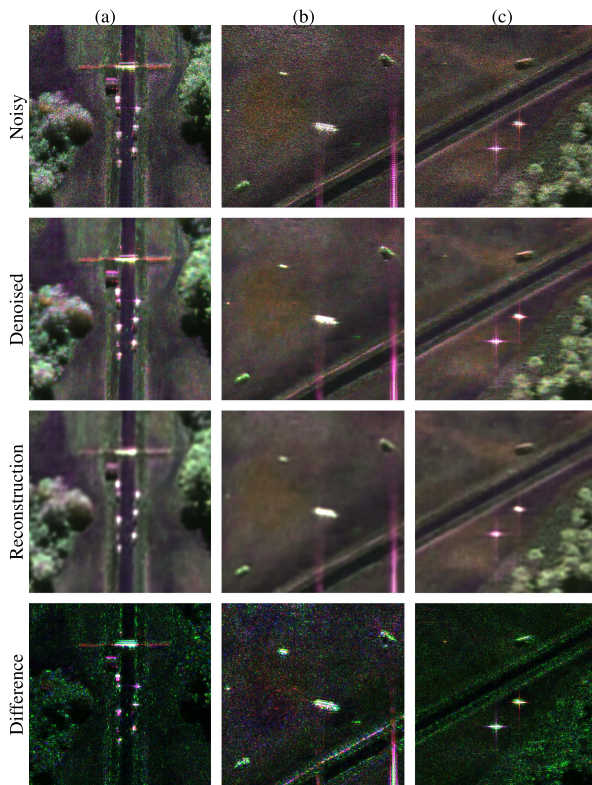
FIGURE 6. Evolution of the different steps of the process for three different areas of the SAR image. (left): (a) SAR image containing a fence and vehicles. (b) SAR image containing vehicles and trihedral sidelobes corresponding to purple lines. (c) SAR image containing three vehicles.
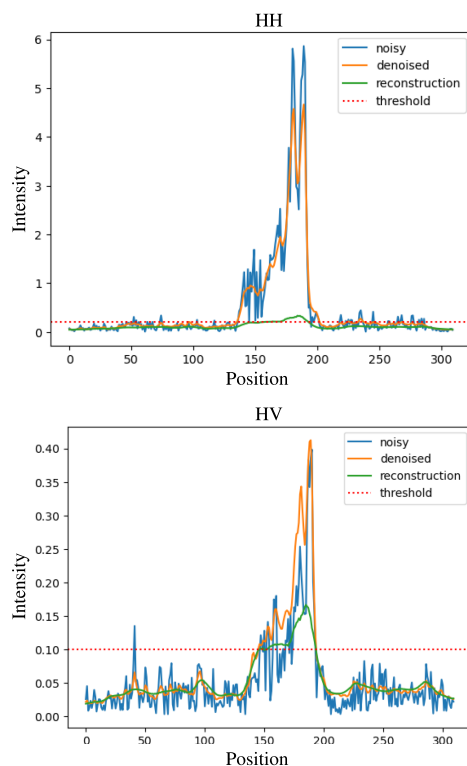
FIGURE 7. Comparisons of horizontal profiles on a strong scatterer of the image (a) in Fig. 6 in *HH* and *HV* polarization.

Algorithm 1: Anomaly Detection Pseudocode in Python Style.

```
# SAR2SAR: despeckling network
# AAE: reconstruction network
# k: semi-kernel size
# scm: compute the sample covariance
 matrix
# norm: Frobenius norm
for x_noisy in eval_data:
  x = SAR2SAR(x_noisy)
  x_rec = AAE(x)
  (_,m,n) = x.shape #
   channels,height,width
  ano_map = zeros(m,n)
  # Compute anomaly score for each
   pixel
  for i in range(m):
    for j in range(n):
      # Crop to compute scm
      x_c = x[:,i-k:i+k+1,j-k:j+k+1]
      x_r_c =
       x_rec[:,i-k:i+k+1,j-k:j+k+1]
      c, c_r = scm(x_c),scm(x_r_c)
      ano_map[i,j] = norm(c -
       c_r).pow(2)
  m,M = min(ano_map), max(ano_map)
  ano_map = (ano_map - m) / (M - m)
  save(ano_map)
```

between $\hat{\Sigma}^X$ and $\hat{\Sigma}^{\hat{X}}$. The Euclidean metric (12) highlights the difference in intensity between the two covariance matrices characterizing each pixel and its reconstructed value while preserving a low PFA. The pseudo-code for the proposed anomaly detection method is detailed in Algorithm 1.

VI. EXPERIENCES AND ANALYSIS

In this section, we experiment with the proposed algorithm on the ONERA SETHI SAR dataset. First, the analysis of the de-speckling network is presented. We then evaluate the change detection method and compare it with a standard approach.

A. DESPECKLING QUALITY

The proposed algorithm is applied to images that are decomposed into patches of size 256×256.

For phase A, we first need to compute a speckle-free image by averaging co-registered images acquired at different dates. Every polarization has been averaged, and we then used the algorithm MuLoG [21] to remove the remaining speckle. It gives us training data of 3052 patches grouped in batches of size 32, which leads to a total of 955 batches. The network is then trained for 20 epochs.

Phases B and C use images from the dataset previously described, and the pre-processing step is the one described in Phase B and C, Section III. There are four piles of two images,
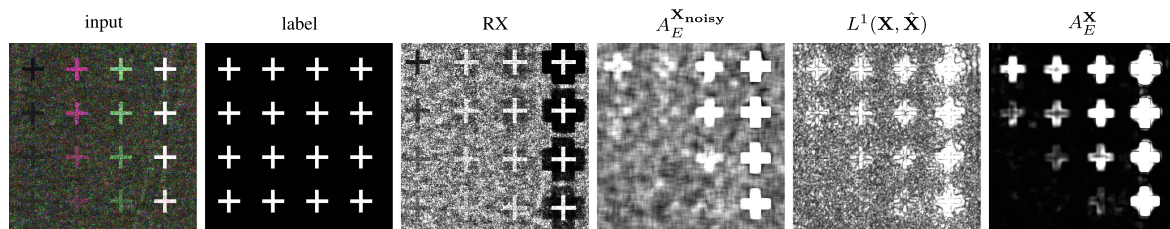
FIGURE 8. Comparison between different anomaly maps on a real SAR image with fake anomalies and its corresponding label.

one for each polarization, and each image is decomposed into 33047 patches with a batch size of 32, which leads to a total of 1032 batches. The network is trained for 10 epochs in phases B and C.

Fig. 5 presents the result of the despeckling process for HH polarization, and the corresponding ratio:

$$r(k, l) = I_s(k, l)/I_{ds}(k, l), \qquad (13)$$

where $I_s(k, l)$ and $I_{ds}(k, l)$ represent the intensity of the original image and the speckle-free image respectively for the pixel at position (k, l).

This test image, relatively heterogeneous, comprises vegetation, roads, thin lines, and a strong scatterer annotated as a vehicle. The image dynamic has not been altered, and because there is no structure in the ratio image, we may confidently affirm that our despeckling network has succeeded in only removing the speckle.

The speckle on its own should follow an exponential distribution with its parameter equal to one. In the histogram in Fig. 5, we found that, as expected, the actual distribution follows an exponential probability density function.
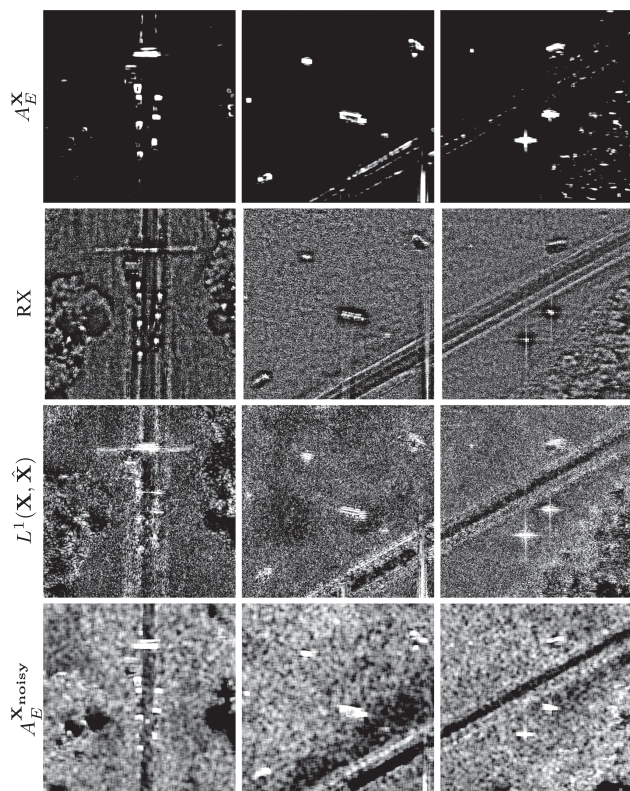
B. ANOMALY DETECTION

To assess the quality of our results, they are compared with the so-called Reed-Xiaoli detector [7] on complex-valued SAR images. The algorithm consists in estimating locally, around a pixel characterized by its polarimetric response $\mathbf{x}$, the associated mean vector $\hat{\boldsymbol{\mu}}$ and the covariance matrix $\hat{\boldsymbol{\Sigma}}$ of its surrounding background and then to test if this pixel under test is belonging, or not, to this background. The anomaly score is computed through the well-known Mahalanobis distance:

$$RX(\mathbf{x}) = (\mathbf{x} - \hat{\boldsymbol{\mu}})^H \hat{\boldsymbol{\Sigma}}^{-1} (\mathbf{x} - \hat{\boldsymbol{\mu}}), \qquad (14)$$

The parameters are locally estimated through the Gaussian SMV and SCM estimates. An exclusion window prevents the use of anomalous data in parameter estimation (guard cells). The semi-kernel size is 12 for the estimation window and 8 for the exclusion window. We define the semi-kernel size as the value k for a kernel of size $2k + 1$.

Additional comparisons will be made using our proposed AAE algorithm without the despeckling pre-processing and the L^1 norm between $\mathbf{X}$ and $\hat{\mathbf{X}}$.

To run the proposed algorithm, we first need to train the AAE on the dataset described in Section II.II-B

FIGURE 9. Comparison between different anomaly maps obtained from the images described in Fig. 6.
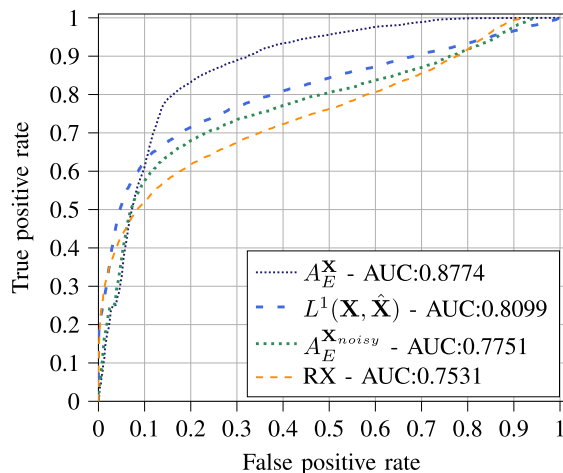
FIGURE 10. ROC curve and value of the AUC for the anomaly detection algorithms displayed Fig. 8.

1) AAE TRAINING AND ARCHITECTURE

The training process is self-supervised. The training dataset has no labels and potentially can contain anomalies. In this case, the evaluation and training datasets can be mixed because the training process requires only unlabeled data. In the same way, as for the despeckling algorithm, a log transformation is used, and the input data are normalized between 0 and 1. These images are decomposed into patches of size 64×64. The sliding window has a stride of 16. This leads to a total of 163785 patches grouped in batches of size 128, so we get 1279 batches for one training epoch with a total of 20 epochs.

The architecture of the network is illustrated in Fig. 2. It has been designed according to [37]. To update the weights, we use the optimizer Adam [38] with a cyclical learning rate [39] that goes linearly from 10^{-3} to 10^{-2}. It takes 2558 batches to go from one value to another. A complete cycle will take four epochs. This method helps to have a robust training phase that will converge even if we do not know the perfect learning rate for the network. For the change detection method, a semi-kernel of size 5 is used for the SCM estimation (detailed in Algorithm 1).

2) EVALUATION DATASET

The proposed method has been qualitatively evaluated on a known area of the dataset described in Section II.II-B. Since there is no ground truth available, the boundary between the anomaly and the background is not apparent, which makes annotation almost impossible in many cases. The area is composed of known anomalies (vehicles) intentionally placed by the ONERA during a measurement campaign.

3) QUALITATIVE RESULTS

The evaluation of noisy images is displayed in Fig. 6 with their denoised and reconstructed versions. The map named *Difference* represents the absolute value between Denoised and Reconstruction maps. We remark that the reconstructed images are blurry, but all the essential structures are kept.

The despeckling pre-processing is not removing any structure, even the smallest one, like the green line in the bottom right of the image (b). In all the reconstructed images, the intensity of each point-like anomaly is greatly reduced. Fig. 7 illustrates this fact and shows a comparison of horizontal profiles at the location characterizing a strong scatterer of image (a). Red dotted lines correspond to the visualization threshold defined in (3). This intensity attenuation is fundamental to the algorithm because the change detection is only applied between the denoised and the reconstructed image in order to highlight potential anomalies.

Fig. 9 presents a comparison of anomaly detection results obtained with different methods. All the results are here clipped at the highest 1% values, see (9).

The anomaly maps obtained with the methods $A_E^{\mathbf{X_{noisy}}}$, the Reed-Xiaoli RX detector and $L^1(\mathbf{X}, \hat{\mathbf{X}})$ have a similar drawback characterizing the presence of a large number of false detections. The proposed method $A_E^{\mathbf{X}}$ is shown to have better performance for the same PFA.

4) QUANTITATIVE RESULTS

For quantitative evaluation, we have embedded synthetic test patterns with different intensity levels in a true anomaly-free crop SAR image.

The fake anomaly map results are represented in Fig. 8. The values are based on true intensities characterizing the dataset with a decreasing significance from top to bottom. In this figure, anomaly detection maps are clipped at 10% of the higher value because some anomalies have a value close to the background. This makes the detection harder; thus, we reduce the dynamic range to have a better representation. One advantage of autoencoder-based detection is detecting abnormal areas of low intensities (left cross), contrary to the RX detector. The proposed $A_E^{\mathbf{X}}$ method outperforms all the other ones and even for the detection of low-intensity patterns. This is illustrated in Fig. 10, which shows the overall performance obtained in terms of Area Under the Curve (AUC). The proposed $A_E^{\mathbf{X}}$ still has the better performance.

VII. CONCLUSION

In this article, we propose a novel anomaly detection algorithm. It is mainly based on the use of adversarial autoencoders. They have not been widely used in the SAR community because of the speckle noise, which dramatically degrades algorithm performance. Thanks to the recent advances in deep learning despeckling algorithms specifically developed for SAR images, we can now efficiently develop new algorithms based on these methods to enhance anomaly detection performance.

Our proposed algorithm outperforms the conventional Reed-Xiaoli method since it can detect abnormal areas with low-intensity values. Because this self-supervised strategy does not require labeled data, it can easily be extended to another type of data as long as the anomaly quantity remains negligible.

The envisaged perspectives concern the improvement of change detection techniques that can help in giving better detection performance as well as in regulating the False Alarm Rate.

REFERENCES

[1] D. W. J. Stein, S. G. Beaven, L. E. Hoff, E. M. Winter, A. P. Schaum, and A. D. Stocker, "Anomaly detection from hyperspectral imagery," *IEEE Signal Process. Mag.*, vol. 19, no. 1, pp. 58–69, Jan. 2002.

[2] S. Matteoli, M. Diani, and G. Corsini, "A tutorial overview of anomaly detection in hyperspectral images," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 25, no. 7, pp. 5–28, Jul. 2010.

[3] A. Alpers, B. Holt, and K. Zeng, "Oil spill detection by imaging radars: Challenges and pitfalls," *Remote Sens. Environ.*, vol. 201, pp. 133–147, 2017.

[4] M. D. Graziano, M. Grasso, and M. D'Errico, "Performance analysis of ship wake detection on Sentinel-1 SAR images," *Remote Sens.*, vol. 9, no. 11, 2017, Art. no. 1107.

[5] W. D. Fisher, T. K. Camp, and V. V. Krzhizhanovskaya, "Anomaly detection in Earth dam and levee passive seismic data using support vector machines and automatic feature selection," *J. Comput. Sci.*, vol. 20, pp. 143–153, 2017.

[6] I. Scollar, A. Tabbagh, A. Hesse, and I. Herzog, *Archaeological Prospecting and Remote Sensing*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[7] I. S. Reed and X. Yu, "Adaptive multiple-band CFAR detection of an optical pattern with unknown spectral distribution," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 38, no. 10, pp. 1760–1770, Oct. 1990.

[8] J. Frontera-Pons, M. A. Veganzones, S. Velasco-Forero, F. Pascal, J.-P. Ovarlez, and J. Chanussot, "Robust anomaly detection in hyperspectral imaging," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Quebec, Canada, 2014, pp. 4604–4607.

[9] M. A. Veganzones, J. Frontera-Pons, F. Pascal, J.-P. Ovarlez, and J. Chanussot, "Binary partition trees-based robust adaptive hyperspectral RX anomaly detection," in *Proc. IEEE Int. Conf. Image Process.*, 2014, pp. 5077–5081.

[10] E. Terreaux, J.-P. Ovarlez, and F. Pascal, "Anomaly detection and estimation in hyperspectral imaging using RMT tools," in *Proc. IEEE 6th Int. Workshop Comput. Adv. Multi-Sensor Adaptive Process.*, Cancun, Mexico, 2015, pp. 169–172.

[11] J. Frontera-Pons, M. A. Veganzones, F. Pascal, and J.-P. Ovarlez, "Hyperspectral anomaly detectors using robust estimators," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 2, pp. 720–731, Feb. 2016.

[12] A. W. Bitar, L.-F. Cheong, and J.-P. Ovarlez, "Sparse and low-rank matrix decomposition for automatic target detection in hyperspectral imagery," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 8, pp. 5239–5251, Aug. 2019.

[13] Y. Haitman, I. Berkovich, S. Havivi, S. Maman, D. G. Blumberg, and S. R. Rotman, "Machine learning for detecting anomalies in SAR data," in *Proc. IEEE Int. Conf. Microw., Antennas, Commun. Electron. Syst.*, Tel-Aviv, Israel, 2019, pp. 1–5.

[14] S. Akçay, A. Atapour-Abarghouei, and T. P. Breckon, "Skip-GANomaly: Skip connected and adversarially trained encoder-decoder anomaly detection," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, 2019, pp. 1–8.

[15] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-AnoGAN: Fast unsupervised anomaly detection with generative adversarial networks," *Med. Image Anal.*, vol. 54, pp. 30–44, May 2019.

[16] S. Sinha et al., "Variational autoencoder anomaly detection of avalanche deposits in satellite SAR imagery," in *Proc. 10th Int. Conf. Climate Inform.*, 2020, pp. 113–119.

[17] S. Mabu, S. Hirata, and T. Kuremoto, "Anomaly detection using convolutional adversarial autoencoder and one-class SVM for landslide area detection from synthetic aperture radar images," *J. Robot., Netw. Artif. Life*, vol. 8, no. 2, pp. 139–144, 2021.

[18] C.-A. Deledalle, L. Denis, and F. Tupin, "Iterative weighted maximum likelihood denoising with probabilistic patch-based weights," *IEEE Trans. Image Process.*, vol. 18, no. 12, pp. 2661–2672, Dec. 2009.

[19] C.-A. Deledalle, L. Denis, and F. Tupin, "NL-InSAR: Nonlocal interferogram estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1441–1452, Apr. 2011.

[20] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, "NL-SAR: A unified nonlocal framework for resolution-preserving (Pol)(In)SAR denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2021–2038, Apr. 2015.

[21] C.-A. Deledalle, L. Denis, S. Tabti, and F. Tupin, "MuLoG, or how to apply Gaussian denoisers to multi-channel SAR speckle reduction?," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4389–4403, Sep. 2017.

[22] E. Dalsasso, L. Denis, and F. Tupin, "SAR2SAR: A semi-supervised despeckling algorithm for SAR images," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 14, pp. 4321–4329, 2021.

[23] E. Dalsasso, L. Denis, M. Muzeau, and F. Tupin, "Self-supervised training strategies for SAR image despeckling with deep neural networks," in *Proc. IEEE 14th Eur. Conf. Synthetic Aperture Radar*, Leipzig, Germany, 2022, pp. 1–6.

[24] E. Dalsasso, L. Denis, and F. Tupin, "As if by magic: Self-supervised training of deep despeckling networks with MERLIN," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, pp. 1–13, 2022, Art. no. 4704713.

[25] J. S. Lee and E. Pottier, *Polarimetric Radar Imaging: From Basics to Applications*, Boca Raton, FL, USA: CRC Press: 2009.

[26] J. W. Goodman, "Some fundamental properties of speckle," *J. Opt. Soc. Amer.*, vol. 66, no. 11, pp. 1145–1150, Nov. 1976.

[27] R. Baqué, P. Dreuillet, and H. Oriot, "SETHI: Review of 10 years of development and experimentation of the remote sensing platform," in *Proc. IEEE Int. Radar Conf.*, 2019, pp. 1–5.

[28] S. Angelliaume, X. Ceamanos, F. Viallefont-Robinet, R. Baqué, P. Déliot, and V. Miegebielle, "Hyperspectral and radar airborne imagery over controlled release of oil at sea," *Sensors*, vol. 17, no. 8, Aug. 2017, Art. no. 1772.

[29] L. Pallotta, "Reciprocity evaluation in heterogeneous polarimetric SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022, Art. no. 4000705.

[30] J. Lehtinen et al., "Noise2Noise: Learning image restoration without clean data," in *Proc. 35th Int. Conf. Mach. Learn.*, 2018, vol. 80, pp. 2965–2974.

[31] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," in *Proc. Int. Conf. Learn. Representations*, San Juan, Puerto Rico, 2016, pp. 1–16.

[32] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, pp. 139–144, Oct. 2020.

[33] A. Mian, G. Ginolhac, J.-P. Ovarlez, A. Breloy, and F. Pascal, "An overview of covariance-based change detection methodologies in multivariate SAR image time series," in *Chapter 3 in Change Detection and Image Time Ser. Anal., Unsupervised Methods, Vol. 1, ISTE/Wiley Encyclopedia of Sci. - Remote Sens. Imagery*, A. M. Atto, F. Bovolo, and L. Bruzzone, Eds., 2021.

[34] L. M. Novak, "Coherent change detection for multi-polarization SAR," in *Proc. IEEE Conf. Rec. 39th Asilomar Conf. Signals, Syst. Comput.*, 2005, pp. 568–573.

[35] W. Förstner and B. Moonen, *Geodesy-The Challenge of the 3rd Millennium* (A Metric for Covariance Matrices, Book Chapter) E. W. Grafarend, F. W. Krumm, and V. S. Schwarze, Eds. Berlin, Heidelberg, Germany: Springer, 2003.

[36] I. L. Dryden, A. Koloydenko, and D. Zhou, "Non-euclidean statistics for covariance matrices, with applications to diffusion tensor imaging," *Ann. Appl. Statist.*, vol. 3, no. 3, pp. 1102–1123, 2009.

[37] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. Int. Conf. Learn. Representations*, 2016, pp. 1–16.

[38] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Representations*, San Diego, CA, USA, 2015, pp. 1–15.

[39] L. N. Smith, "Cyclical learning rates for training neural networks," in *Proc. Winter Conf. Appl. Comput. vis. IEEE*, Santa Rosa, CA, 2017, pp. 464–472.

MAX MUZEAU (Graduate Student Member, IEEE) received the Engineering degree in signal processing and computer vision from Télécom Saint-Etienne, Saint-Étienne, France, and the M.S. degree in advanced imaging and material appearance from University Jean-Monnet, Saint-Étienne, in 2021. He is currently working toward the Ph.D. degree with the ONERA (The French Aerospace Lab) and Centralesupélec in the SONDRA Laboratory. His research interests include remote sensing, deep learning, computer vision, signal processing, and their applications to earth observation, especially synthetic aperture radar imagery.

CHENGFANG REN (Member, IEEE) received the Agrégation degree in applied physics and the M.Sc. degrees in electrical engineering from École Normale Supérieure de Cachan, France, in 2011 and 2012, respectively, and the Ph.D. degree from Paris-Sud XI University, France, in 2015. From 2015 to 2016, he was a Postdoctoral Research Scientist with the GIPSA Laboratory, Saint Martin d'Hères, France. He is currently an Assistant Professor with SONDRA, the French-Singaporean Laboratory of CentraleSupĺec, France. His research interests include statistical signal processing, and machine and deep learning for radar applications.

SÉBASTIEN ANGELLIAUME received the Engineering degree from the École Nationale Supérieure d'Ingénieur de Constructions Aéronautiques, Toulouse, France, in 2003, and the M.S. degree from the École Nationale Supérieure d'Electrotechnique, d'Electronique, d'Informatique, d'Hydraulique et des Télécommunications, Toulouse, in 2003. Since 2006, he has been with the Electromagnetism and Radar Department, Office National d'Études et de Recherches Aérospatiales (ONERA - The French Aerospace Lab), Salon de Provence, France. He has been involved in the ONERA synthetic aperture radar airborne platform RAMSES and SETHI, developing science applications. Since 2020, his research at ONERA focuses mainly on low-frequency SAR remote sensing.

MIHAI DATCU (Fellow, IEEE) received the M.S. and Ph.D. degrees in electronics and telecommunications from the University Politehnica 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. From 1992 to 2002, he had a longer Invited Professor Assignment with the Swiss Federal Institute of Technology (ETH Zurich), Zurich, Switzerland. He was a Visiting Professor with the University of Oviedo, Oviedo, Spain; University Louis Pasteur; International Space University, Strasbourg; University of Siegen, Siegen, Germany; University of Innsbruck, Innsbruck, Austria; University of Alcala, Alcala de Henares, Spain; University Tor Vergata, Rome, Italy; Universidad Pontificia de Salamanca, Madrid, Spain; University of Camerino, Camerino, Italy; and Swiss Center for Scientific Computing, Manno, Switzerland. Since 1981, he has been a Professor with the Department of Applied Electronics and Information Engineering, Faculty of Electronics, Telecommunications and Information Technology, UPB. Since 1993, he has been a Scientist with German Aerospace Center (DLR), Cologne, Germany. 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. Since 2011, he has been leading the Immersive Visual Information Mining Research Laboratory, Munich Aerospace Faculty. He is the Director of the Research Center for Spatial Information, UPB. 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, Telecom Paris, Paris, France, a collaboration of DLR with the French Space Agency (CNES), Paris. He has been a Professor with the DLR-CNES Chair, ParisTech, Paris Institute of Technology, Telecom Paris. He has initiated the European frame of projects for image information mining (IIM) and is involved in research programs for information extraction, data mining and knowledge discovery, and data understanding with the European Space Agency (ESA), NASA, and in a variety of national and European projects. He and his team have developed and are developing the operational IIM processor in the Payload Ground Segment systems for the German missions, TerraSAR-X, TanDEM-X, and the ESA Sentinel-1 and Sentinel-2. He is a Senior Scientist with the Remote Sensing Technology Institute, DLR, and was involved in the DLR-ONERA Joint Virtual Center for AI in Aerospace. His research interests include explainable and physics-aware artificial intelligence, smart sensor design, and quantum machine learning with applications in earth observation. Dr. Datcu is a Member of the ESA Working Group Big Data from Space and Visiting Professor with ESA's Φ-Lab. He was the recipient of the Best Paper Award and IEEE Geoscience and Remote Sensing Society Prize, in 2006, National Order of Merit with the rank of Knight, for outstanding international research results, awarded by the President of Romania, in 2008, and Romanian Academy Prize Traian Vuia for the development of the SAADI image analysis system and his activity in image processing, in 1987. He was also the recipient of the Chaire d'Excellence Internationale Blaise Pascal 2017 for international recognition in the field of data science in earth observation and IEEE GRSS David Landgrebe Award in recognition of his 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 was a Co-organizer for international conferences and workshops and the Guest Editor of a special issue on AI, Big Data, or Quantum ML of the IEEE and other journals.

JEAN-PHILIPPE OVARLEZ (Member, IEEE) was born in Denain, France, in 1963. He received the Engineering degree jointly from École Supérieure d'Electronique Automatique et Informatique, Paris, France, the Diplôme d'Etudes Approfondies degree in signal processing from the University of Paris XI, Orsay, France, in 1987, and the Ph.D. degree in physics from the University of Paris VI, Paris, in 1992. In 2011, he obtained a Research Directorship Habilitation thesis in signal processing from the University of Paris-Sud, France, and his qualification for an University Professor position. In 1992, he joined the Electromagnetic and Radar Division, French Aerospace Lab (ONERA), Palaiseau, France, where he is currently the Research Director and a Member of the Scientific Committee, ONERA Physics Domain. Since 2008, he has been attached part-time to CentraleSupélec SONDRA Lab, in charge of Signal Processing activities supervision. His research interests include statistical signal processing for radar and SAR applications, such as time-frequency, imaging, detection, and parameter estimation. In 2015, he became a Member of a Special Area Team of Theoretical and Methodological Trends in Signal Processing, EURASIP, and Treasurer of the IEEE GRSS French Chapter in 2016.