Change Detection Based on Convolutional Neural Networks Using Stacks of Wavelength-Resolution Synthetic Aperture Radar Images

João G. Vinholi [ID], Bruna G. Palm [ID], *Member, IEEE*, Danilo Silva [ID], *Member, IEEE*, Renato Machado [ID], *Senior Member, IEEE*, and Mats I. Pettersson [ID], *Senior Member, IEEE*

Abstract— This article presents two supervised change detection algorithms (CDAs) based on convolutional neural networks (CNNs) that use stacks of coregistered wavelength-resolution (WR) synthetic aperture radar (SAR) images to detect changes in an image under monitoring. The additional information of a scene of interest provided by SAR image stacks can be explored to enhance the performance of CDAs. In particular, stacks of images with similar statistics can be obtained for ultrawideband (UWB) very high-frequency (VHF) SAR systems, as they produce images highly stable in time. The proposed CDAs can be summed up into four stages: difference image (DI) formation, semantic segmentation, clustering, and change classification. The CNN-GSP algorithm is based on a ground scene prediction (GSP) image, which is used as a reference to form a DI. A CNN-based model then analyzes the DI. The CNN-MDI algorithm feeds multiple DIs with identical monitored images to a CNN-based model, which will concurrently analyze their features. Tests with CARABAS-II data show that the proposed CDAs can outperform other state-of-the-art algorithms that also use stacks of WR-SAR images. Beyond that, the proposed algorithms outperformed a CNN-based CDA that does not use image stacks, which shows that CNN-based algorithms can use the additional information provided by stacks of SAR images to reduce false alarm occurrences while increasing the probability of detection of changes.

Index Terms— CARABAS-II, classification, convolutional neural network (CNN), deep learning, remote sensing, semantic segmentation.

I. INTRODUCTION

SYNTHETIC aperture radar (SAR) data is an essential tool in remote sensing applications [1] due to its capability of providing suitable visual information acquisition, independent of weather and illumination conditions, and wide terrain coverage in a short observation time [2]. Generally, SAR data analysis and processing include tasks such as: 1) image modeling [3], [4]; 2) identification and classification of distinct ground type [5], [6]; and 3) change detection [7], [8].

Especially, the use of statistical models and machine learning techniques can be applied as venues with accurate results for the above SAR-related challenges, as presented in [9], [10], [11], and [12].

In particular, change detection algorithms (CDAs) aim at identifying changes in the same scene obtained at different times, and it is an important research topic of remote sensing understanding [13]. Furthermore, changes identification is commonly applied to civil and military applications, such as disaster management [14], urban planning [15], and human-made target detection [16], [17]. Specifically, ultrawideband (UWB) very high-frequency (VHF) SAR systems are high-resolution systems widely employed for natural disaster monitoring, foliage-penetrating applications, and detection of concealed targets [18]. These systems can be denoted as wavelength-resolution (WR) SAR systems [19], [20], [21], [22], [23] since they are characterized by a large fractional bandwidth and a wide antenna bandwidth wavelength of the transmitted pulses [24].

A critical characteristic of WR-SAR systems is that the speckle noise does not significantly influence the acquired images [25]. This characteristic exists since only a single scatterer is present in the resolution cell. In other words, small scatterers in the ground area of interest do not contribute to the backscattering for low-frequency WR radar systems. Consequently, small structures, such as tree branches and leaves, are not shown in the SAR image [26]. Moreover, large scatterers are associated with low-frequency components and, therefore, they tend to be less influenced by environmental effects and are stable in time. Hence, an image package with similar statistics can be obtained through multipasses with identical heading and incidence angle at a given ground area [21]. SAR image stack is widely studied for SAR systems with high resolution (see [27], [28], [29]); however, the use of image stacks in WR-SAR images for change detection applications is a topic of interest that needs to be more deeply explored.

An example of a low-frequency WR UWB VHF SAR system is the CARABAS-II. This system has a large fractional bandwidth that operates at 20–86-MHz, horizontal (HH) polarization [30], [31]. An SAR image dataset was made public in [32]. The image scene is located in northern Sweden

João G. Vinholi and Renato Machado are with the Department of Telecommunications, Aeronautics Institute of Technology (ITA), São José dos Campos 12228 900, Brazil (e-mail: jgvinholi@gmail.com; rmachado@ita.br).

Bruna G. Palm is with the Department of Telecommunications, Aeronautics Institute of Technology (ITA), São José dos Campos 12228 900, Brazil, and also with the Department of Mathematics and Natural Sciences, Blekinge Institute of Technology (BTH), 371 79 Karlskrona, Sweden (e-mail: bruna.palm@bth.se).

Danilo Silva is with the Department of Electrical and Electronic Engineering, Federal University of Santa Catarina (UFSC), Florianópolis 88040 900, Brazil (e-mail: danilo.silva@ufsc.br).

Mats I. Pettersson is with the Department of Mathematics and Natural Sciences, Blekinge Institute of Technology (BTH), 371 79 Karlskrona, Sweden (e-mail: mats.pettersson@bth.se).

Digital Object Identifier 10.1109/TGRS.2022.3211010

and is the same for all 24 available images. It was considered an incidence angle of 58°. Each image has 25 military vehicles placed in the ground scene in four different target deployments and three imaging geometries [30], [31]. The changes related to this dataset are related to the target's location and orientation for each deployment. Efficient change detection methods of targets concealed by forests are widely explored in the literature (see [21], [22], [23], [30], [31], [33], [34], [35]). All these studies consider as input a difference image (DI), obtained by a simple subtraction between two images (reference and monitored image). However, more information can be obtained using image stacks of the CARABAS-II dataset aiming at improving CDAs.

Additionally, another possible venue for change detection in UWB WR-SAR images is the use of a supervised method. With that, empirical-based knowledge about the sought targets can be considered as prior information for obtaining the changes in the evaluated images [17], [23].

Thus, in this article, we propose two supervised convolutional neural network (CNN)-based CDAs, named CNN-MDI and CNN-ground scene prediction (GSP), that use stacks of WR-SAR images. The stacks provide additional information about the scene that is explored to achieve better performance indicators. The first approach aims to maximize the detection performance at the cost of higher numerical complexity. In contrast, the latter aims to achieve good performance indicators while maintaining a lower numerical cost. The proposed CDAs search for positive changes in the monitored image and can be summarized into four stages: 1) DI formation; 2) semantic segmentation; 3) clustering; and 4) classification of changes. The CNN-GSP uses a GSP image, produced by fusing a WR-SAR stack into a single image, as a reference for change detection, which is then fed to the CDA presented in [36]. On the other hand, the CNN-MDI algorithm uses the images contained in a WR-SAR data stack as references to generate multiple DIs with identical monitored images. These DIs are then concurrently analyzed by a CNN-based model to locate relevant changes.

Tests executed with CARABAS-II data suggest that the proposed algorithms can outperform other state-of-the-art CDAs for WR-SAR images. Also, the algorithms outperformed the best-performing CNN-based CDA for pairs of WR-SAR images, showing that the additional information provided by stacks of WR-SAR images enabled the CNNs to distinguish false alarms from real changes better. In a direct comparison between the two proposed CDAs, the CNN-MDI yielded higher performance indicators. However, the CNN-GSP is less numerically intensive, considering that it requires fewer neural network evaluations per prediction.

A. Motivations

The fact that WR-SAR systems operate at much lower frequencies when compared to more usual SAR bands, like C and X, brings some challenges for change detection tasks. The transmitted pulses do not reflect on small scatterers, and thus the produced images do not show finer details of the textures and objects present in the scene of interest. Beyond that, the

only available dataset of WR-SAR images, CARABAS-II, has a small number of images that share multiple similarities, e.g., the geographical location and the position and shapes of targets in the scene. For these reasons, using traditional deep CNN models with hundreds of thousands of parameters would undoubtedly lead to overfitting and an unnecessary high numerical complexity. Moreover, such CNN models have high receptive fields, leading them to learn the peculiarities of the training dataset, e.g., the gridlike disposition of changes, leading to more overfitting. To the best of our knowledge, no other publication addresses deep learning-based CDAs that use stacks of WR SAR images. Thus, we present two CDAs that consider all the peculiarities of these systems. We present the performance evaluation over the only publicly available dataset for this class of SAR system. Multiple CNN architectures were not tested since the main objective of this article is to evaluate how much the use of stacks of WR-SAR images can improve the performance of a baseline CNN-based CDA for WR-SAR images, presented in [36].

B. Contributions

The main contributions of this article are as follows.

1) An investigation of whether using stacks of multiple WR-SAR images to detect changes in a scene of interest with CNN-based algorithm results in performance improvements compared to when only two images are used per detection.
2) The proposal of two CNN-based CDAs that use stacks of WR-SAR images. One aims to reach the highest performance indicators, at the cost of higher numerical complexity. In contrast, the other tries to improve performance while keeping the numerical complexity as close as possible to when only two images are used per detection. Both algorithms extend on the structure of the best-performing CNN-based CDA for WR-SAR images, presented in [36].
3) Achieve, by an expressive margin, the best-performance indicator in any scenario where WR-SAR images were used to detect changes.
4) The execution of a statistical evaluation of the obtained results to quantify how likely the performance improvements obtained in the executed tests represent real gains and, thus, how likely are stacks of WR-SAR able to improve the performance of CNN-based CDAs.

II. RELATED WORKS

An alternative to enhancing CDAs' performance (CDA) performance for WR-SAR images is considering more than two images as input. For example, a study using a small stack (three images) was employed in a CDA with a noise canceller approach [21]. In [37], an image stack considering eight images of the CARABAS-II dataset is applied to obtain a GSP image. In the same paper, the GSP was used as a reference image of an ordinary CDA, which was enough to improve the performance of the algorithm. Both methods were based on pixel-by-pixel DIs and resulted in a good tradeoff between detection probability and false alarm rate (FAR).

More recently, Ramos et al. [38] proposed a CDA based on robust principal component analysis, using image stacks to detect changes in WR SAR images. As a result, the FAR was reduced, keeping the probability of detection high.

A study using a small stack (three images) was employed as a noise canceller algorithm in a CDA for CARABAS-II dataset in [21]. In [37], an image stack considering eight images of the CARABAS-II dataset is applied to obtain a GSP image. This image was used as a reference image in a CDA based on subtractions and thresholding operations. The collection of images was considered to eliminate clutter and noise in the reference image, improving detection performance. Both methods were based on pixel-by-pixel DIs and resulted in a good tradeoff between detection probability and FAR.

CDAs for UWB WR-SAR images based on traditional signal processing approaches, such as likelihood ratio test [33], generalized likelihood ratio test [39], Bayes' theorem [35], noise-canceling [19], can be benefited from the use of image stack since more information is available to deal with two main challenges: 1) noise fluctuations over multitemporal images and 2) elongated structures reported within the images, which are the main source of false alarms [19]. On the other hand, supervised methods can also be an alternative to overcome this issue. Supervised methods can excel in CDAs since they explore the prior information for obtaining an exhaustive description of the changes in the evaluated images [17]. More precisely, such algorithms are trained based on sliding windows that comprise both each pixel under test and their neighborhood, acquiring empirical-based knowledge about the sought targets [23].

In [34], a CDA based on a logistic regression model was considered to detect the military vehicles in UWB VHF WR-SAR images, showing that the FAR can be significantly reduced, even without a stage for threshold selection. Another approach to improve change detection in VHF WR-SAR images is the use of CNNs, due to the great potential for learning complex spatial patterns and its stability. Since its feature extraction process is based on collecting visual features from the images, this approach is extensively employed for object classification in optical images [23]. Furthermore, in SAR systems using higher frequency bands, CNNs have been employed for change detection [40] and object classification [41].

In [23], [36], and [42], CNNs were used for change detection in UWB WR-SAR images. In particular, Campos et al. [23] considers a CNN as a filter to reduce the occurrence of false alarms, whereas Vint et al. [42] proposes the creation of a synthetic dataset of targets with a generative adversarial network (GAN), which are subsequently fed to a CNN-based classifier. The CDA presented in [36] has reached high-performance indicators in tests executed over the CARABAS-II dataset. Given the peculiarities of this dataset and the small number of available images, popular end-to-end segmentation models—usually made of hundreds of thousands of trainable parameters—may be unable to generalize well enough to reach acceptable performance with data acquired by different WR-SAR systems. For this reason, the architecture of this CDA has been custom-made to be robust to such problem.

Both Campos et al. [23] and Vinholi et al. [36] have reached superior performance in terms of probability of detection and FAR, when compared with other approaches not based on CNNs, evidencing how CNNs have the power to distinguish real changes from false alarms and learn complex patterns in SAR DIs. In [43], a combination of image stacks and supervised algorithms is introduced for VHF SAR ground surveillance. However, to the best of our knowledge, the use of CNNs for WR-SAR image stacks change detection is not addressed in the literature and this article aims to propose a first treatment.

III. PRELIMINARIES

A. CARABAS-II Dataset

The dataset considered in this article was obtained from CARABAS-II, a Swedish UWB VHF SAR system. The images are available in [32] and are fully discussed in [30] and [31]. The dataset is composed of 24 coregistered magnitude single-look SAR images, where each image cover a scene of size 2 km × 3 km, and have almost no speckle noise, since the CARABAS-II is a low-frequency WR system. As reported in [30], the spatial resolution of CARABAS-II is approximately 2.5 m in azimuth and 2.5 m in range.

The dataset can be split into stacks composed of images captured with the same flight angle and has backscattering stable in time; hence, only target changes are expected within the image stacks. The stacks have images with the same flight geometry but associated to four different targets' deployments (missions 1–4) in the ground scene. Stack 1 has the images associated with passes 1 and 3; Stack 2 to passes 2 and 4; and Stack 3 to passes 5 and 6.

Each image corresponds to an area of 6 km² and is defined as a matrix of 3000 × 2000 pixels, in which each pixel size is 1 m × 1 m. The ground scene of the images is mostly composed of boreal forest with pine trees; fences, power lines, and roads are also present in the scene. Additionally, military vehicles were deployed in the SAR scene and placed uniformly, in a manner to facilitate their identifications in the tests [31]. Each image shows 25 targets, obscured by foliage, of three different sizes: 1) ten small vehicles with a square design, with dimensions 4.4 × 1.9 × 2.2 m; 2) eight truck-sized vehicles with dimensions 6.8 × 2.5 × 3.0 m; and 3) seven trucks with dimensions 7.8 × 2.5 × 3.0 m [31]. It was considered an incidence angle of 58° for all acquisitions.

For illustration, four images of the CARABAS-II dataset are shown in Fig. 1. The regions where the military vehicles were deployed on the ground scene are highlighted by the blue circles. In Fig. 1(a)–(d), the military vehicles were oriented in a south-western, north-western, south-western, and western headings, respectively. In Fig. 1(a) and (b), the military vehicles were deployed in the top left of the scene, while in Fig. 1(c) and (d), they can be found in the bottom right of the ground scene.

The algorithms proposed in this article are inspired by two previously published studies [36], [37]. The contributions presented in these papers are briefly discussed below.
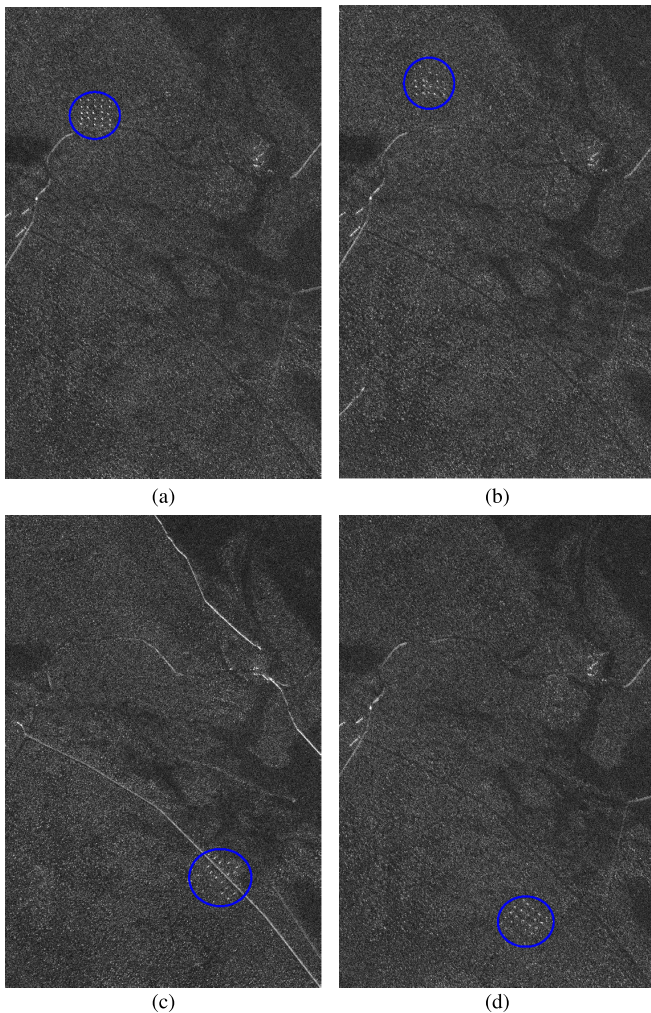
Fig. 1. Example of images from the CARABAS II dataset for the four target orientation headings. The blue circles highlight the regions where the military vehicles were deployed in the ground scenes. (a) South-Western Heading. (b) North-Western Heading. (c) South-Western Heading. (d) Western Heading.

B. Change Detection Based on Convolutional Neural Networks

Vinholi et al. [36] proposed a CDA for WR-SAR images based on CNNs. The presented algorithm consists of a semantic segmentation stage and a classification stage. As illustrated in Fig. 2, the CDA takes as input a DI—obtained by the subtraction of a standalone WR-SAR reference image from a monitored image—and outputs a list containing the central points of all detected changes. The reasons for the choice of an architecture containing two custom built neural networks with a relatively low number of trainable parameters for this problem, instead of just one classical end-to-end segmentation network are twofold.

1) To avoid data generalization issues by greatly reducing the number of trainable parameters of the Segmentation CNN and limiting its overall receptive field. These characteristics are important to assure that the model does not learn from the peculiarities contained in the low number of available WR-SAR images. As can observed in Fig. 1, the targets to be detected as changes in the dataset have many similarities between themselves: they

are all pointlike bright targets, distributed in a 2-D grid geometry. More traditional semantic segmentation architectures, like the U-Net [44], could not be adopted in this scenario without generalization concerns. That is, it would not be expected that the trained model would perform similarly with data from other WR-SAR systems. The high receptive field would lead to learning the geometrical features of the small dataset, whereas the much greater number of parameters trained in a small dataset with similar images would lead to further generalization problems.

2) To improve performance, by employing a more numerically complex classification CNN as a false alarm filter. As shown in [36], this second network has the potential to reduce significantly the number of false alarms for a given probability of detection.

A CNN with few parameters, the so-called Segmentation CNN, searches the DI for pixels that are likely parts of relevant changes, producing a pixelwise semantic segmentation probability map image. Each pixel of this map is a value between 0 and 1, corresponding to the probability that the pixel is part of a relevant change. The map is then binarized by a thresholding operation, where positive pixels—classified as potential changes of interest—are mapped to 1, and negative pixels—classified as clutter—are mapped to 0. In summary, for a pixel at the position (j, k) in the monitored image, the binary segmentation output at position (j, k) is computed as

$$\hat{y}_{j,k} = \begin{cases} 1, & \text{if } \tilde{y}_{j,k} > \omega_1 \\ 0, & \text{otherwise} \end{cases} \tag{1}$$

where $\tilde{y}_{j,k} \in [0, 1]$ is the corresponding CNN output, and ω_1 is the segmentation threshold. Table I shows the architecture of the Segmentation CNN.

The produced binary map is fed to the density-based spatial clustering of applications with noise (DBSCAN) algorithm [45] to locate clusters of positive pixels. This algorithm searches for clusters containing at least $n_{\min}$ neighbor points, while the distance between these neighbors is at most ϵ meters. Small patches centralized at the central points of the found clusters are extracted from the DI and are sent to another CNN—named Classification CNN. The network outputs a number between 0 and 1 for each patch analyzed, representing the probability that a relevant change exists inside the patch. Finally, each patch is classified as containing or not containing a change of interest by a thresholding operation. In other words, the classification decision is defined as

$$\hat{y} = \begin{cases} 1, & \text{if } \tilde{y} > \omega_2 \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

where $\tilde{y}$ is the CNN output after analyzing a patch, and ω_2 is the classification threshold. Table II shows the architecture of the Classification CNN. All hidden layers of both networks are activated by the rectified linear unit (ReLU) function, with the exception of the last ones, activated by the Sigmoid function.

The Segmentation CNN is trained with WR-SAR DIs. The ground truth associated with each DI is composed of binary
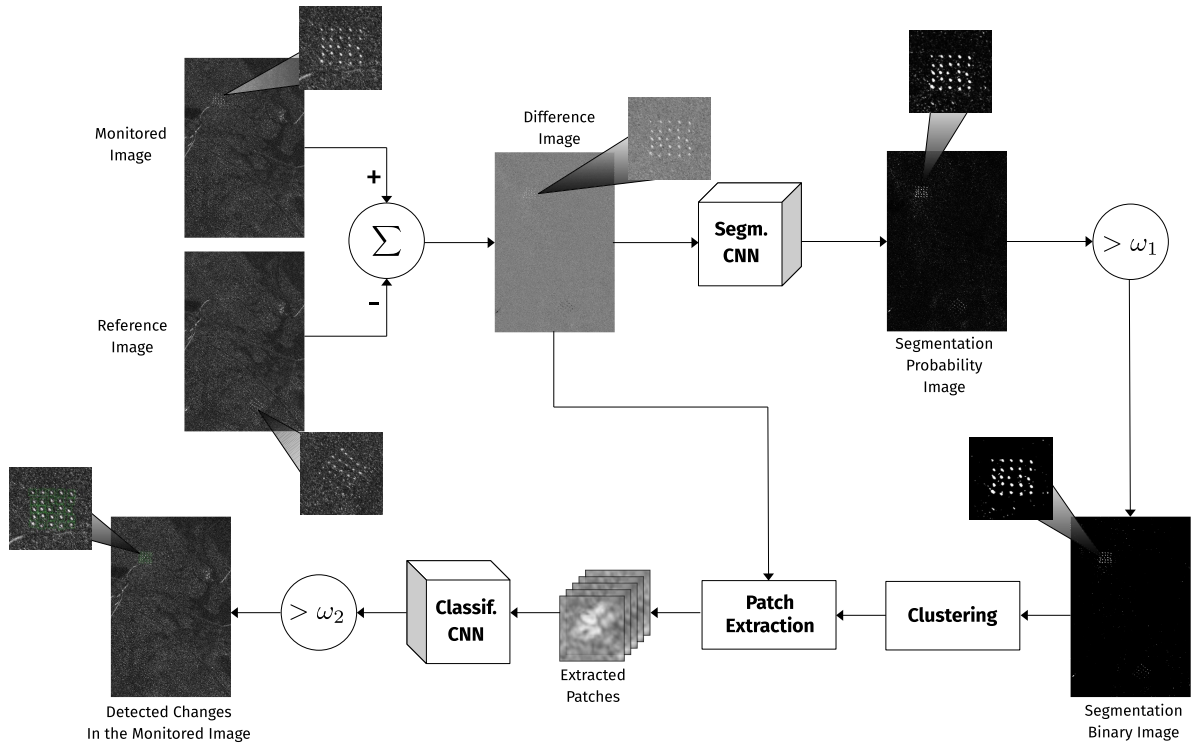
Fig. 2. Summary of the CDA presented in [36]. The inputs are the monitored and the reference images; both are WR-SAR images of the same scene, captured with the same flight geometry. The outputs are the individual positioning of the changes of interest in the monitored image. Rectangles represent the data processing steps, tridimensional boxes represent CNNs, and circles represent elementwise mathematical operations.

TABLE I
ARCHITECTURE OF THE SEGMENTATION CNN

Layer	Dimensions of Each Filter	Number of Filters	Stride	Padding	Size of Output	Num. of Trainable Param.
Input	—	—	—	—	$3000 \times 2000 \times 1$	0
Conv.	$5 \times 5 \times 16$	16	1	4	$3000 \times 2000 \times 16$	416
Conv.	$1 \times 1 \times 16$	16	1	0	$3000 \times 2000 \times 16$	272
Dropout $P = 0.3$	—	—	—	—	$3000 \times 2000 \times 16$	0
Conv.	$3 \times 3 \times 16$	8	1	2	$3000 \times 2000 \times 8$	1160
Conv.	$1 \times 1 \times 8$	1	1	0	$3000 \times 2000 \times 1$	9
						Total: 1857

masks whose dimensions are identical to the dimensions of the original WR-SAR images. Nonzero elements in these masks represent the location of pixels that are parts of relevant changes. For the CARABAS-II dataset, squares of 1's were centered at the central points of appropriate changes—vehicles concealed under foliage—contained in the DIs. The TGB11 vehicles were marked as 3×3 squares, while both the TGB30 and TGB40 vehicles were marked as 5×5 squares.

The classification training set consists of small patches of 34×34 pixels extracted from the WR-SAR DIs. These patches are big enough to cover typical changes present in WR-SAR DIs. On the one hand, negative patches—those containing no relevant changes—are extracted by moving a sliding window over the DIs. On the other hand, positive patches are extracted as patches centered on the central points of relevant changes. Values of 0 and 1 are, respectively, assigned to each negative and positive patch. Classification data augmentation took place to increase the number of positive examples and avoid over-fitting the training data. Also, the addition of Gaussian noise and rotation were performed over some extracted patches.

The area occupied by targets of interest present in WR-SAR images is usually small compared to the total size of the illuminated scene. This result has a severe imbalance in both networks. To address this issue, both CNNs are trained with the Balanced Focal Loss cross-entropy function [46], which has been known for speeding up training and increasing the performance of deep neural networks trained with imbalanced

TABLE II
ARCHITECTURE OF THE CLASSIFICATION CNN

Layer	Dimensions of Each Filter	Number of Filters	Stride	Padding	Size of Output	Num. of Trainable Param.
Input	—	—	—	—	$34 \times 34 \times 1$	0
Conv.	$3 \times 3 \times 1$	16	1	2	$34 \times 34 \times 16$	160
Batch Norm.	—	—	—	—	$34 \times 34 \times 16$	64
Conv.	$3 \times 3 \times 16$	16	1	2	$34 \times 34 \times 16$	2320
Max Pool.	$2 \times 2 \times 1$	1	2	0	$17 \times 17 \times 16$	0
Batch Norm.	—	—	—	—	—	64
Conv.	$3 \times 3 \times 16$	32	1	2	$17 \times 17 \times 32$	4640
Max Pool.	$2 \times 2 \times 1$	1	2	1	$9 \times 9 \times 32$	0
Batch Norm.	—	—	—	—	—	128
Conv.	$3 \times 3 \times 32$	64	1	2	$9 \times 9 \times 64$	18496
Max Pool.	$2 \times 2 \times 1$	1	2	1	$5 \times 5 \times 64$	0
Batch Norm.	—	—	—	—	—	256
Conv.	$3 \times 3 \times 64$	64	1	0	$3 \times 3 \times 64$	36928
Aver. Pool.	$3 \times 3 \times 1$	1	3	0	$1 \times 1 \times 64$	0
Dropout $P = 0.3$	—	—	—	—	$1 \times 1 \times 64$	0
Conv.	$1 \times 1 \times 64$	1	1	0	$1 \times 1 \times 1$	65
						Total: 63121

datasets. It is defined as

$$f(p_y) = -\alpha_y (1 - p_y)^\gamma \log(p_y) \qquad (3)$$

where p_y is the probability of the predicted label being equal to the ground truth $y \in \{0, 1\}$, $\alpha_y \in [0, 1]$ is the class imbalance weighting factor corresponding to the label equal to y, and γ is the focusing parameter, used to tune the importance given to harder to learn examples. The focusing parameter was set to $\gamma = 2$ for both networks, while the weighting factors were set to $\alpha_1 = 0.9999$ and $\alpha_0 = 0.0001$ for the Segmentation CNN, and to $\alpha_1 = 0.9$ and $\alpha_0 = 0.1$ for the Classification CNN. These values were selected by evaluating multiple choices of hyperparameters over the validation set and choosing those that yielded the best-performance indicators.

To the best of our knowledge, for the CARABAS-II dataset, the algorithm presented in [36] outperformed every existing CDA at the time of publication, achieving a detection probability of 99% at an FAR of 0.0833/km². Also, higher performance was obtained for most operating points even when the Classification CNN was removed, i.e., when the potential relevant changes localized by the DBSCAN algorithm were assumed to be correct. The obtained results evince that CNNs can learn complex patterns and distinguish real changes from false alarms in WR-SAR DIs.

C. Ground Scene Prediction Based on Image Stacks

Palm et al. [37] considered several different statistical methods, such as median, trimmed mean, and autoregressive models, for GSP in WR SAR images. This article evaluated the performance over the CARABAS-II dataset of a simple CDA consisting of thresholding and morphological operations applied to a DI when using GSP images as reference images. Tests showed how the additional information of the scene provided by stacks of SAR images could help significantly
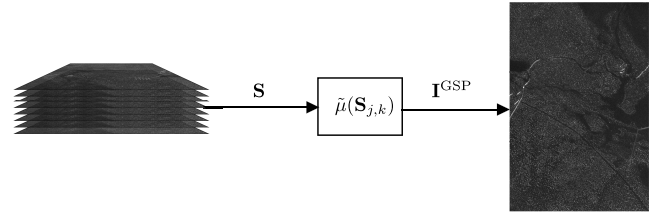

Fig. 3. Definition of the median GSP method [37]. A stack of WR-SAR images $\mathbf{S}$ is processed by the median operator, which generates the GSP image $\mathbf{I}^{\text{GSP}}$. This operation is denoted as $\tilde{\mu}(\mathbf{S}_{j,k})$, for $j = 1, 2, \ldots, H$ and $k = 1, 2, \ldots, W$.

decrease FARs compared to when standalone WR-SAR images are used as reference images.

The median operation yielded the best-performance indicators in the executed tests. Formally, this GSP method is defined as follows. Let $\mathbf{I}_i \in \mathbb{R}^{H \times W}$ be the ith image in a stack of WR-SAR images $\mathbf{S} = [\mathbf{I}^{(1)}, \mathbf{I}^{(2)}, \ldots, \mathbf{I}^{(n)}] \in \mathbb{R}^{H \times W \times n}$. The pixelwise median operation is then applied to $\mathbf{S}$, producing the matrix $\mathbf{I}^{\text{GSP}} \in \mathbb{R}^{H \times W}$. The individual elements $I_{j,k}^{\text{GSP}}$ are given by

$$\begin{aligned} I_{j,k}^{\text{GSP}} &= \tilde{\mu}(\mathbf{S}_{j,k}) \\ &= \tilde{\mu}\left(\left[I_{j,k}^{(1)}, I_{j,k}^{(2)}, \ldots, I_{j,k}^{(n)} \right] \right) \end{aligned} \qquad (4)$$

where $\tilde{\mu}(\cdot)$ is the median operator. This process is illustrated in Fig. 3.

IV. METHODOLOGY

Stacks of WR-SAR images provide additional scene information that could be explored by a CNN-based CDA to better distinguish relevant changes from false alarms. However, to the best of our knowledge, an investigation of whether this information improves the performance of CNN-based CDAs for WR-SAR systems has not been presented in the literature.
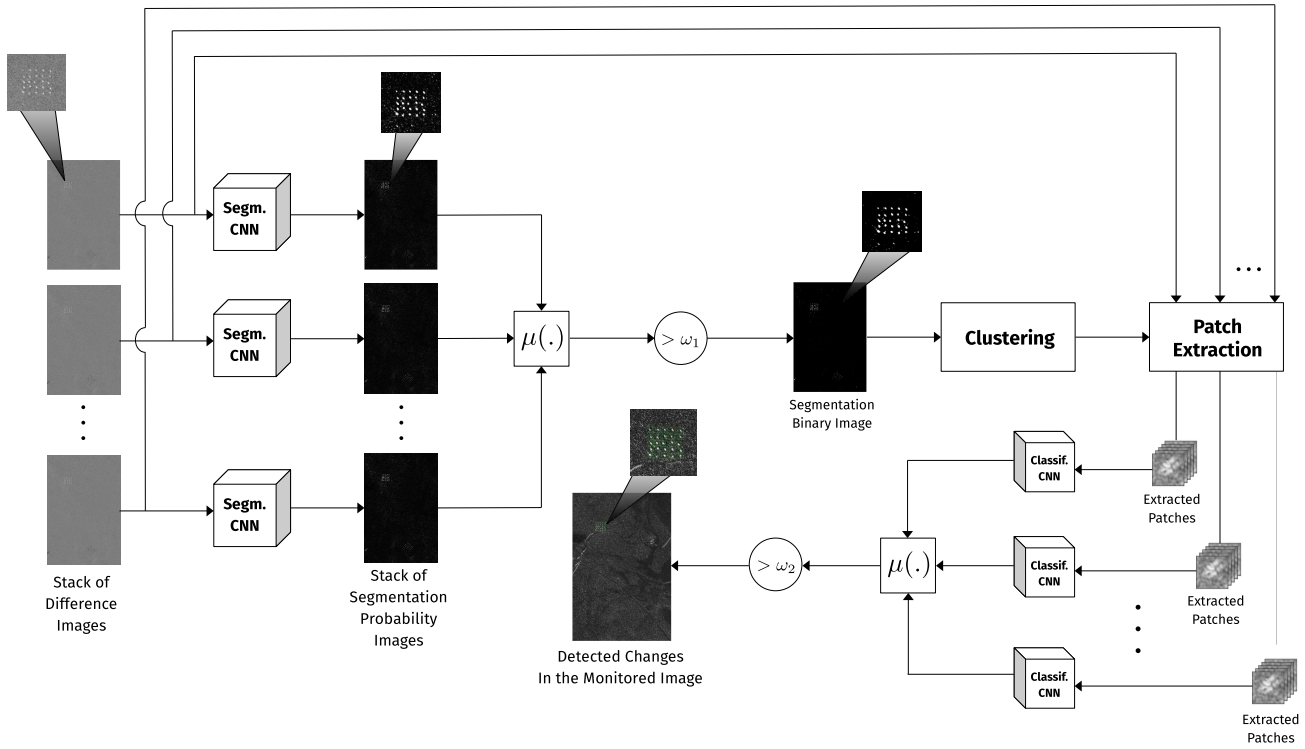
Fig. 4. Illustration of the steps of the CDA proposed in Section IV-B. The input of the algorithm is a stack of WR-SAR DIs with identical monitored images. The outputs are the individual positioning of the changes of interest in the monitored image. Rectangles represent the data processing steps, tridimensional boxes represent CNNs, and circles represent elementwise mathematical operations. The symbol $\mu(.)$ denotes the median operation over the pixels dimension.

For that, we propose two CNN-based CDAs using WR-SAR image stacks.[1]

A. CNN-Based Change Detection Using a Ground Scene Prediction

In this first algorithm, a GSP image is generated with the method discussed in Section III-C and is used as a reference image for the CNN-Based CDA in Section III-B. The sequential steps of this CDA are as follows.

1) *GSP Generation:* The pixelwise median operation is applied to a stack of n coregistered WR-SAR images of the same scene, with identical flight geometry, to generate a GSP image.
2) *DI Formation:* The GSP image is subtracted from a monitored WR-SAR image, forming a DI. The DI is normalized to zero mean and unit variance.
3) *Change Detection:* The CNN-based CDA discussed in Section III-B analyzes the produced DI and outputs a list with the location of the detected relevant changes in the monitored image.

B. CNN-Based Change Detection Using Multiple Difference Images

As illustrated in Fig. 4, this algorithm does not use a GSP image as a reference image. Instead, a stack of WR-SAR images generates MDIs with identical monitored images, each with a different reference image. These DIs are jointly

analyzed by a CNN-based model with a few distinctions from the one used in the GSP-based algorithm; this model outputs a list with the central points of the relevant changes detected in the monitored image. The sequential steps of this CDA are as follows.

1) *DIs Formation:* A stack of n coregistered WR-SAR images of the same scene, with identical flight geometry, is used to form n DIs; each of these DIs is defined by the subtraction of an image contained in the stack from a fixed monitored WR-SAR image. Consequently, all the produced DIs have the same monitored image while the reference image changes. These images are then normalized to zero mean and unit variance.
2) *Fusion of Semantic Segmentation Maps:* The CNN shown in Table I individually analyzes the n normalized DIs, outputting n probability images with the same dimensions of the DIs. Since each DI uses a different WR-SAR image as a reference, the generated probability maps differ from each other. For this reason, they can provide additional information about the scene.

   The n generated images are fused into a single probability image by applying the median operation over the pixels dimension. We have chosen the median over other fusion methods so that outlier pixels do not influence the resulting fusion image. After, a binary segmentation mask is formed by thresholding each value of the fused probability image. The pixels in the monitored image linked to nonzero elements in the binary mask are classified as changes.

TABLE III
SELECTED HYPERPARAMETERS FOR THE PROPOSED CDAS

Segmentation CNN	
Learning Rate Schedule	$l_{r0} = 5 \times 10^{-3}, d_r = 0.97$
Training Epochs	60
Focal Loss	$\alpha_0 = 0.0001$ $\alpha_1 = 0.9999, \gamma = 2$
Classification CNN	
Learning Rate	$l_r = 1 \times 10^{-4}$
Training Epochs	70
Focal Loss	$\alpha_0 = 0.1000$ $\alpha_1 = 0.9000, \gamma = 2$
Clustering	
Min. Number of Points per Cluster	$n_{min} = 8$
Max. Distance Between Points	$\epsilon = 1\,\mathrm{m}$

3) *Localization of Regions of Interest:* The regions in the scene that possibly contain relevant changes are identified; this takes place by applying the DBSCAN clustering algorithm [45] to the binary mask, which searches for clusters of nonzero elements. The central location of the found clusters is assumed to be the central points of the regions of interest (ROIs) in the scene— where relevant changes may be present. Then, small patches centered on the central points of the ROIs are extracted from each of the n DIs, producing n patches per ROI.

4) *Fusion of Classification Outputs:* For each ROI found in the scene, another CNN, whose architecture is identical to the network shown in Table II, analyzes the n extracted patches.

The Segmentation and Classification CNNs of both proposed CDAs share the same architecture and hyperparameters. Adam optimizer [47] and Glorot weight initialization [48] are used by both CNNs. For the Segmentation CNN, the learning rate is adjusted by an exponential decay schedule. That is, the learning rate is initialized as l_{r0} and decays exponentially every epoch by a fixed factor d_r. The Classification CNN is trained with a fixed learning rate. The selected hyperparameters are presented in Table III. To avoid overfitting to the CARABAS-II dataset, we performed the hyperparameter search with a subset of the available data, as explained in Section V-B. Note that altering the dimensions of the input data delivered to the Segmentation CNN would not change its intended behavior, considering that both its input and output have identical dimensions and that it is a fully convolutional network. Meanwhile, the input shape of the Classification CNN has been chosen to be big enough to entirely contain typical changes present in WR-SAR DIs.

V. EXPERIMENTS AND RESULTS

A. Train and Test Data Partitioning

The usual splitting of a dataset into a training set and a test set may not be a good choice for small datasets like the CARABAS-II, since it could lead to highly biased performance indicators [49]. To use all the available data for both training and testing, k-fold cross-validation is applied to the evaluated data. This cross-validation procedure consists of partitioning a dataset into k non-overlapping sets (folds) so that k different train-test data splits can be defined [49]. In each of the splits, one of the k sets is selected as the test set, while the other $k-1$ sets are used as the train set. Since the CARABAS-II dataset has four different target deployments, we set $k = 4$; this way, the dataset is separated into folds (subsets) of monitored images of the same flight mission (target deployment). This splitting criteria guarantees that, for each fold, the target positioning of the test set is not seem in any of the images available in the training set; avoiding, thus, leaking this information to the training data.

Considering that the proposed algorithms have as inputs DIs produced by different procedures, two distinct datasets based on the 24 CARABAS-II WR-SAR images need to be defined. Table IV shows one of the four different train/test data splits, where Fold 4 (highlighted in gray) is selected as the test set, and Folds 1–3 are employed as the train set. The stacks used to produce the GSPs employed as reference images in the CNN-GSP algorithm are described in the third and fourth columns. The fifth column shows the reference images considered in the CNN-MDI algorithm to form the DIs associated with each one of the monitored images.

1) Dataset I: This dataset is used to train and test the CNN-GSP algorithm. It consists of 24 DIs created by the procedure discussed in Section IV-A. As shown in Table IV, the stacks present in the train set are not generated using images from the test set. For this reason, each stack defined in the test set contains seven images, whereas those defined in the train set contain five images. Therefore, the stacks associated with the monitored images change for each one of the selected k-fold train/test split, and the GSPs need to be recreated. This distinction between folds is necessary to avoid data leakage between the test sets and the train sets.

2) Dataset II: This dataset is used for training and testing the CNN-MDI algorithm. It consists of 108 DIs with 24 unique monitored images. As shown in Table IV, the train/test data split is similar to the split of Dataset I, except that now MDIs generated by the steps discussed in Section IV-B are associated with each monitored image. Due to the constraints imposed to avoid data leakage, the folds selected for training contain fewer DIs when compared to the fold selected for testing. Based on tests executed during the development phase, the classification training data (patches) is augmented with random noise addition, random rotation, and random vertical flipping, aiming at improving the performance of the proposed method. The possible angles of rotation are $\{0, \pi/2, \pi, 3\pi/2\}$ radians, selected equiprobably, whereas flipping is performed along the vertical axis with a probability of $1/2$. Noise matrices, whose elements are i.i.d. random variables that follow a normal

TABLE IV
EXAMPLE OF A CROSS-VALIDATION TRAIN/TEST DATA SPLIT. THE TEST SET FOR THIS CASE IS HIGHLIGHTED IN GRAY

Fold	Monitored Image	Stack of Reference Images (CNN-MDI)	Stack Used to Produce the GSP (CNN-GSP)
1 (Train)	M2P1	M3P1, M3P3, M4P1, M4P3	M2P1, M3P1, M3P3, M4P1, M4P3
	M2P2	M3P2, M3P4, M4P2, M4P4	M2P2, M3P2, M3P4, M4P2, M4P4
	M2P3	M3P1, M3P3, M4P1, M4P3	M2P3, M3P1, M3P3, M4P1, M4P3
	M2P4	M3P2, M3P4, M4P2, M4P4	M2P4, M3P2, M3P4, M4P2, M4P4
	M2P5	M3P5, M3P6, M4P5, M4P6	M2P5, M3P5, M3P6, M4P5, M4P6
	M2P6	M3P5, M3P6, M4P5, M4P6	M2P6, M3P5, M3P6, M4P5, M4P6
2 (Train)	M3P1	M2P1, M2P3, M4P1, M4P3	M3P1, M2P1, M2P3, M4P1, M4P3
	M3P2	M2P2, M2P4, M4P2, M4P4	M3P2, M2P2, M2P4, M4P2, M4P4
	M3P3	M2P1, M2P3, M4P1, M4P3	M3P3, M2P1, M2P3, M4P1, M4P3
	M3P4	M2P2, M2P4, M4P2, M4P4	M3P4, M2P2, M2P4, M4P2, M4P4
	M3P5	M2P5, M2P6, M4P5, M4P6	M3P5, M2P5, M2P6, M4P5, M4P6
	M3P6	M2P5, M2P6, M4P5, M4P6	M3P6, M2P5, M2P6, M4P5, M4P6
3 (Train)	M4P1	M2P1, M2P3, M3P1, M3P3	M4P1, M2P1, M2P3, M3P1, M3P3
	M4P2	M2P2, M2P4, M3P2, M3P4	M4P2, M2P2, M2P4, M3P2, M3P4
	M4P3	M2P1, M2P3, M3P1, M3P3	M4P3, M2P1, M2P3, M3P1, M3P3
	M4P4	M2P2, M2P4, M3P2, M3P4	M4P4, M2P2, M2P4, M3P2, M3P4
	M4P5	M2P5, M2P6, M3P5, M3P6	M4P5, M2P5, M2P6, M3P5, M3P6
	M4P6	M2P5, M2P6, M3P5, M3P6	M4P6, M2P5, M2P6, M3P5, M3P6
4 (Test)	M5P1	M2P1, M2P3, M3P1, M3P3, M4P1, M4P3	M5P1, M2P1, M2P3, M3P1, M3P3, M4P1, M4P3
	M5P2	M2P2, M2P4, M3P2, M3P4, M4P2, M4P4	M5P2, M2P2, M2P4, M3P2, M3P4, M4P2, M4P4
	M5P3	M2P1, M2P3, M3P1, M3P3, M4P1, M4P3	M5P3, M2P1, M2P3, M3P1, M3P3, M4P1, M4P3
	M5P4	M2P2, M2P4, M3P2, M3P4, M4P2, M4P4	M5P4, M2P2, M2P4, M3P2, M3P4, M4P2, M4P4
	M5P5	M2P5, M2P6, M3P5, M3P6, M4P5, M4P6	M5P5, M2P5, M2P6, M3P5, M3P6, M4P5, M4P6
	M5P6	M2P5, M2P6, M3P5, M3P6, M4P5, M4P6	M5P6, M2P5, M2P6, M3P5, M3P6, M4P5, M4P6

distribution with expectation $\mu = 0$ and standard deviation $\sigma = 5$, are generated and added to each patch fed to the classification CNN. The dimensions of the generated matrices are identical to those of the training patches. The augmentation happens simultaneously with training and, as a result of the addition of random noise, each training patch fed to the classification CNN is different from patches previously fed.

B. Development and Validation

The development of the proposed CDAs took place by evaluating multiple possibilities of model architectures and hyperparameters and selecting those that yielded the best performance over a set of validation data. To avoid overfitting, this set comprises only the first three flight missions—Mission 2 to 4—leaving out Mission 5 for the final test. Thus, we performed k-fold cross validation with $k = 3$ in the development phase. The folds were defined similar to how they were described in the final test, except that we did not enforce the constraints against data leakage between test and train sets. Thus, the images from the flight mission selected for test in a particular train/test split are also present in the stacks of the train set.

C. Performance Evaluation

In this section, the results of the performance evaluation of both proposed algorithms are presented and discussed. We also compare the obtained results with those of other CDAs that were also tested with the CARABAS-II dataset [21], [36], [37]. Two common choices of performance metrics for SAR change detection tasks are the probability of detection P_d and the FAR [21], [23], [31], [35], [36], [37]. P_d is defined as the ratio between the number of detected targets and the number of targets contained in an image, and FAR is defined as the ratio between the number of false alarms (false positives) and the area under surveillance.

We present in Fig. 5 the performance of the CNN-GSP and CNN-MDI CDAs in terms of receiver operating characteristic (ROC) curves. The points of operation contained in these curves are defined by fixing the segmentation threshold ω_1 while varying the classification threshold ω_2. Therefore, each curve is associated with a fixed ω_1. The selection of an optimal segmentation threshold for both algorithms happened in the development phase by searching for the thresholds that produced the ROC curves with the highest areas under the curve (AUC) values. We calculated the AUCs for FAR $\in [0, 0.8/\text{km}^2]$, and the best-performing segmentation thresholds were 0.5 and 0.575 for the CNN-GSP and CNN-MDI algorithms, respectively.

To fairly compare the performances of the CNN-GSP and CNN-MDI algorithms with the CDA [36], we employed a methodology very similar to the one described in Section V-A. The exception is that only 24 DIs were used, each one
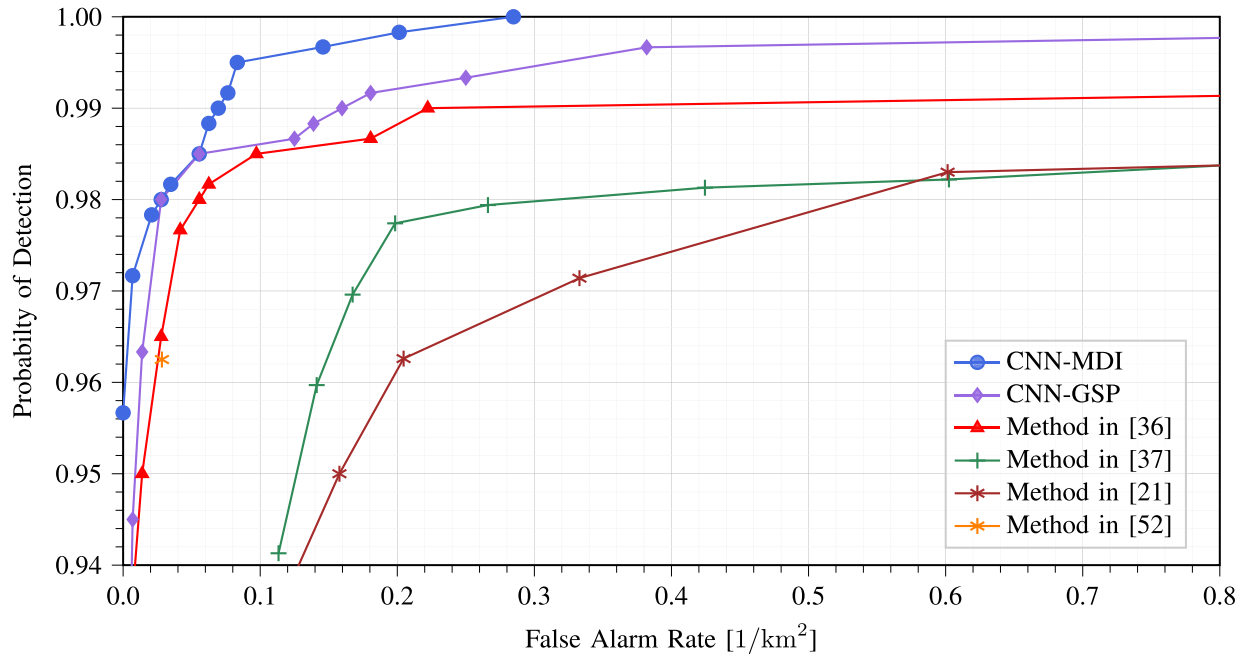
Fig. 5. Comparison of the ROC curves of the CNN-GSP and CNN-MDI algorithms with the curves from [21], [36], [37], and [23]. The proposed algorithms outperform by a good margin the CNN-based methods in [36] and [23], which shows that stacks of WR-SAR images can be explored by CNNs to improve detection performance. The new CDAs also outperformed the methods in [21] and [37], which are based on classic detection techniques and use stacks of WR-SAR images.

associated univocally with a monitored image. This distinction is due to the fact that in [36], only one DI is associated with each monitored image. The reference images linked with each monitored image have been equiprobably selected from all possibilities to form the DIs while assuring that each one has been paired with a reference image captured with the same flight angle. Additionally, the hyperparameter selection is identical to the one shown in Table III. The procedure for the segmentation threshold selection followed the methodology described in Section IV-B, and the optimal ω_1 was set equal to 0.5 aiming at having the highest AUC in the validation stage. Note that this allows us to be certain that no advantage is given to either the stacks-based algorithms or the method introduced in [16]. Beyond that, by doing a fair comparison, it is expected that both the CNN-GSP and CNN-MDI stacks-based algorithms outperform [16] since more information about the scene is available in the proposed algorithms.

Fig. 5 shows that the algorithms CNN-GSP and CNN-MDI outperform the methods presented in [37], [21], [23], and [36] by a considerable margin for the tested points of operation. This suggests that the additional information provided by stacks of WR-SAR images can lead to fewer false alarm occurrences, especially for higher probabilities of detection ($P_d > 0.99$). We also can observe that only the CNN-MDI algorithm reached $P_d = 1$ with a relatively low FAR of $0.285/\text{km}^2$ and achieved a probability of detection higher than 0.9 without any false alarms ($P_d = 0.96$). Despite not showing performance indicators as good as CNN-MDI, the CNN-GSP algorithm displayed a noticeable advantage over the CDA in [36], while requiring fewer neural network evaluations when compared with the CNN-MDI.

At $P_d = 0.99$, for instance, the CNN-GSP and CNN-MDI algorithms, respectively, reached FAR values 28% and 69% smaller than the value reached in [36]. For comparison purposes, we added the ROC curves for two other CDAs based on stacks of WR-SAR images. One of them was presented in [37] and discussed in Section III-C, while the other one is presented in [21]. These two algorithms underperformed the three CNN-based CDAs—CNN-GSP, CNN-MDI, and the algorithm in [36]. However, since these are not machine learning algorithms, they do not need to be trained. Thus, for them, fewer steps are needed to reach a production-ready state.

In general, change detection methods have good performance for most of the CARABAS-II dataset image pairs. However, for few of them, the performance is degraded significantly either in terms of high FAR or undetected rates [19]. One particular example is related to Mission 3 and Pass 5, in which Vu et al. [21] and Palm et al. [37] achieved 16 and 15 detected targets, respectively, and 6 false alarms. On the other hand, the proposed CNN-MDI ($\omega_1 = 0.571$ and $\omega_2 = 0.42$) and CNN-GSP ($\omega_1 = 0.5$ and $\omega_2 = 0.772$) detected 22 and 18 targets, respectively, and had just one false alarm in the same image. These results highlight that the additional information of the scene provided by the image stacks combined with the complex patterns learned by the CNNs can help to significantly decrease FARs, distinguishing real changes from false alarms.

Table V shows the performance per fold of the proposed algorithms for two particular points of operation. The CNN-GSP algorithm is evaluated at $\omega_1 = 0.500, \omega_2 = 0.775$, and the CNN-MDI at $\omega_1 = 0.575, \omega_2 = 0.425$. We can see that fold 4, which is not used in the development phase, reached higher than average performance for both algorithms. This
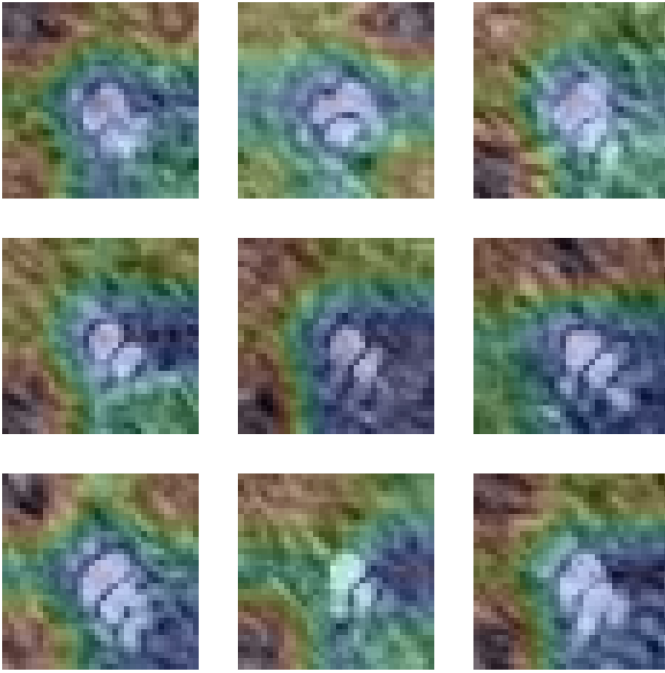
Fig. 6. Heat maps generated with the Grad-CAM [50] technique in nine target-containing patches extracted from CARABAS-II DIs. The algorithm has been executed using the trained classification CNN. The hottest the color of a region, the higher is the influence of that region to the classification CNN prediction.

TABLE V

PERFORMANCE PER FOLD OF THE PROPOSED CDAs. FOLD 4, HIGHLIGHTED IN GRAY, WAS NOT USED TO DEVELOP THE PROPOSED ALGORITHMS AND HAS BEEN RESERVED FOR THE FINAL TEST

Fold	CNN-MDI $(\omega_1 = 0.575, \omega_2 = 0.425)$		CNN-GSP $(\omega_1 = 0.500, \omega_2 = 0.775)$	
	P_d	FAR	P_d	FAR
1	1.0000	0.0833	1.0000	0.0833
2	0.9800	0.1389	0.9400	0.0556
3	1.0000	0.0556	1.0000	0.0833
4	1.0000	0.0556	1.0000	0.0000
Total	0.995	0.0833	0.985	0.0556

suggests that the neural networks did not overfit to Folds 1–3, used to develop the algorithms and search for hyperparameters.

Fig. 6 shows heat maps over nine patches centralized at different targets contained in CARABAS-II DIs. These maps were generated using the Grad-CAM algorithm [50], which quantifies how much each region of an image influences the prediction of a neural network-based classifier. This is achieved by measuring the gradients of a particular convolutional layer—usually the last one before the output layer—with respect to the loss function when a particular input is given to the network. Then, the gradients are processed and resized to the dimensions of the input, to generate a heat map. The
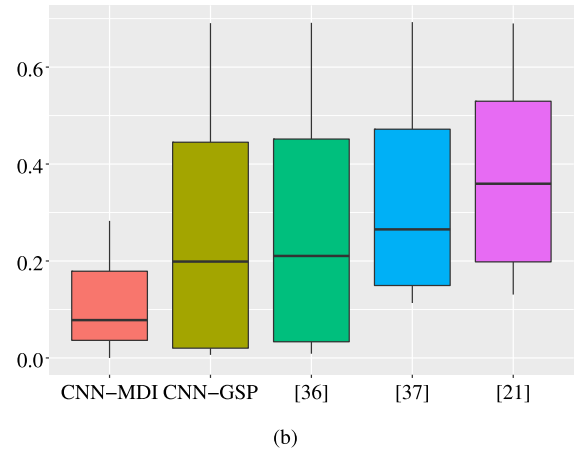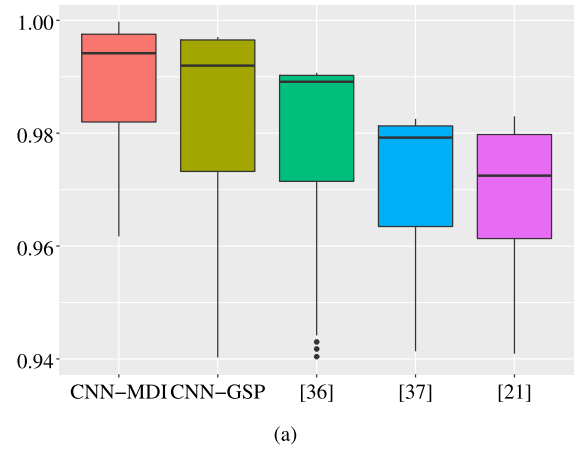

(a)

(b)

Fig. 7. Box-plot charts of the probability of detection values and FARs of the data extracted from the ROC curves presented in Fig. 5. (a) Probability of Detection. (b) FAR [1/km^2].

samples displayed in Fig. 6 show that the classification CNN is usually more influenced by the regions that surround the targets but not so much by the targets themselves. This implies that this network is not heavily influenced by the peculiarities and forms of each target but instead by their surroundings in the image.

D. Statistical Evaluation

To further evaluate the performance of the proposed methods, we considered the Kruskal–Wallis test and the posthoc Dunn's test to identify if the derived methods present significantly different mean behavior in terms of the probability of detection and FAR in comparison with the approaches described in [21], [36], and [37]. Both tests are widely employed for comparisons purposes of machine learning schemes in different signal processing applications (see [51], [52], [53], [54], [55]). We also considered a visual inspection of the data through box-plots.

To perform the tests and create the charts, we extracted the data from the ROC curves presented in Fig. 5 for FAR $\in$ [0/km^2, 0.285/km^2], observing that the CNN-MDI algorithm reaches $P_d = 1$ at FAR $= 0.285$/km^2. We performed a linear interpolation with the ROC curves to increase the available point numbers of operation used in the calculations. The significance level was set to 0.05, a convenient cutoff level to reject the null hypothesis [56].

TABLE VI

KRUSKAL–WALLIS AND DUNN'S TEST RESULTS, WHERE
p-VALUES $< \alpha = 0.05$, BEING α THE SIGNIFICANCE
LEVEL, ARE HIGHLIGHTED IN GRAY

Probability of Detection

	Kruskal-Wallis p-value < 0.0001			
	CNN-MDI	CNN-GSP	Method in [36]	Method in [37]
CNN-GSP	0.0782			
Method in [36]	0.0002	0.0098		
Method in [37]	< 0.0001	< 0.0001	0.0002	
Method in [21]	< 0.0001	< 0.0001	< 0.0001	0.5109

False Alarm Rate [$1/\mathrm{km}^2$]

	Kruskal-Wallis p-value < 0.0001			
	CNN-MDI	CNN-GSP	Method in [36]	Method in [37]
CNN-GSP	0.0018			
Method in [36]	0.0003	0.5198		
Method in [37]	< 0.0001	0.0061	0.0284	
Method in [21]	< 0.0001	< 0.0001	0.0003	0.1893

Fig. 7 displays the box-plot charts of the probability of detection values and FAR of the data extracted from the ROC curves presented in Fig. 5. Visually, we can verify that [21] and [37] show a similar mean behavior for both probability of detection and FAR. Additionally, we can visually highlight that: 1) CNN-MDI and CNN-GSP and 2) CNN-GSP and [36] have a similar mean behavior.

Table VI displays the p-values of the Kruskal–Wallis and Dunn's tests—p-values $< \alpha = 0.05$, α being the significance level, are highlighted in gray. The results presented in Table VI corroborate with the visual inspection insights. For example, in terms of probability of detection, CNN-MDI and CNN-GSP presented similar mean behavior. On the other hand, both proposed methods significantly excelled all the other tested tools. The CNN-MDI performed better in terms of FAR compared to all the tested methods, and the CNN-GSP improved the performance compared to [21] and [37]. The statistical tests highlight that the derived schemes presented competitive performance when compared with recently published results.

E. Time Evaluation

To better visualize the differences in numerical cost between the proposed algorithms, Table VII shows the average running time for both methods. The standard deviation of the running time is also presented. The results were obtained by measuring the time it took for each of the 24 possible stack combinations, described in Table V, to be analyzed by both algorithms. The longer average running time of the CNN-MDI can be linked to the fact that it takes as inputs $N-1$ DIs, N being the size of the stack, whereas CNN-GSP only takes one, meaning that the Segmentation CNN has to analyze $N-1$ times more images.

TABLE VII

TIME EVALUATION RESULTS, CALCULATED WITH MEASUREMENTS
OF THE RUNNING TIME FOR THE CNN-GSP AND
CNN-MDI ALGORITHMS OVER THE 24 STACKS
PRESENTED IN TABLE V

Method	Total Time Mean $\pm$ Standard Deviation
CNN-GSP	17.428 ± 1.173 seconds
CNN-MDI	4.476 ± 1.532 seconds

Also, the Classification CNN has to analyze $N-1$ patches for each possible change/target location. On the other hand, this added numerical complexity resulted in higher performance indicators in the executed tests.

VI. CONCLUSION

In this article, two CNN-based CDAs using stacks of WR SAR images were proposed. Their purpose is to look for new elements in a particular scene by analyzing DIs. They can be divided into four main steps: 1) DI formation; 2) semantic segmentation; 3) clustering; and 4) classification of changes. The CNN-GSP algorithm forms a DI using a GSP image sent to a CNN-based CDA to locate relevant changes. On the other hand, the CNN-MDI algorithm employs MDIs of the same scene in a CNN-based CDA. Both methods were tested in the CARABAS-II data yielding higher detection probabilities and lower FARs compared to other state-of-the-art algorithms. By applying the Kruskal–Wallis and Dunn's tests, it is shown that the performance improvements are statistically significant. Moreover, when the two algorithms were compared, CNN-MDI showed better performance indicators, at the cost of a higher computational complexity.

REFERENCES

[1] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*. Rijeka, Croatia: SciTech, 2004.

[2] G. Melamed, S. R. Rotman, D. G. Blumberg, and A. J. Weiss, "Anomaly detection in polarimetric radar images," *Int. J. Remote Sens.*, vol. 33, no. 4, pp. 1164–1189, Feb. 2012.

[3] B. G. Palm, F. M. Bayer, and R. J. Cintra, "Improved point estimation for the Rayleigh regression model," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–4, 2022.

[4] H. Sportouche, J.-M. Nicolas, and F. Tupin, "Mimic capacity of Fisher and generalized gamma distributions for high-resolution SAR image statistical modeling," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 12, pp. 5695–5711, Dec. 2017.

[5] C. Tison, J. M. Nicolas, F. Tupin, and H. Maitre, "A new statistical model for Markovian classification of urban areas in high-resolution SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 42, no. 10, pp. 2046–2057, Oct. 2004.

[6] J. Inglada and G. Mercier, "A new statistical similarity measure for change detection in multitemporal SAR images and its extension to multiscale change analysis," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 5, pp. 1432–1445, May 2007.

[7] G. Mercier, G. Moser, and S. B. Serpico, "Conditional copulas for change detection in heterogeneous remote sensing images," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 5, pp. 1428–1441, May 2008.

[8] Y. Zhao, B. Huang, and H. Song, "A robust adaptive spatial and temporal image fusion model for complex land surface changes," *Remote Sens. Environ.*, vol. 208, pp. 42–62, Apr. 2018.

[9] J. Naranjo-Torres, J. Gambini, and A. C. Frery, "The geodesic distance between G_I^0 models and its application to region discrimination," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 10, no. 3, pp. 987–997, Mar. 2017.

[10] H. Wang and K. Ouchi, "Accuracy of the K-distribution regression model for forest biomass estimation by high-resolution polarimetric SAR: Comparison of model estimation and field data," *IEEE Trans. Geosci. Remote Sens.*, vol. 46, no. 4, pp. 1058–1064, Apr. 2008.

[11] L. E. Falqueto, J. A. S. Sa, R. L. Paes, and A. Passaro, "Oil rig recognition using convolutional neural network on Sentinel-1 SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1329–1333, Aug. 2019.

[12] L. Gomez, L. Alvarez, L. Mazorra, and A. C. Frery, "Fully PolSAR image classification using machine learning techniques and reaction-diffusion systems," *Neurocomputing*, vol. 255, pp. 52–60, Sep. 2017.

[13] H. Mahmoudzadeh, "Digital change detection using remotely sensed data for monitoring green space destruction in Tabriz," *Int. J. Environ. Res.*, vol. 1, pp. 35–41, Jan. 2007.

[14] S. Martinis, A. Twele, and S. Voigt, "Unsupervised extraction of flood-induced backscatter changes in SAR data using Markov image modeling on irregular graphs," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 1, pp. 251–263, Jan. 2011.

[15] C. Pratola, F. D. Frate, G. Schiavon, and D. Solimini, "Toward fully automatic detection of changes in suburban areas from VHR SAR images by combining multiple neural-network models," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2055–2066, Apr. 2013.

[16] A. Vidal and M. R. Moreno, "Change detection of isolated housing using a new hybrid approach based on object classification with optical and TerraSAR-X data," *Int. J. Remote Sens.*, vol. 32, no. 24, pp. 9621–9635, Dec. 2011.

[17] L. Jia, M. Li, P. Zhang, and Y. Wu, "SAR image change detection based on correlation kernel and multistage extreme learning machine," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 10, pp. 5993–6006, Oct. 2016.

[18] L. M. Ulander, W. E. Pierson, M. Lundberg, P. Follo, P.-O. Frolind, and A. Gustavsson, "Performance of VHF-band SAR change detection for wide-area surveillance of concealed ground targets," *Proc. SPIE*, vol. 5427, pp. 259–271, Sep. 2004.

[19] V. T. Vu, R. Machado, M. I. Pettersson, P. Dammert, and H. Hellsten, "False alarm reduction in wavelength-resolution SAR change detection using adaptive noise canceler," *IEEE Trans. Geosci. Remote Sens.*, vol. 55, no. 1, pp. 591–599, Jan. 2017.

[20] D. I. Alves et al., "A statistical analysis for wavelength-resolution SAR image stacks," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 2, pp. 227–231, Feb. 2020.

[21] V. T. Vu, "Wavelength-resolution SAR incoherent change detection based on image stack," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 7, pp. 1012–1016, Jul. 2017.

[22] V. T. Vu, N. R. Gomes, M. I. Pettersson, P. Dammert, and H. Hellsten, "Bivariate gamma distribution for wavelength-resolution SAR change detection," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 1, pp. 473–481, Jan. 2019.

[23] A. B. Campos, M. I. Pettersson, V. T. Vu, and R. Machado, "False alarm reduction in wavelength–resolution SAR change detection schemes by using a convolutional neural network," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, 2020, Art. no. 4004805.

[24] L. M. Ulander, P.-O. Froelind, A. Gustavsson, H. Hellsten, and B. Larsson, "Detection of concealed ground targets in CARABAS SAR images using change detection," in *Proc. SPIE, Algorithms Synth. Aperture Radar Imag. VI*, vol. 3721, E. G. Zelnio, Ed., Orlando, FL, USA, Aug. 1999, pp. 243–252. [Online]. Available: http://proceedings.spiedigitallibrary.org/proceeding.aspx?articleid=989291, doi: 10.1117/12.357641.

[25] Z. Xu, "Wavelength-resolution SAR speckle model," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[26] R. Machado, M. I. Pettersson, V. T. Vu, P. Dammert, and H. Hellsten, "The stability of UWB low-frequency SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 8, pp. 1114–1118, Aug. 2016.

[27] F. Baselice, G. Ferraioli, and V. Pascazio, "Markovian change detection of urban areas using very high resolution complex SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 5, pp. 995–999, May 2014.

[28] S. Montazeri, X. X. Zhu, M. Eineder, and R. Bamler, "Three-dimensional deformation monitoring of urban infrastructure by tomographic SAR using multitrack TerraSAR-X data stacks," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 12, pp. 6868–6878, Dec. 2016.

[29] Y. Wang, X. X. Zhu, and R. Bamler, "An efficient tomographic inversion approach for urban mapping using meter resolution SAR image stacks," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 7, pp. 1250–1254, Jul. 2014.

[30] L. M. H. Ulander, M. Lundberg, W. Pierson, and A. Gustavsson, "Change detection for low-frequency SAR ground surveillance," *IEE Proc.-Radar Sonar Navigat.*, vol. 152, no. 6, pp. 413–420, Dec. 2005.

[31] M. Lundberg, L. M. Ulander, W. E. Pierson, and A. Gustavsson, "A challenge problem for detection of targets in foliage," *Proc. SPIE*, vol. 6237, May 2006, Art. no. 62370K.

[32] SDMS. (2020). *Sensor Data Management System Public Web Site*. [Online]. Available: https://www.sdms.afrl.af.mil/index.php

[33] N. R. Gomes, P. Dammert, M. I. Pettersson, V. T. Vu, and H. Hellsten, "Comparison of the Rayleigh and K-distributions for application in incoherent change detection," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 5, pp. 756–760, May 2019.

[34] R. D. Molin, R. A. S. Rosa, F. M. Bayer, M. I. Pettersson, and R. Machado, "A change detection algorithm for SAR images based on logistic regression," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 1514–1517.

[35] D. I. Alves et al., "Wavelength-resolution SAR change detection using Bayes' theorem," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 13, pp. 5560–5568, 2020.

[36] J. G. Vinholi, D. Silva, R. Machado, and M. I. Pettersson, "CNN-based change detection algorithm for wavelength-resolution SAR images," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[37] B. G. Palm et al., "Wavelength-resolution SAR ground scene prediction based on image stack," *Sensors*, vol. 20, no. 7, pp. 1–16, 2020.

[38] L. P. Ramos et al., "A wavelength-resolution SAR change detection method based on image stack through robust principal component analysis," *Remote Sens.*, vol. 13, no. 5, p. 833, Feb. 2021.

[39] L. M. Ulander and M. Lundberg, "Modeling of change detection in VHF- and UHF-band SAR," in *Proc. 7th Eur. Conf. Synth. Aperture Radar*. Friedrichshafen, Germany : VDE, Jun. 2008, pp. 1–4.

[40] F. Gao, X. Wang, Y. Gao, J. Dong, and S. Wang, "Sea ice change detection in SAR images based on convolutional-wavelet neural networks," *IEEE Geosci. Remote Sens. Lett.*, vol. 16, no. 8, pp. 1240–1244, Aug. 2019.

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

[42] D. Vint, M. Anderson, Y. Yang, C. Ilioudis, G. Di Caterina, and C. Clemente, "Automatic target recognition for low resolution foliage penetrating SAR images using CNNs and GANs," *Remote Sens.*, vol. 13, no. 4, p. 596, Feb. 2021. [Online]. Available: https://www.mdpi.com/2072-4292/13/4/596

[43] W. Ye, C. Paulson, and D. Wu, "Target detection for very high-frequency synthetic aperture radar ground surveillance," *IET Comput. Vis.*, vol. 6, no. 2, pp. 101–110, Mar. 2012.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.

[45] M. Ester, H. Kiregel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining (KDD)*. Amsterdam, The Netherlands: Elsevier, 1996, pp. 1–6.

[46] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.

[47] D. P. Kingma and J. L. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–15.

[48] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," *J. Mach. Learn. Res.*, vol. 9, pp. 249–256, May 2010.

[49] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2017.

[50] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-CAM: Visual explanations from deep networks via gradient-based localization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 618–626.

[51] T. Murata et al., "Salivary metabolomics with alternative decision tree-based machine learning methods for breast cancer discrimination," *Breast Cancer Res. Treatment*, vol. 177, no. 3, pp. 591–601, Oct. 2019.

[52] A. Emami, N. Kunii, T. Matsuo, T. Shinozaki, K. Kawai, and H. Takahashi, "Seizure detection by convolutional neural network-based analysis of scalp electroencephalography plot images," *NeuroImage: Clin.*, vol. 22, Jan. 2019, Art. no. 101684.

[53] E. J. Sutton et al., "Breast cancer molecular subtype classifier that incorporates MRI features," *J. Magn. Reson. Imag.*, vol. 44, no. 1, pp. 122–129, Jul. 2016.

[54] F. D. Paiva, R. T. N. Cardoso, G. P. Hanaoka, and W. M. Duarte, "Decision-making for financial trading: A fusion approach of machine learning and portfolio selection," *Expert Syst. Appl.*, vol. 115, pp. 635–655, Jan. 2018.

[55] M. Taghavi et al., "Machine learning-based analysis of CT radiomics model for prediction of colorectal metachronous liver metastases," *Abdominal Radiol.*, vol. 46, no. 1, pp. 249–256, Jan. 2021.

[56] R. A. Fisher, "Statistical methods for research workers," in *Breakthroughs in Statistics*. Edinburgh, U.K.: Springer, 1992, pp. 66–70.

João G. Vinholi received the B.Eng. degree in electrical engineering from the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, in 2020. He is currently pursuing the Ph.D. degree in electronics and computer engineering with the Aeronautics Institute of Technology (ITA), São José dos Campos, Brazil.

He is also with ICEYE Luxembourg, Luxembourg City, Luxembourg, as a Machine Learning Engineer, where he conducts the research and development of machine learning-based tools for remote sensing applications. His research interests include machine learning, deep learning, computer vision, digital signal processing, and their applications for remote sensing.

Bruna G. Palm (Member, IEEE) received the B.Sc. degree in statistics from the Federal University of Santa Maria (UFSM), Santa Maria, Brazil, in 2014, and the D.Sc. degree in statistics from the Federal University of Pernambuco (UFPE), Recife, Brazil, in 2020.

From February 2018 to January 2019, she was a Guest Ph.D. Researcher with the Blekinge Institute of Technology (BTH), Karlskrona, Sweden. From April 2020 to March 2021, she was a Research Fellow with the Department of Telecommunications, Aeronautics Institute of Technology (ITA), São José dos Campos, Brazil. From August 2021 to November 2021, she was a Visiting Research Fellow at BTH in partnership with Saab AB, Stockholm, Sweden. She is currently an Associate Senior Lecturer at BTH. Her main research interests include data science, regression/dynamic models, statistical computing, parametric inference, SAR data, and statistical signal/image processing.

Dr. Palm's Ph.D. dissertation received the Commission Outstanding Award from the Brazilian Society of Applied and Computational Mathematics, Brazil, in 2021.

Danilo Silva (Member, IEEE) received the B.Sc. degree from the Federal University of Pernambuco (UFPE), Recife, Brazil, in 2002, the M.Sc. degree from the Pontifical Catholic University of Rio de Janeiro (PUC-Rio), Rio de Janeiro, Brazil, in 2005, and the Ph.D. degree from the University of Toronto, Toronto, ON, Canada, in 2009, all in electrical engineering.

From 2009 to 2010, he was a Post-Doctoral Fellow with the University of Toronto; the École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland; and the State University of Campinas (UNICAMP), Campinas, Brazil. In 2010, he joined the Department of Electrical and Electronic Engineering, Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, where he is currently an Associate Professor. His research interests include machine learning, computer vision, signal processing, and their applications in medicine and healthcare. His previous research interests include wireless communications, channel coding, information theory, and network coding.

Dr. Silva is a member of the Brazilian Telecommunications Society (SBrT). He was a recipient of a CAPES Ph.D. Scholarship in 2005, the Shahid U.H. Qureshi Memorial Scholarship in 2009, and an FAPESP Post-doctoral Scholarship in 2010. He also served as the Technical Program Co-Chair for the XXXVIII Brazilian Symposium on Telecommunications and Signal Processing (SBrT 2020).

Renato Machado (Senior Member, IEEE) received the B.S.E.E. degree from São Paulo State University (UNESP), Ilha Solteira, Brazil, in 2001, and the M.Sc. and Ph.D. degrees in electrical engineering from the Federal University of Santa Catarina (UFSC), Florianópolis, Brazil, in 2004 and 2008, respectively.

From November 2013 to February 2015, he was a Visiting Research Fellow with the Blekinge Institute of Technology (BTH), Karlskrona, Sweden, in partnership with Saab AB, Stockholm, Sweden. From August 2009 to December 2017, he was an Assistant (2008–2016) and Associate (2017) Professor at the Federal University of Santa Maria, Santa Maria, Brazil, where he lectured many courses in bachelor and graduate programs and assumed different positions in the institution, namely, a Researcher Leader of the Communications and Signal Processing Research Group, a Coordinator of the Telecommunications Engineering Program, and the Director of the Aerospace Science Laboratory. Since December 2017, he has been an Associate Professor with the Aeronautics Institute of Technology (ITA), São José dos Campos, Brazil. He is the Research leader of the Digital and Signal Processing Laboratory, ITA, and the SAR and Radar Signal Processing Laboratory, ITA. He is a Principal Investigator of the Brazilian Institute of Data Science (BI0S), Campinas, Brazil, a Research Center for Applied Science and Technology in Artificial Intelligence, FAPESP/MCTI/CGI. From August 2021 to July 2022, he was the Head of the Graduate Program in Electronics and Computer Engineering, ITA, where he has been the Head of the Electronic Engineering Division since June 2022. His research interests include SAR processing, SAR image processing, change detection, radar signal processing, and digital signal processing.

Mats I. Pettersson (Senior Member, IEEE) received the M.Sc. degree in engineering physics, the Licentiate degree in radio and space science, and the Ph.D. degree in signal processing from the Chalmers University of Technology, Gothenburg, Sweden, in 1993, 1995, and 2000, respectively.

He was with Mobile Communication Research at Ericsson, Lund, Sweden. He was with the Swedish Defense Research Agency (FOI) for ten years in Linköping, Sweden. At FOI, he focused on ultra-wideband low-frequency SAR systems. Since 2005, he has been with the Blekinge Institute of Technology (BTH), Karlskrona, Sweden, where he is currently a Full Professor. He has authored or coauthored over 200 scientific publications, of which more than 60 are in peer-reviewed scientific journals. His research is related to radar surveillance and remote sensing. His main research interests include SAR processing, space–time adaptive processing (STAP), high-resolution SAR change detection, automotive radar, radio occultation, THz SAR, and computer vision.