Multi-Objective CNN-Based Algorithm for SAR Despeckling

Sergio Vitale, *Graduate Student Member, IEEE*, Giampaolo Ferraioli, *Senior Member, IEEE*, and Vito Pascazio, *Senior Member, IEEE*

***Abstract*— Deep learning (DL) in remote sensing has nowadays become an effective operative tool: it is largely used in applications, such as change detection, image restoration, segmentation, detection, and classification. With reference to the synthetic aperture radar (SAR) domain, the application of DL techniques is not straightforward due to the nontrivial interpretation of SAR images, especially caused by the presence of speckle. Several DL solutions for SAR despeckling have been proposed in the last few years. Most of these solutions focus on the definition of different network architectures with similar cost functions, not involving SAR image properties. In this article, a convolutional neural network (CNN) with a multi-objective cost function taking care of spatial and statistical properties of the SAR image is proposed. This is achieved by the definition of a peculiar loss function obtained by the weighted combination of three different terms. Each of these terms is dedicated mainly to one of the following SAR image characteristics: spatial details, speckle statistical properties, and strong scatterers identification. Their combination allows balancing these effects. Moreover, a specifically designed architecture is proposed to effectively extract distinctive features within the considered framework. Experiments on simulated and real SAR images show the accuracy of the proposed method compared with the state-of-art despeckling algorithms, both from a quantitative and qualitative point of view. The importance of considering such SAR properties in the cost function is crucial for correct noise rejection and details preservation in different underlined scenarios, such as homogeneous, heterogeneous, and extremely heterogeneous.**

***Index Terms*— Convolutional neural network (CNN), deep learning (DL), despeckling, image restoration, statistical distribution, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) imaging system produces images affected by a multiplicative noise, called speckle, creating a succession of strong and weak backscatterings. The presence of the speckle impairs the performance of several tasks, such as detection, segmentation, and classification; indeed, a despeckling operation is crucial for the interpretation of SAR images.

Manuscript received April 24, 2020; revised August 17, 2020; accepted October 20, 2020. Date of publication November 16, 2020; date of current version October 26, 2021. *(Corresponding author: Sergio Vitale.)*

Sergio Vitale and Vito Pascazio are with the Dipartimento di Ingegneria, University Parthenope, 80143 Naples, Italy (e-mail: sergio.vitale@uniparthenope.it; vito.pascazio@uniparthenope.it).

Giampaolo Ferraioli is with the Dipartimento di Scienze e Tecnologie, University Parthenope, 80143 Naples, Italy (e-mail: giampaolo.ferraioli@uniparthenope.it).

Digital Object Identifier 10.1109/TGRS.2020.3034852

The high number of studies and algorithms developed in the last forty years testifies to the importance of this topic. Despite the great understanding of the speckle and its characteristics, despeckling is still an open issue far from being solved.

The first solutions work in the spatial domain, such as [1]–[5] and [6], [7]: the first ones are based on a minimum-mean-square error (MMSE), while the second ones propose a maximum *a posteriori* (MAP) filter. These methods produce intense smoothing for reducing speckle in homogeneous areas that cannot be applied on the edges.

Since the early 1990s, despeckling techniques operating in a transformed domain have been proposed [8]–[11]. Filters based on such an approach often operate a homomorphic transformation in order to work with additive noise. These solutions embody strong spatial adaptability in order to better preserve edges, which is a crucial issue in SAR despeckling.

A new research line in the despeckling domain has been drawn by the nonlocal methods that have shown very effective performances in preserving details while removing noise [12]. Such methods look for similar patches in the image and merge them in order to produce target pixels. Usually, differently from the previous solutions, statistics of the speckle and of the SAR backscattering [13], [14] are taken into account for the definition of patch similarity. Several algorithms have been defined within the nonlocal paradigm, mainly by differentiating the choice of the similarity criterion or the merging function. For example, the nonlocal paradigm based on different SAR similarity distances is applied by the methods proposed in [15] and [16]. Whereas, a ratio-based metric is used in [17] and [18]. Hybrid approaches arose, such as [19]–[21], that join the nonlocal paradigm with the wavelet transform. A detailed review of the aforementioned despeckling filters can be found in [22] and [23].

In the last years, deep learning (DL) is showing great performance in many natural image processing tasks, such as classification, detection, segmentation, and not less denoising. Indeed, also remote sensing community is starting to exploit the potential of this approach, even if many difficulties arise due to the difference between natural and remote sensed images.

Recently, several DL solutions have been proposed for SAR despeckling. Such methods are data-driven: differently from the previous classical approaches, it is mandatory to have a data set composed of many couples of noisy inputs and noise-free images (references). Since for SAR despeckling,

a noise-free reference is not available, the first issue for such methods is the construction of simulated data set.

Mainly, DL despeckling algorithms rely on the simulation of fully developed speckle multiplied to the grayscale version of an optical image that, at the same time, serves as a clean reference for the network. For the sake of simplicity, this approach is referred to as *synthetic approach* in the following of this article. Among them, we recall [24]–[28]. In [24], a simple residual CNN composed of eight layers is proposed, while a CNN with dilated convolution in order to increase the receptive field and skip connections for avoiding vanishing gradient is presented in [25]. In [26], the use of U-Net has been proposed. In [27], the Mulog [29] framework combined with an additive white Gaussian noise (AWGN) denoising CNN is adapted for SAR. Later, in [28], the same method is proposed trying to combine DL and the NL paradigm through a postclassification of the filtered image.

Moreover, instead of using a synthetic approach, in other techniques, such as [30] and [31], the multitemporal average version of SAR acquisition serves as a reference. Always for the sake of simplicity, this approach is referred to as *multitemporal approach* in the following of this article. Real data have been also used for training a CNN as in [32] following the *Noise2Noise* scheme [33]. In such a scheme, the network learns to predict the clean image by using as input-reference data two noisy images with the same underlying clean data but different independent realizations of noise.

Most of these proposals focus only on the definition of the architecture and use very similar cost functions, not taking into account statistical properties of the SAR image and the presence of strong scatterers, demanding their knowledge to the features extraction from the training data. In [25] and [32], the MSE is used as cost function. In [24] and [26], the MSE is combined with a total variation regularization. A smoothed L_1 loss adapted to the speckle noise case has been considered in [30]. The first attempt to include the first-order statistics of the speckle was proposed in [34], whereas, in [31], a cost function is used based on statistic similarity.

In this article, a CNN for SAR despeckling that takes into account the statistical properties of the SAR image has been proposed. The network is a 17 layers CNN with a skip connection trained with the synthetic approach. Beyond the proposed architecture, the main contribution is in the definition of a multi-objective cost function given by a combination of three terms, each designed for a precise goal. Indeed, each of these terms takes care, respectively, of spatial details, statistical properties, and strong scatterers identification.

The rest of this article is organized as follows. The description of the method and related contribution is in Section II. Experimental results and discussion are presented in Section III. Conclusion is presented in Section IV. An ablation study of the cost function has been carried out in the Appendix.

II. Methodology

In this section, the proposed method is described: first, the acquisition model and the statistics of the acquired SAR image is presented; then, the definition of the data simulation process, of the proposed architecture, and of the multi-objective cost function are detailed. Finally, the contribution of this article is highlighted.

A. Signal Statistical Description

The interpretation of the SAR image is challenging due to the geometrical properties of the SAR imaging system and to the presence of speckle. Indeed, speckle is a multiplicative noise produced by interference among the backscatterings of the objects inside a resolution cell of the sensor [35]. The generic SAR image can be expressed as follows:

$$Y = X \cdot N \tag{1}$$

where Y is the SAR image, X is the noise-free image, and N is the speckle.

The statistical distribution of the speckle is well known under certain conditions. Three main cases can be considered: homogeneous, heterogeneous, and extremely heterogeneous areas. Homogeneous areas (such as fields, roads, and so on) are characterized by the lack of dominant scatterers, and the surface X can be considered stationary. This is the case of the fully developed hypothesis for the speckle N, whose amplitude follows the square root of Gamma distribution [14]:

$$p_N(n, L) = \frac{2L^L}{\Gamma(L)} n^{2L-1} e^{-Ln^2} \qquad n, L > 0 \tag{2}$$

where L is the number of looks of the SAR image and $\Gamma(\cdot)$ is the Gamma function. This probability density function (pdf) in case of single look becomes a Rayleigh distribution.

Heterogeneous (tree and forest) and extremely heterogeneous areas (urban) are characterized by objects with shape and dimension that produce geometrical distortions and strong backscattering (e.g., multiple bounces, layover, and shadowing). In heterogeneous areas, the speckle can still be considered Gamma distributed, but the surface is not stationary anymore. In an extremely heterogeneous area, the hypothesis of distributed scatterers is not valid anymore due to the presence of dominant ones. Indeed, the speckle does not follow anymore the fully developed hypothesis [36].

The statistical distribution of SAR backscattering Y in different scenarios is provided in [14], where the use of the square root of generalized inverse Gaussian distribution $\mathcal{G}_A(\alpha, \gamma, \lambda, L)$ as a general model for the amplitude return of SAR backscattering is proposed. Frery *et al.* [14] prove that the distribution of the SAR return of homogeneous (HO), heterogeneous (H), and extremely heterogeneous (EH) areas are a particular case of this distribution depending on the parameter subspace. An extension of this classification considering several possible scenarios has been recently proposed in [37].

According to [14], in the subspace ($\alpha > 0, \gamma = 0, \lambda > 0, L > 0$), when α and λ tends to infinite, the distribution tends to a square root gamma $\Gamma^{1/2}(L, L/\beta)$ as in (2) with β being the estimation $E[Y^2]$ of the second-order statistic. Such distribution describes the return from HO areas.

Moreover, Frery *et al.* [14] have proved that the SAR return Y in the H area follows the $\mathcal{K}_A(\alpha, \lambda, L)$ distribution. This is the distribution the $\mathcal{G}_A(\alpha, \gamma, \lambda, L)$ tends to, when the parameter subspace is always ($\alpha > 0, \gamma = 0, \lambda > 0, L > 0$).

Fig. 1. RGB samples of Merced Land Use data set.

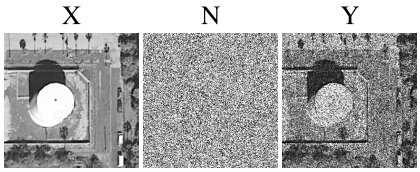

Fig. 2. Simulation process, from left to right: noise-free reference, simulated noise, and simulated SAR image.

For EH areas, the amplitude distribution of the SAR image can be described according to $G_A^0(\alpha, \gamma, L)$, that is the distribution $\mathcal{G}_A(\alpha, \gamma, \lambda, L)$ tends to, when the parameter subspace is $(\alpha < 0, \gamma > 0, \lambda = 0, L > 0)$.

B. Data Simulation

In this section, the data simulation process adopted for the training of the proposed CNN-based despeckling algorithm is illustrated.

Thousands of noise-free images from the optical UC Merced Land Use data set [38] have been considered. This data set is typically considered for classification purposes thanks to the presence of hundreds of images belonging to different classes. Samples of this data set are shown in Fig. 1.

The optical images have been converted from the RGB domain to the grayscale one, obtaining the noise-free references X. The speckle noise N has been generated under the fully developed hypothesis in case of a single-look image according to (2). The final noisy image Y has been obtained by simply multiplying the noise-free image by the speckle, as in Fig. 2.

Considering that the Merced Land Use data set is composed of several scenarios (such as agricultural field, baseball diamonds, forest, residential areas, and so on), the simulation process transforms all these data into noisy images whose distribution belongs to the $\Gamma^{1/2}$ distribution (HO areas) or to $\mathcal{K}_A$ distribution (H areas) that both are a particular case of the generalized inverse Gaussian distribution. In Fig. 3, the distributions of two samples of the data set are shown. The magenta solid curve represents the distribution of a simulated image taken from the "agricultural" class of the data set. In this case, the surface X is almost homogeneous, and the distribution of the resulting simulated Y fits the $\Gamma^{1/2}$ (magenta dashed). At the same time, the black solid curve represents the distribution of a simulated image taken from the "forest" class. In this case, the texture X cannot be considered homogeneous, but some fluctuation had to be taken into account. Indeed, the distribution fits well the $\mathcal{K}_A$ curve. This process does not allow to simulate the EH case, where the speckle is not fully developed mainly due to the presence of dominant scatterers and geometrical distortions.

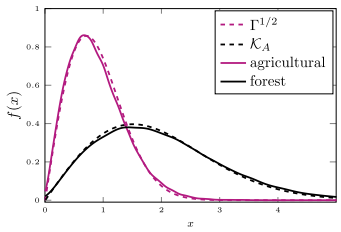

Fig. 3. Comparison between two different samples of training data set: agricultural sample (magenta solid) and forest (black solid). In dashed, the theoretical $\Gamma^{1/2}$ and $\mathcal{K}_A$ distributions.

From the whole data set, $57\,526 \times 64 \times 64$ amplitude patches for the training and $14\,336 \times 64 \times 64$ for the validation have been extracted.

C. Network Architecture

The design of the proposed network architecture comes from the results achieved in our previous works [34], [39], where ten layers of CNNs with different cost functions were proposed.

Starting from the result of [39], the proposed neural network is composed of 17 convolutional layers. For each layer, we consider rectified linear unit (ReLU) as activation function [40], but for the last. In all the layers, batch normalization [41] is performed except for the first and the last ones. In addition, skip connections, which have shown great utility in training deep networks [42], are introduced in the inner layers.

Given the previous remarks, the output of layer k can be expressed as

$$z_k = f_k(\Phi_k, z_{k-1}) = \begin{cases} \sigma(w_k * Y + b_k), & k = 1 \\ BN[\sigma(w_k * z_{k-1} + b_k)] + \alpha f_{k-3}(\Phi_{k-3}, z_{k-4}) & 1 < k < D \\ (w_k * z_{k-1} + b_k), & k = D \end{cases} \quad (3)$$

where

$$\alpha = \begin{cases} 1, & \langle k-1 \rangle_3 = 0 \\ 0, & \text{otherwise} \end{cases}$$

with number of layers $D = 17$, and (w_k, b_k, Φ_k, z_k) the weights, the bias, the set of parameters and the output of layer k, respectively. BN stays for batch normalization and $\sigma(\cdot) = \max(0, \cdot)$ is the ReLU activation function. The operation $\langle k-1 \rangle_3$ is the reminder of the division $(k-1)/3$.

Based on this network architecture, given a couple of samples (Y, X), where Y is the noisy image and X acts as a reference, the final estimated clean image is $\widehat{X} = z_D$.

For each layer, 64 features maps are extracted except for the last one that has to fit the single-channel output. All the convolutional kernel have dimension 3×3. In Fig. 4, a scheme of the network is shown. The scheme of the residual block is defined in Fig. 5.

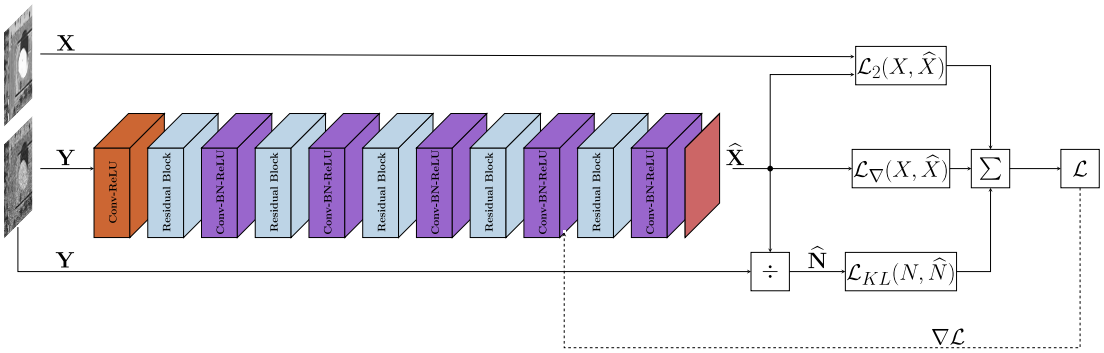

Fig. 4. Network architecture: all the layers have 64 features maps with a 3×3 convolutional kernel. The first layer (in orange) is followed by the ReLU activation function. After, there is an alternation of the residual block (in light blue) and inner layers with ReLU and batch normalization (in purple), while the last layer (in red) has neither activation function nor normalization. The cost function is a linear combination of three terms.

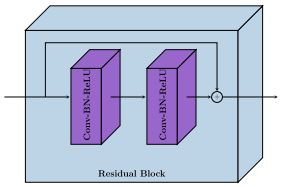

Fig. 5. Residual Block is composed of two Conv-BN-ReLU layers and a skip connection that sums the input to the output of second layer.

D. Cost Function

In the proposed algorithm, the aim is to propose a cost function that takes care of both spatial and statistical properties of the SAR images. The defined multi-objective cost function $\mathcal{L}$ is a linear combination of three terms, each of them specifically dedicated to catch and to preserve information from the SAR image. Specifically

$$\mathcal{L} = \mathcal{L}_2 + \lambda_{\mathrm{KL}}\mathcal{L}_{\mathrm{KL}} + \lambda_\nabla \mathcal{L}_\nabla \tag{4}$$

$$\mathcal{L}_2 = \mathcal{L}_{\mathrm{MSE}} = \|\widehat{X} - X\|^2 \tag{5}$$

$$\mathcal{L}_{\mathrm{KL}} = D_{\mathrm{KL}}(\widehat{N}, N_{\mathrm{teo}}) \tag{6}$$

$$\mathcal{L}_\nabla = \|\nabla X - \nabla \widehat{X}\|^2. \tag{7}$$

$\mathcal{L}_2$ is the MSE between the reference X and filtered image $\widehat{X}$; $\mathcal{L}_{\mathrm{KL}}$ is the Kullback–Leibler divergence (D_{KL}) between the distribution of estimated noise $\widehat{N} = Y/\widehat{X}$ and the theoretical one N_{teo}, whose definition is provided in Section III-A; $\mathcal{L}_\nabla$ is the MSE between the gradient of the reference X and gradient of the filtered image $\widehat{X}$.

Let us consider each of the three terms separately.

Naturally, the goal is to train the network to generate an output as similar as possible to the reference. To this aim, the $\mathcal{L}_2$ term directly compares the output $\widehat{X}$ with the reference X, and it is responsible for spatial reconstruction.

Despite the importance of reducing spatial distortion, taking into account the properties of the noise within the despeckling operation is crucial, as shown by different methods, such as [17] and [18]. For this reason, the $\mathcal{L}_{\mathrm{KL}}$ term that takes into account the statistical properties of the noise has been introduced.

The $\mathcal{L}_{\mathrm{KL}}$ is the Kullback–Leibler divergence computed between the pdf of the estimated ratio image (the ratio between the SAR image and the estimated noise-free one) and the theoretical fully developed speckle (in our case, a Rayleigh distribution with parameter $\sigma = 1/\sqrt{2}$). The goal is to train the network to produce an output whose ratio image follows the statistical properties of the speckle.

The introduction of the $\mathcal{L}_\nabla$ term is twofold: improving the edge preservation [43] and dealing with dominant scatterers in real images. $\mathcal{L}_\nabla$ compares the gradients of $\widehat{X}$ with a gradient of X. The gradient gives information on the edges but, obviously, is not exactly an edge detector. It highlights transitions in images and so tends to identify the presence of structures. Therefore, if from one side, it trains the network in preserving edges, on the other, it helps the network in identifying and isolating strong scatterers.

In the Appendix, an ablation study on the effects of these three terms has been proposed.

E. Identification of Not Fully Developed Areas

The presence of strong scatterers is challenging for all the filters, and their filtering policy is still an open issue within the despeckling community [35].

As reported in the literature, these points should be left unfiltered or at least processed in a different way [44]. Some methods, such as SAR-BM3D, NOLAND, and fast adaptive nonlocal SAR despeckling (FANS), filter them by aggregation of similar patches selected through a statistical approach; other methods, such as [45], do not filter them at all.

These points are related to EH areas (usually urban), where, as pointed out by Frery *et al.* [14] and Tison *et al.* [36], the speckle is not fully developed anymore.

As reported in Section II-B, the only distribution not included in the training is the one for the EH areas. However, the defined cost function allows easy detection of such areas. Indeed, if from one side the presence of $\mathcal{L}_{\mathrm{KL}}$ encourages the filtering under the fully developed hypothesis, from the other side, $\mathcal{L}_\nabla$ tends to preserve edges and to identify structures. Their combination (together with $\mathcal{L}_2$) highlights the presence of such points producing strong values on the ratio image. As a matter of fact, on these points, the ratio image of an ideal filter should not show a Rayleigh distribution. Thus, the appearance of such points on the ratio image can be considered as a positive issue. Actually, it allows identifying such points, having a different statistical distribution from the trained one

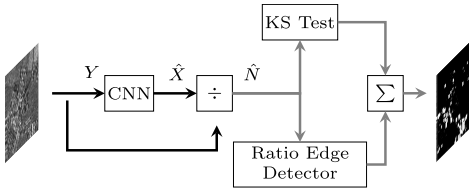

Fig. 6. Flowchart for the identification of not fully developed areas from the ratio image.

(i.e., it allows to automatically identify points belonging to EH areas).

The identification of such EH points is performed directly from the ratio image produced by the proposed CNN. For this goal, a combination of the ratio edge detector proposed in [4] and a Kolmogorov–Smirnov test on the ratio image produced by our algorithm is applied.

The former aims to highlight the edges and remaining structures in the ratio image, the latter detects the area where the predicted speckle is not fully developed by comparison through a threshold at patch level of ratio image distribution and the fully developed one. In Fig. 6, a flowchart of this detection process is shown. The knowledge of those pixels allows the final user to decide which filtering policy to be adopted (left unfiltered, define a specific statistical-based filter, using a multitemporal approach, and so on) [11].

F. Contribution

In this section, the contribution of the proposed method, called multi-objective network (MONet) for SAR despeckling, is described, and its innovative issues are highlighted. The proposed MONet shares some points with image despeckling CNN (IDCNN) [24], SAR-CNN [30], and SAR-dilated residual network (DRN) [25]. Indeed, the proposed CNN has 17 layers, such as SAR-CNN, and also skip connections are added in the inner layers, like in SAR-DRN. Differently from SAR-DRN, a deeper network has been preferred to dilate convolutions. Deeper networks allow us to extract more features and to add more abstractions, facilitating the exploitation of the data and the network generalization. The depth has been set experimentally: in [46], it has been proved that a deeper network gives better results.

The main innovation consists in the definition of the cost function: a combination of the $\mathcal{L}_2$ norm with other terms is used for the reconstruction. While IDCNN combines the $\mathcal{L}_2$ with the total variation in order to provide smooth results, in the proposal, the term $\mathcal{L}_\nabla$ for edges preservation and dominant scatterers identification has been considered. Moreover, a statistical term $\mathcal{L}_{\mathrm{KL}}$ for speckle properties preservation is added in the combination that leads to the whole cost function. In Table I, the differences among the aforementioned methods are summarized.

III. Experimental Results

In order to validate the method, experiments have been carried out on both simulated and real data. Both quantitative analysis, based on performance indexes, and qualitatively analysis, based on visual inspection, have been conducted.

TABLE I
Main Differences Among Compared DL Methods

Method	Depth	Skip Connection	Spatial Loss	Statistical Loss
ID-CNN	10	×	$\mathcal{L}_2 + TV$	×
SAR-CNN	17	×	smoothed $\mathcal{L}_1$	×
SAR-DRN	7	✓	$\mathcal{L}_2$	×
MONet	17	✓	$\mathcal{L}_2 + \mathcal{L}_\nabla$	$\mathcal{L}_{KL}$

For comparison, two different families of despeckling algorithms have been considered: nonlocal and DL-based ones. In particular, NL algorithms have been addressed since they are often considered in the literature as a benchmark for evaluating achievable performances. Between the available NL algorithms, we considered FANS [21], SAR-BM3D [19], and NOLAND [18], while the DL-based algorithms have been considered in order to compare the performances of the proposed algorithm with methods sharing the same philosophy. In particular, ID-CNN and SAR-DRN as DL methods have been used.

For each NL method, the parameters have been set accordingly to those suggested in the relative articles. While given that the DL solutions are data-driven, in order to have a fair comparison, the CNN-based solutions have been retrained on our same data set following the description in the relative articles.

For this reason, we did not compare with the solution based on the multitemporal approach [30]–[32], because a fair comparison is not possible using training on simulated data. Moreover, for the rest of DL articles, the code or traning data set are not publicly available.

The proposed network is trained with mini batch of 128 samples, using the Adam optimizer [47] with parameter $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The learning rate is set to $\eta = 0.0001$ for the first 87 epochs, and after the training is refined for the other 35 epochs with a learning rate scaled by 10. The lambdas parameter for the cost function has been empirically set for balancing their effects: $\lambda_{\mathrm{KL}} = 10^4$ and $\lambda_\nabla = 1$. The framework used for the implementation is Theano, running on Python. Both training and testing have been carried out on a GeForce GTX 1080Ti GPU with 11 GB of memory. The code of the proposed method is available in *https://github.com/sergiovitale/MONet-SAR-Despceckling-CNN-Theano-implementation*.

A. Metrics

For numerical evaluation both reference and no-reference metrics have been considered. As reference metrics, the *structural similarity* (SSIM) index, the MSE, and the *signal-to-noise ratio* (SNR) have been used for evaluating results on the simulated data set, where a reference is available.

1) **SSIM** measures the similarity between $\widehat{X}$ and X from a perceptual point of view. The ideal filter would produce SSIM $= 1$.
2) **MSE** measures the average similarity between $\widehat{X}$ and X. The ideal value is zero.
3) **SNR** measures the SNR and gives us information about the capability of noise suppression. The higher SNR, the better the filter.

Regarding no-reference metrics, the equivalent number of looks (ENL), the M-index, the Haralick homogeneity δh, the residual ENL $r_{\widehat{\text{ENL}}}$, the mean of the ratio μ_N, and the Kullback–Leibler divergence D_{KL} are considered.

1) **ENL** is an indicator of noise suppression in homogeneous areas. Once a homogeneous area has been selected, the ENL computes the ratio between the squared power of the mean and the variance of the filtered image (both in intensity format)

$$\text{ENL} = \frac{E[\widehat{X}^2]^2}{\text{Var}(\widehat{X}^2)}. \tag{8}$$

Higher is the ENL, the greater is the noise suppression.

2) **M-index** [48] is a combination of three factors δh, $r_{\widehat{\text{ENL}}}$, and r_μ:
   a) δh is based on the Haralick homogeneity texture [49] and it is the distance between the homogeneity h_0 of ratio image compared with the homogeneity h_g of the random permuted the ratio image itself. It is compute as $\delta h = ((|h_0 - h_g|)/h_0)$, with

$$h_z = \sum_i \sum_j \frac{1}{1+(i-j)^2} \cdot p_z(i,j) \tag{9}$$

   where $p_z(i, j)$ is the gray scale level co-occurrence matrix of the ratio image z at an arbitrary position. δh computes a sort of correlation of the ratio image and give us information of remaining structures that should not be present after an ideal filtering. The ideal filter will produce $\delta h = 0$.
   b) $r_{\widehat{\text{ENL}}}$ is the residual ENL and once n homogeneous patches are selected the ENL computed on ratio and SAR image are compared

$$r_{\widehat{\text{ENL}}} = \frac{1}{n} \sum_{i=1}^{n} \frac{|\widehat{\text{ENL}}_{\text{noisy}}(i) - \widehat{\text{ENL}}_{\text{ratio}}(i)|}{\widehat{\text{ENL}}_{\text{noisy}}(i)}. \tag{10}$$

   The ideal filter will produce $r_{\widehat{\text{ENL}}}$ equal to 0.
   c) r_μ is the function of the mean ratio μ_N computed on the same patches selected for the $r_{\widehat{\text{ENL}}}$

$$r_\mu = \frac{1}{n} \sum_{i=1}^{n} |1 - \mu_N(i)|. \tag{11}$$

   The ideal filter will produce r_μ equal to zero.

   The ideal filter will produce an M-index equal to zero.

3) The $\mathbf{D}_{\text{KL}}$ computes the distance between the statistical distribution of the ratio image with the theoretical Rayleigh distribution

$$D_{\text{KL}}(\hat{N}, N_{\text{teo}}) = \sum_i P_{\hat{N}}(i) \log_2 \left(\frac{P_{\hat{N}}(i)}{P_{N_{\text{teo}}}(i)} \right) \tag{12}$$

   where $P_{\hat{N}}$ is the pdf of the predicted speckle and $P_{N_{\text{teo}}}$ is the pdf of the theoretical noise. Under the fully developed hypothesis, an ideal filter will produce a $D_{\text{KL}} = 0$

Clearly, other indexes could have been adopted and considered. We focus on these ones since they are largely and commonly adopted by the community.

TABLE II
NUMERICAL ASSESSMENT ON SIMULATED DATA SET: THE VALUE ARE AVERAGED ON THE WHOLE SIMULATED TESTING DATA SET COMPOSED OF 100 IMAGES

	SSIM	**SNR**	**MSE**	ENL
FANS	.7049	8.0432	.00482	**822**
SAR-BM3D	.7379	8.4251	.00438	240
NOLAND	.6847	7.4712	.00544	84
ID-CNN	.7231	8.3644	.00437	144
SAR-DRN	<u>.7437</u>	<u>8.7240</u>	<u>.00406</u>	374
MONet	**.7510**	**8.8555**	**.00395**	<u>580</u>

B. Simulated Results

For the simulation, 100 single-look amplitude images of size 256×256 have been selected. These belong to five classes (20 for each class) of the Merced Land Use data set not used during the training phase. In Table II, the numerical evaluation for reference metrics and ENL, averaged on the whole data set, is shown. The best solution is expressed in bold, the second best is underlined.

Regarding the reference metrics SSIM, SNR, and MSE, it is evident that DL solutions outperform the other given that they are trained on a data set with the same properties as the testing one. The proposed MONet outperforms all the DL and NL methods. The best NL solution on simulated experiments is SAR-BM3D. Regarding the ENL, FANS performs largely better than the others, followed by the MONet.

Numerical assessment is not enough, and visual inspection is essential for understanding the performance of a filter. Four different images with different textures are shown in Fig. 7 for qualitative analysis. Together with the noisy images (first column), the noise-free reference images are reported. Columns from 3 to 8 show the filtering results of the different considered approaches.

Among the NL methods, FANS is over smoothed, losing many spatial details, but with good edge preservation. NOLAND and SAR-BM3D are very close to each other with good detail preservation, but both of them produce some artifacts on homogeneous areas that impair the edges preservation. Among the CNN methods, MONet shows the best performance on spatial details and edges preservation. IDCNN and SAR-DRN are very close to each other, with the former producing a filtered image still a bit noisy and the latter producing some distortions on the edges.

In general, the proposed solution seems to produce the most similar image to the reference, showing very good noise suppression without losing details and good edges preservation.

Regarding the computational efficiency, the processing time is clearly related to the number of parameters the network is composed of. Being our network deeper compared with the others, it allows extracting more representative features, resulting in a better generalization at the cost of higher computational time. Anyway, the algorithm guarantees a fast processing time: for example, a 3000×3000 is processed in approximately 4 s.

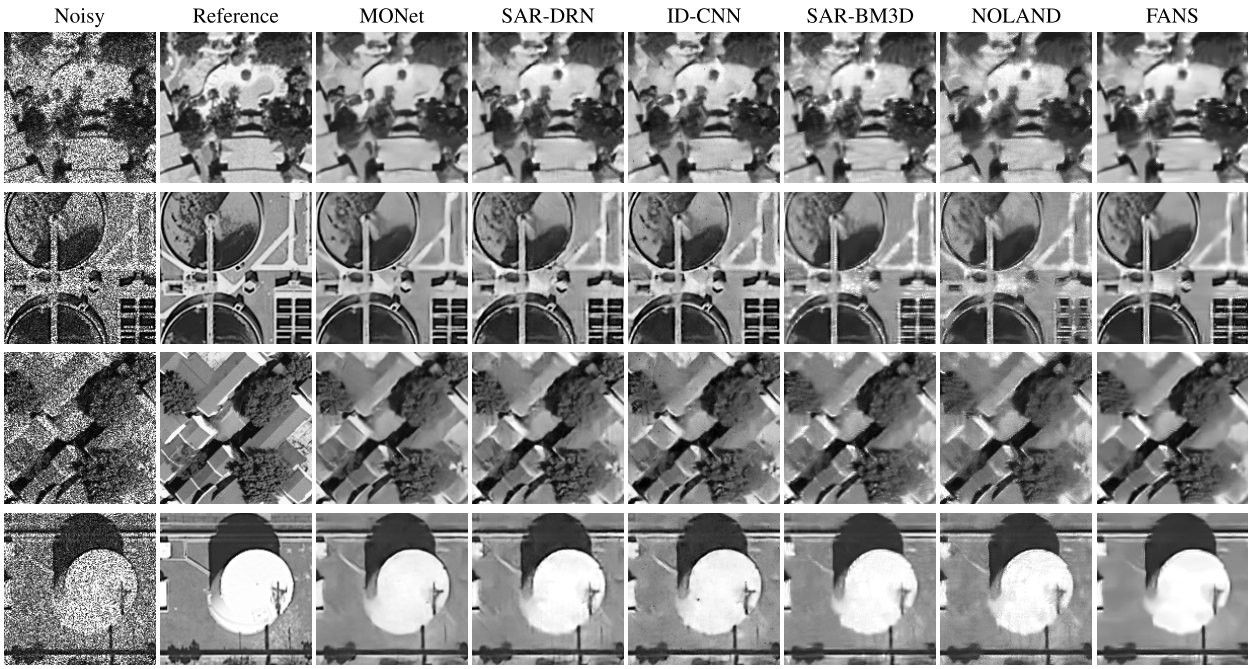

Fig. 7. Results on a subset of the simulated images, from left to right: simulated noisy image, noise-free reference, MONet, SAR-DRN, ID-CNN, SARBM3D, NOLAND, and FANS.

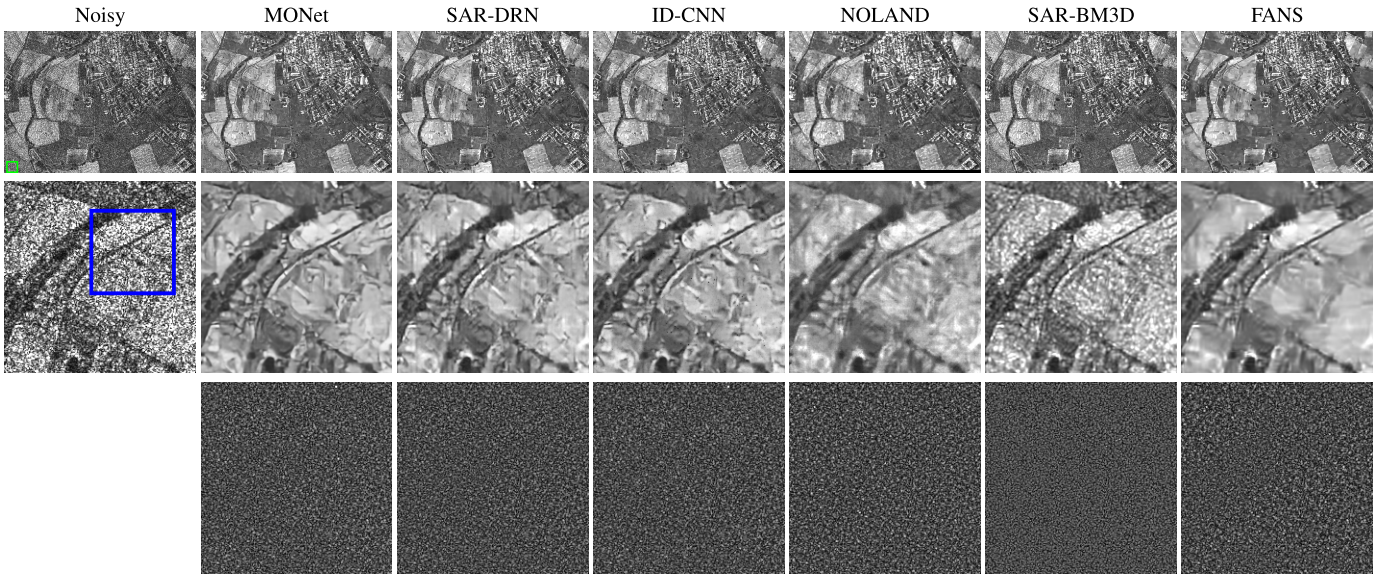

Fig. 8. Results on CSK image: scene under test in the top row; details of the image in the second row; and corresponding ratio image in the third row.

C. Result on Real-SAR Images

In this section, we show the results of the proposed algorithm on real-SAR data. Four different test cases have been considered using images acquired by different sensors [COSMO-SkyMed (CSK), TerraSAR-X, RADARSAT2, and Sentinel-1), working bands (X and C), acquisition modalities (strip map and interferometric wide swath), resolutions (three and five meters), and polarization (HH and VH). These have been done in order to show the independence of the achievable results with respect to the considered data sets.

In Figs. 8–11, the noisy images (first column) and the results on the considered algorithms (columns 2–7) are shown. In the first row, the results of the whole image are shown. In the second row, the results on a particular patch of the whole image are presented. The corresponding ratio images are in the third row. Because of the lack of a reference, it is difficult to find a metric that can fairly evaluate the filters given they rely on certain mathematical assumptions for the speckle that is not sure are confirmed in the real-SAR image under test. For this reason, the evaluation of filtering performance mostly relies on visual inspection, considering the ability to suppress noise while preserving objects in the scene. To this aim, also the ratio images produced by each method are shown. As noticed in the simulated results, FANS has good edges preservation but produces oversmoothed results on homogeneous areas.

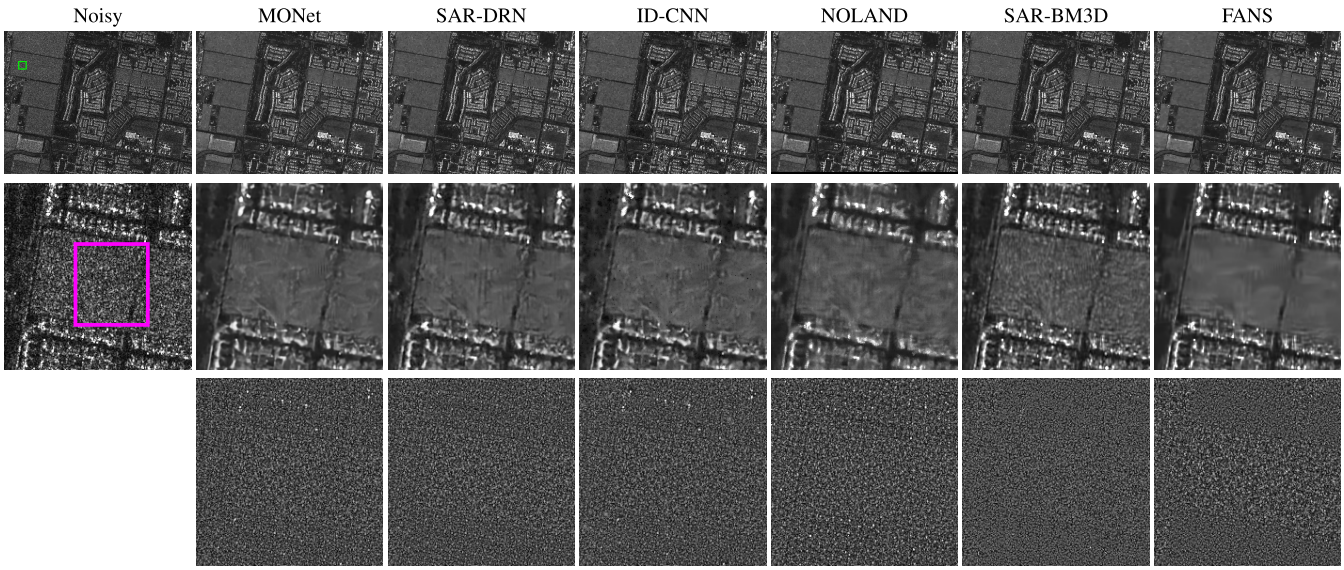

Fig. 9. Results on RADARSAT2 image: scene under test in the top row; details of the image in the second row; and corresponding ratio image in the third row.

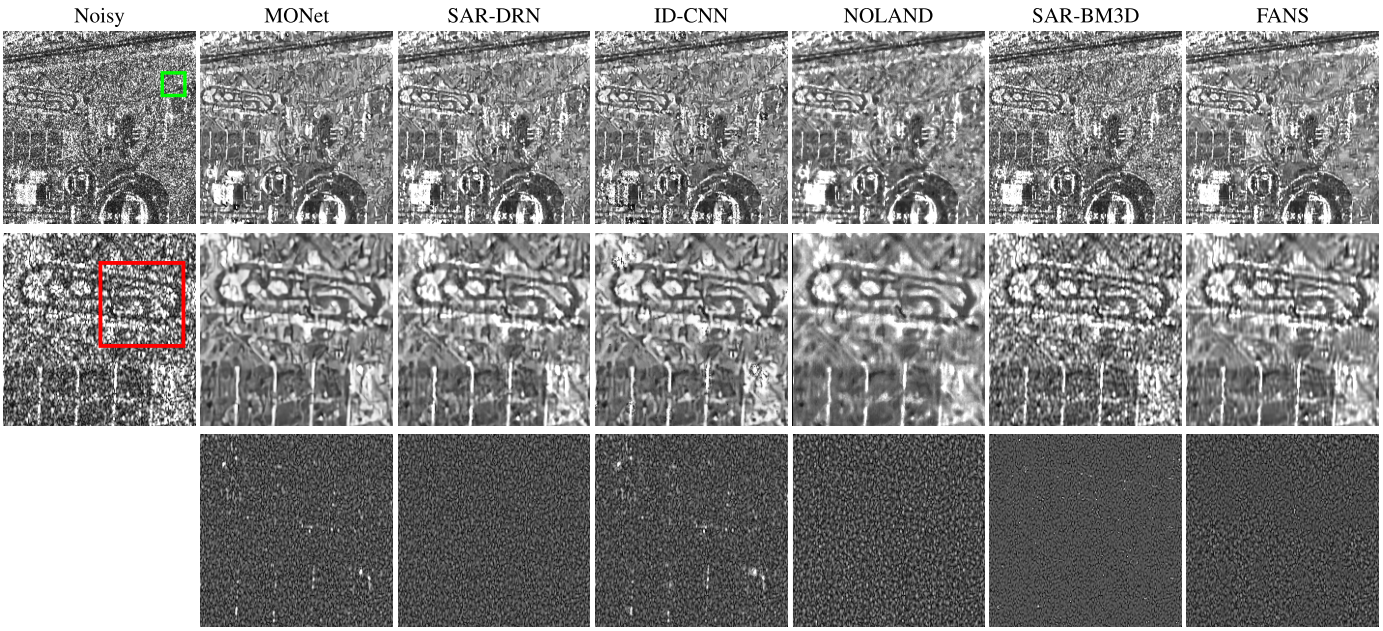

Fig. 10. Results on TSX image: scene under test in the top row; details of the image in the second row; and corresponding ratio image in the third row.

NOLAND better preserves spatial details than FANS, but it is still smooth. SAR-BM3D has the best edges and objects preservation among the NL filters, but the noise is still present on the filtered images. In general, the DL solutions try to more suppress the noise compared with the NL approaches. MONet shows a good tradeoff between noise suppression and edges preservation: in homogeneous areas, noise is removed without losing many spatial details. Moreover, at the same time, the edges are quite well preserved. Similar considerations can be done for SAR-DRN and IDCNN, but both of them produces some artifacts: the former introduces a vertical texture in all the images and produces disturbed edges, generating less clean images; the latter has good edges preservation but less suppresses the noise with respect the other two and produce some black spots.

These considerations can be appreciated on the details and on the relative ratio images shown in Figs. 8–11.

For example, in the COSMO-SkyMed image of Fig. 8, the boundaries of the road are retained quite well from MONet, and homogeneous areas are reach of spatial details not deleted by the noise removal.

These spatial details barely appear in the NL approaches (except for SAR-BM3D), while edges are well defined.

Moving to the ratio image, it must be recalled that an ideal filter should produce an uncorrelated ratio image: more correlation and more structures are visible in the ratio, worse

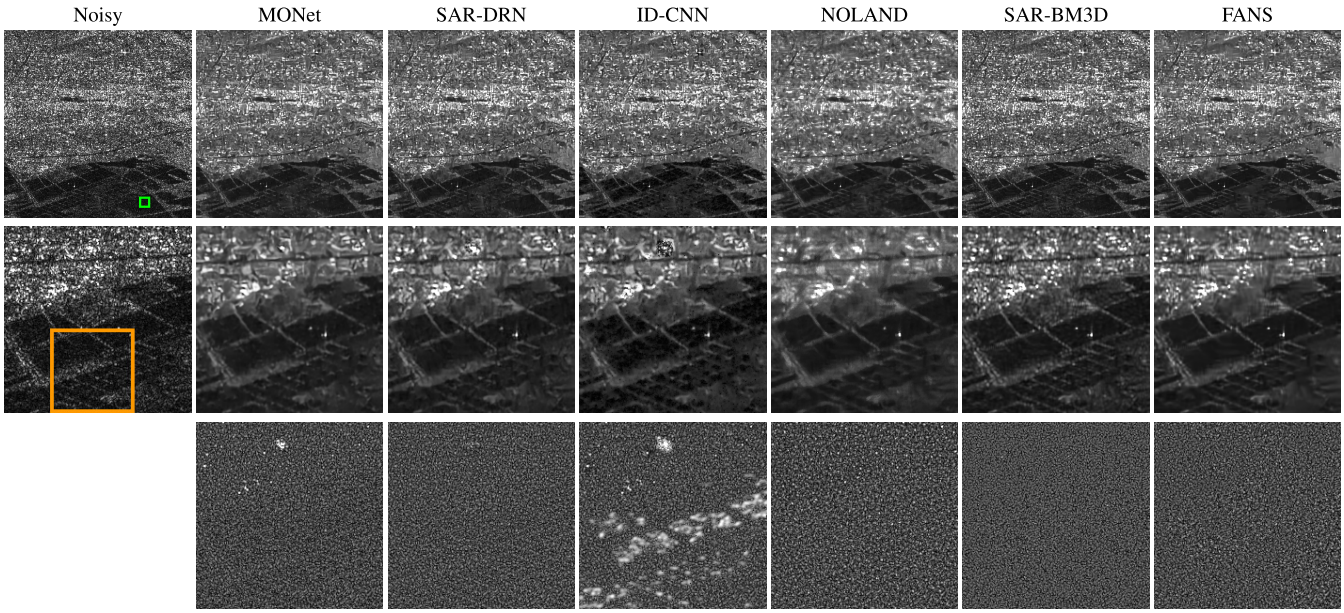

Fig. 11. Results on Sentinel-1 image: scene under test in the top row; details of the image in the second row; and corresponding ratio image in the third row.

is the filtering effect. From the ratio images, the road is more visible for SAR-DRN respects the others: meaning it is heavily filtered and not well preserved. The ratio images of IDCNN and MONet are very similar to each other with some emergent structure for the former. The NOLAND ratio image looks almost uncorrelated, but it is characterized by a large granularity typical of generalized smoothness. Contrary, SAR-BM3D has a ratio image with a very tiny granularity typical of good object preservation but also of not perfect noise suppression. From the FANS ratio image, it is easily observable a different behavior of the filter in different areas: large granularity on homogeneous areas proving its oversmoothing effect and very tiny granularity in correspondence of not homogeneous areas.

The same behavior can be appreciated on the RADARSAT2 image in Fig. 9. Even if some structures are more highlighted in the ratio image for MONet, it is still going to have better edges preservation than other methods, except for SAR-BM3D and FANS. At the same time, these two solutions still present their limitation: the presence of residual noise for SAR-BM3D and oversmoothing for FANS. In Fig. 10, the results for the TerrraSAR-X (TSX) image are shown. This image is very challenging for all the solutions, but generally, the previous considerations are still valid. SAR-BM3D is still noisy, FANS tends to oversmooth, and NOLAND has good detail preservation. The results of MONet and SAR-DRN are similar, while ID-CNN produces many artifacts. Observing the ratio images, the proposed solution produces less structure compared with other DL methods, meaning better preservation of details. It can be noted, in correspondence of urban structures, the appearance of textures within the proposed solutions. Such effect can be expected due to invalid fully developed hypotheses for such areas (i.e., extremely heterogeneous areas). This point will be better discussed and analyzed in Section II-E.

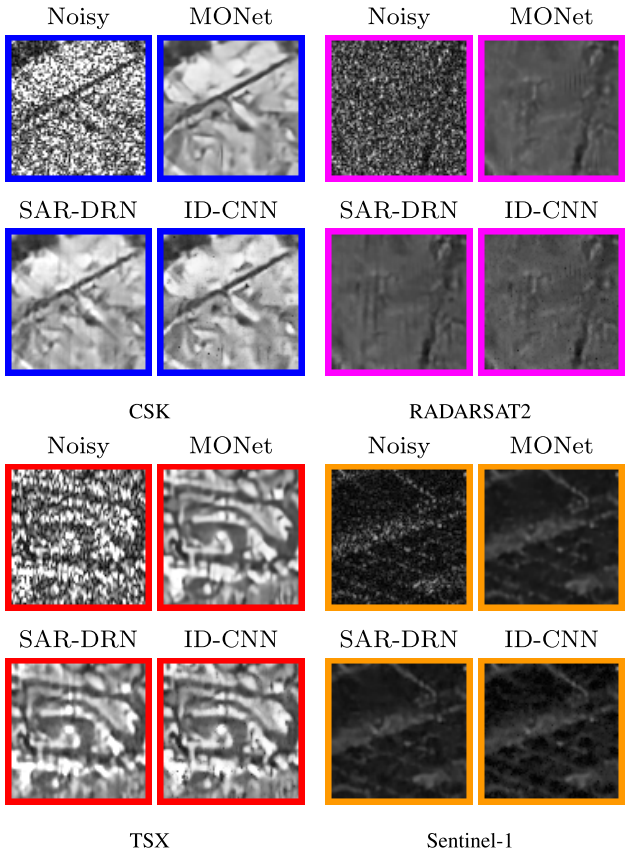

Fig. 12. Zoomed-in detail of blue, magenta, red, and orange boxes of Figs. 8–11 from CSK, RADARSAT2, TSX, and Sentinel-1 data sets, respectively.

The results on Sentinel-1 image are shown in Fig. 11. The NL solutions have good edge preservation and generally good noise suppression. SAR-DRN and MONet have similar performance, while ID-CNN suffers on the points characterized by low amplitude value.

In order to better spot the aforementioned differences among the DL methods, the zooms of the details highlighted in the blue, magenta, red, and yellow square boxes of Figs. 8–11 for the correspondent CSK, RADARSAT2, TSX, and Sentinel-1 data sets are shown in Fig. 12.

For CSK and RADARSAT2 details, MONet has better edges preservation, SAR-DRN shows smoothness and a vertical texture, ID-CNN is a bit more noisy and presents some black artifacts.

Regarding the zoom of TSX, it can be noted how the proposed MONet tends to enforce smoothness on homogeneous areas but, at the same time, preserves edges with slightly better quality compared with SAR-DRN. Indeed, the path inside the zoom is better-retained with respect to IDCNN and SAR-DRN.

The performance of the filters is quite similar to the detail of Sentinel-1, where ID-CNN always shows some black spots.

In the end, the proposed MONet shows edge preservation comparable with NL approaches but with better noise suppression resulting in very good objects and details preservation. Compared with SAR-DRN and IDCNN, it seems that the depth of the network combined with the use of the defined cost function helps in suppressing the noise and, at the same time, keeping intact some details, such as edges and small object.

In addition to visual comparison, the numerical assessment for each site under test has been carried out. The M-index is considered, from which the Haralick homogeneity δh, the residual ENL $r_{\widehat{\text{ENL}}}$, and the mean of the ratio μ_N has been extracted. Moreover, the D_{KL} between the pdf of the predicted speckle and the Rayleigh distribution has been reported.

Regarding the M-index, NOLAND always has the best value, followed by MONet, except for the TSX image, where the second best is FANS. In order to interpret these results, the three factors δh, $r_{\widehat{\text{ENL}}}$, and μ_N, whose M-index is a combination, have been extracted. Lower is δh, less is the remaining structure, and higher is the detail preservation during the noise suppression. Lower is $r_{\widehat{\text{ENL}}}$, the ENLs computed on the ratio image are closer to the ENLs computed to the noisy, meaning better statistical preservation of the noise. MONet always shows the best or the second best value for δh, confirming a better details preservation w.r.t other methods. Indeed, the other methods produce more artifacts, and the ratio images highlight more structures. The reverse is the situation for the performance on the $r_{\widehat{\text{ENL}}}$: MONet is always surpassed by NOLAND and FANS.

In general, from Tables III–V, we can see that DL methods outperform NL methods on δh, but the situation is reverted on $r_{\widehat{\text{ENL}}}$. This can be explained by the fact that DL methods are trained under the fully developed hypothesis that is not correct everywhere inside the images, and therefore, the statistical $r_{\widehat{\text{ENL}}}$ highlights this characteristic.

Moreover, together with the $r_{\widehat{\text{ENL}}}$, the mean value of the ratio images μ_N has been extracted. The ideal filter should produce a mean ratio equal to one. Except for the RADARSAT2, where the proposed solution reaches the lowest performance, in the CSK and TSX, it reaches the best and second-best performance, respectively, confirming a good quality filtering process.

For considering the ability in noise suppression, the ENL on homogeneous areas for each image under test has been computed. The selected areas are highlighted in the green boxes, and the corresponding ENL of the noisy images are shown in Tables III–VI. In general, the ENL performance is very close to all the methods, except for FANS that strongly outperforms the others. This can be explained by the over-smooth behavior of FANS with respect to the other solutions. Regarding the D_{KL}, we can see that NOLAND always has

TABLE III
NUMERICAL ASSESSMENT CSK

	M-index	δh	$r_{\widehat{ENL}}$	μ_N	D_{KL}	ENL
Noisy	-	-	-	-	-	0.995
FANS	16.96	0.2003	**138**	0.8750	0.0224	**36**
SAR-BM3D	19.81	0.0275	368	0.8790	0.1181	7
NOLAND	**10.61**	0.0507	161	0.8841	**0.0068**	15
ID-CNN	12.64	0.0021	250	0.8758	0.0235	12
SAR-DRN	11.14	**0.0001**	222	0.8786	0.0393	15
MONet	10.69	0.0002	213	**0.8903**	0.0178	14

TABLE IV
NUMERICAL ASSESSMENT RADARSAT2

	M-index	δh	$r_{\widehat{ENL}}$	μ_N	D_{KL}	ENL
Noisy	-	-	-	-	-	1
FANS	12.40	0.1412	**106**	0.8786	0.0206	**69**
SAR-BM3D	16.42	0.0500	278	0.8788	0.0906	15
NOLAND	**8.75**	0.0383	136	**0.8844**	**0.0058**	21
ID-CNN	9.52	**0.0001**	257	0.8739	0.0346	18
SAR-DRN	12.90	0.0010	190	0.8779	0.0404	22
MONet	9.42	**0.0001**	188	0.8677	0.0254	29

TABLE V
NUMERICAL ASSESSMENT TSX—TEHRAN

	M-index	δh	$r_{\widehat{ENL}}$	μ_N	D_{KL}	ENL
Noisy	-	-	-	-	-	0.999
FANS	21,96	0.0140	42	0.8583	0.2376	**7**
SAR-BM3D	56.37	0.0882	103	**0.9026**	0.3975	2
NOLAND	**14.07**	0.0460	**23**	0.8944	**0.0193**	6
ID-CNN	26.79	0.0289	50	0.8855	0.0767	5
SAR-DRN	23.96	0.0006	47	0.8829	0.1348	5
MONet	24.34	**0.0001**	48	0.9006	0.0651	5

TABLE VI
NUMERICAL ASSESSMENT SENTINEL-1

	M-index	δh	$r_{\widehat{ENL}}$	μ_N	D_{KL}	ENL
Noisy	-	-	-	-	-	1
FANS	16.66	0.1465	19	0.8729	0.0364	**30**
SAR-BM3D	25.15	0.0673	44	0.8783	0.1702	7
NOLAND	**9.94**	0.0170	**18**	0.8776	**0.0077**	16
ID-CNN	18.39	0.0033	36	**0.9100**	0.0281	12
SAR-DRN	14.02	0.0001	28	0.8731	0.0578	14
MONet	13.99	**0.0000**	27	0.8741	0.0344	14

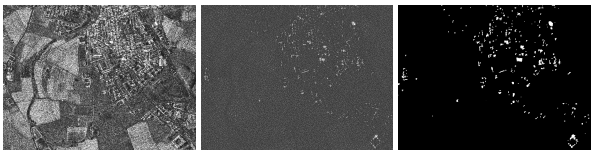

Fig. 13. Result of the detection process on CSK data set, from left to right: SAR image; ratio image produced by MONet; and detection result.

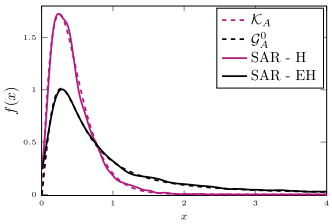

Fig. 14. Comparison between the distributions of two different areas of SAR image (solid) with the theoretical ones (dashed): black solid curve represents the distribution of extremely heterogeneous detected points on SAR image (SAR-EH); the black dashed curve is the theoretical $\mathcal{G}_A^0(-0.5, 0.145, 1)$; magenta solid curve represents the distribution of a heterogeneous area of SAR image (SAR-H); and magenta dashed is the theoretical $\mathcal{K}_A(2, 7.5, 1)$.

the best performance. This thanks to the fact that D_{KL} is included in the similarity research process. Proposed MONet has the second-best performance on CSK and on TSX, while on RADARSAT2, it reaches the third one. Naturally, the D_{KL} results are affected by the presence of not homogeneous areas, and therefore, they are rather general. It is worth to notice that among the DL methods, MONet always has the best D_{KL} index. This means that using a statistical term as $\mathcal{L}_{\mathrm{KL}}$ gives the network added useful statistical information that cannot be acquired only by the data.

D. Identification of Not Fully Developed Areas: Validation

It is worth to notice that our network is trained under the fully developed hypothesis, and the use of $\mathcal{L}_\nabla$ aims in preserving objects, details, and strong scatterers, where that hypothesis is not valid anymore. These points strongly appear in the ratio images produced by the proposed method. As described in Section II-E, an identification step allows isolating such points leaving the user the possibility to decide the filtering policy.

The validation of this procedure is performed in the following only on the CSK data set; however, similar results can be achieved using the other data sets. In Fig. 13, the detection map of not fully developed points are shown for CSK. In Fig. 14, it is shown how the detected points on the SAR image (SAR Extremely Heterogeneous points, SAR-EH) generate a pdf (solid magenta curve) that well fits the theoretical distribution of $\mathcal{G}_A^0(\alpha, \gamma, L)$ (dashed magenta) of [14] as the distribution that better describes such areas. The parameters are empirically estimated as ($\alpha = -0.5, \gamma = 0.145$, and $\lambda = 0, L = 1$).

At the same time, we estimated the distribution of the SAR image in the remaining points (SAR Heterogeneous points, SAR-H): this fits the $\mathcal{K}_A(\alpha, \lambda, L)$ distribution, meaning that all the remaining part of the image belongs to heterogeneous areas. The parameters are empirically set as ($\alpha = 2, \gamma = 0, \lambda = 7.5$, and $L = 1$).

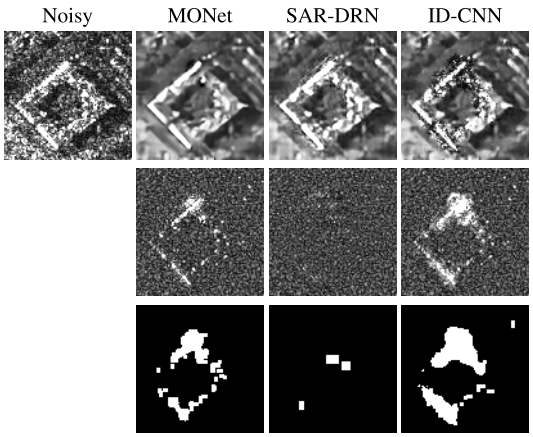

Fig. 15. Comparison of the detection of strong scatterers between the proposed MONet, SAR-DRN, and ID-CNN. In the top noisy and ratio images are shown. In the bottom, the relative detection.

This confirms the fact that our CNN is able to detect the points belonging to the extremely heterogeneous areas directly from the ratio image. Naturally, this issue is in common with all the CNN that use training data simulated under the fully developed hypothesis. Therefore, this procedure could also be extended to other methods, such as ID-CNN and SAR-DRN. In Fig. 15, a patch of CSK is shown with relative detection for the DL methods. First of all, it is important to note the different behavior of three CNNs on strong scatterers: MONet try to isolate the objects by preserving the edges and at the same time produce a strong structure in the ratio; contrary SAR-DRN try to less filter these elements, but some distortion are visible both in the filtered image and in the ratio. ID-CNN produces many artifacts not only in correspondence of the scatterers but also in its neighborhood.

By visually inspecting the amplitude image and the detection results, it is evident that many point scatterers are not correctly identified by SAR-DRN, much false identification is present in ID-CNN, while the MONet seems the most reliable.

IV. Conclusion

In this article, a CNN for SAR despeckling trained on simulated data has been proposed. The nonlinearity introduced by the 17 layers is crucial for feature extraction, while skip connections are used for avoiding the vanishing gradient problem. Beyond the proposed architecture, the main focus is dedicated to the definition of a multi-objective cost function composed of three terms: $\mathcal{L}_2$, $\mathcal{L}_{\mathrm{KL}}$, and $\mathcal{L}_\nabla$. The combination of these three terms allows the preservation of spatial details, statistical properties, edges, and identification of strong scatterers. An ablation study proves how the combination is crucial for taking care of these aspects simultaneously. Experimental validation, both on simulated and real data, shows the advantages of including these SAR image properties in the cost function.

The performance on simulated images shows an improvement with respect to the state-of-the-art, mainly on edges and details preservation. This is also confirmed in real-SAR images, where the results present good noise rejection, edges preservation, and absence of artifacts. This means more clear filtered images with well-retained edges and objects.

TABLE VII

NUMERICAL ASSESSMENT ON SIMULATED DATA SET FOR DIFFERENT COST FUNCTIONS: THE VALUE ARE AVERAGED ON THE WHOLE SIMULATED TESTING DATA SET COMPOSED OF 100 IMAGES. FROM TOP TO BOTTOM: NETWORK TRAINED WITH L_2, L_{kl}, L_∇, AND $\mathcal{L}$

	SSIM	**SNR**	**MSE**
L_2	0,7509	8,8514	0,0040
L_{kl}	0,7514	8,8543	0,0039
L_∇	0,7512	8,8585	0,0039
$\mathcal{L}$	0,7510	8,8555	0,0039

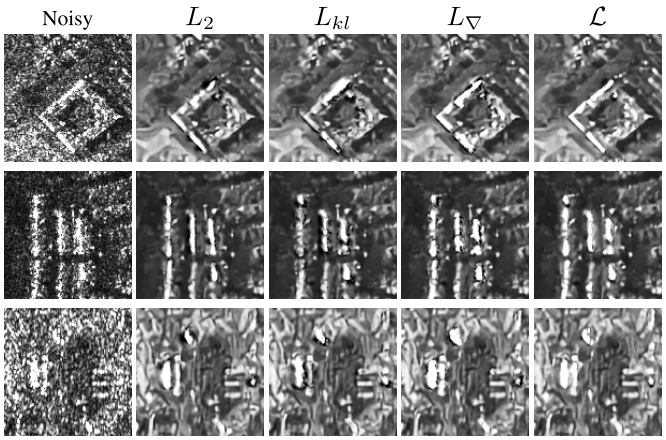

Fig. 16. Details for the different cost function.

Moving to real data, MONet based on the fully developed hypothesis and a cascade of convolutions cannot directly handle the point scatterers, but differently from other CNN-based solutions (such as SAR-DRN and ID-CNN), it is able to identify them. Hence, the knowledge of those pixels allows the final user to decide how to process them (left unfiltered, define a specific statistical-based filter, using a multitemporal approach, and so on).

Being a DL-based method, once the network training is performed, the computational time is limited. Further works will address the possibility of adapting the filter to multilook, multitemporal, and multichannel SAR images.

APPENDIX
ON THE IMPACT OF THE COST FUNCTION

In this section, an ablation study has been carried out in order to assess the impact of the defined cost function. The cost function is given by a combination of the terms in (4). In order to compare the performance and the impact of these three terms, the same architecture is trained on the same data set with a cost function composed once only of the $L_2 = \mathcal{L}_2$ term, once of the combination $L_{\mathrm{kl}} = \mathcal{L}_2 + \lambda_{\mathrm{kl}}\mathcal{L}_{\mathrm{KL}}$, and once with the combination $L_\nabla = \mathcal{L}_2 + \lambda_\nabla \mathcal{L}_\nabla$. These solutions are compared with the proposed method. In Table VII, we summarize the numerical assessment on the same testing data set of Section III-B. The results are almost the same for each solution like there is no difference in introducing such terms in the cost function. It seems that $\mathcal{L}_2$ is enough for the despeckling. However, these are average metrics that do not take into account the details that make the difference between one solution and another. Moving to real data, things largely change.

In Fig. 16, a detail for each data set CSK, RADARSAT2, and TSX are shown. It can be noted how important the impact of the cost function is. Starting from L_2 that tries to preserve spatial details, the use of the KL divergence in L_{kl} helps in filtering the homogeneous areas, but we lose information on strong scatterers producing a soothing effect. In addition, the L_∇ tries to preserve edges but does not consider the speckle properties and tends to create strange artifacts in the neighborhood of the strong scatterers. The proposed cost function $\mathcal{L}$ is able to balance all these effects and to give the best compromise. The use of $\mathcal{L}$ allows the filter the image balancing at the same time the statistical properties of the noise and the details preservation.

ACKNOWLEDGMENT

The COSMO-SkyMed (CSK) data have been provided by Agenzia Spaziale Italiana (ASI) within the project CSK SAR data - Contract n. I/065/09/0. The TerrraSAR-X data have been provided by Deutsches Zentrum für Luft- und Raumfahrt German Aerospace Center (DLR) within the framework of the Project MTH3649. The RS2 data are free and available at https://mdacorporation.com. RADARSAT-2 Data and Products by MacDonald, Dettwiler and Associates, Ltd., Westminster, CO, USA—all rights reserved. RADARSAT is an official mark of the Canadian Space Agency, Longueuil, QC, Canada. Sentinel data are provided by the European Space Agency (ESA)–Copernicus Sentinel Data 2018, processed by ESA.

REFERENCES

[1] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-2, no. 2, pp. 165–168, Mar. 1980.

[2] V. S. Frost, J. A. Stiles, K. S. Shanmugan, and J. C. Holtzman, "A model for radar images and its application to adaptive digital filtering of multiplicative noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-4, no. 2, pp. 157–166, Mar. 1982.

[3] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, "Adaptive noise smoothing filter for images with signal-dependent noise," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-7, no. 2, pp. 165–177, Mar. 1985.

[4] R. Touzi, A. Lopes, and P. Bousquet, "A statistical and geometrical edge detector for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. GRS-26, no. 6, pp. 764–773, Nov. 1988.

[5] A. Lopes, R. Touzi, and E. Nezry, "Adaptive speckle filters and scene heterogeneity," *IEEE Trans. Geosci. Remote Sens.*, vol. 28, no. 6, pp. 992–1000, Nov. 1990.

[6] D. Kuan, A. Sawchuk, T. Strand, and P. Chavel, "Adaptive restoration of images with speckle," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 3, pp. 373–383, Mar. 1987.

[7] A. Lopes, E. Nezry, R. Touzi, and H. Laur, "Maximum a posteriori speckle filtering and first order texture models in SAR images," in *Proc. 10th Annu. Int. Symp. Geosci. Remote Sens.*, 1990, pp. 2409–2412.

[8] H. Guo, J. E. Odegard, M. Lang, R. A. Gopinath, I. W. Selesnick, and C. S. Burrus, "Wavelet based speckle reduction with application to SAR based ATD/R," in *Proc. 1st Int. Conf. Image Process.*, 1994, vol. 1, pp. 75–79.

[9] G. Franceschetti, V. Pascazio, and G. Schirinzi, "Iterative homomorphic technique for speckle reduction in synthetic-aperture radar imaging," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 12, no. 4, pp. 686–694, 1995.

[10] L. Gagnon and A. Jouan, "Speckle filtering of SAR images: A comparative study between complex-wavelet-based and standard filters," in *Proc. SPIE, 5th Int. Soc. Opt. Photon., Wavelet Appl. Signal Image Process.*, vol. 3169, A. Aldroubi, A. F. Laine, and M. A. Unser, Eds., 1997, pp. 80–91.

[11] F. Argenti and L. Alparone, "Speckle removal from SAR images in the undecimated wavelet domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2363–2374, Nov. 2002.

[12] 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.

[13] J. W. Goodman, "Statistical properties of laser speckle patterns," in *Laser Speckle and Related Phenomena*, J. C. Dainty, Ed. Berlin, Germany: Springer, 1984, ch. 2.

[14] A. C. Frery, H.-J. Muller, C. C. F. Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Trans. Geosci. Remote Sens.*, vol. 35, no. 3, pp. 648–659, May 1997.

[15] P. Coupe, P. Hellier, C. Kervrann, and C. Barillot, "Bayesian non local means-based speckle filtering," in *Proc. 5th IEEE Int. Symp. Biomed. Imag., Nano Macro*, May 2008, pp. 1291–1294.

[16] H. Zhong, J. Xu, and L. Jiao, "Classification based nonlocal means despeckling for SAR image," in *Proc. Int. Soc. Opt. Photon., Autom. Target Recognit. Image Anal. (MIPPR)*, T. Zhang, B. Hirsch, Z. Cao, and H. Lu, Eds., vol. 7495, 2009, pp. 231–238.

[17] H. Feng, B. Hou, and M. Gong, "SAR image despeckling based on local homogeneous-region segmentation by using pixel-relativity measurement," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 7, pp. 2724–2737, Jul. 2011.

[18] G. Ferraioli, V. Pascazio, and G. Schirinzi, "Ratio-based nonlocal anisotropic despeckling approach for SAR images," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 10, pp. 7785–7798, Oct. 2019.

[19] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage," *IEEE Trans. Geosci. Remote Sens.*, vol. 50, no. 2, pp. 606–616, Feb. 2012.

[20] H. Zhong, Y. Li, and L. C. Jiao, "SAR image despeckling using Bayesian nonlocal means filter with sigma preselection," *IEEE Geosci. Remote Sens. Lett.*, vol. 8, no. 4, pp. 809–813, Jul. 2011.

[21] D. Cozzolino, S. Parrilli, G. Scarpa, G. Poggi, and L. Verdoliva, "Fast adaptive nonlocal SAR despeckling," *IEEE Geosci. Remote Sens. Lett.*, vol. 11, no. 2, pp. 524–528, Feb. 2014.

[22] R. Touzi, "A review of speckle filtering in the context of estimation theory," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 11, pp. 2392–2404, Nov. 2002.

[23] C.-A. Deledalle, L. Denis, G. Poggi, F. Tupin, and L. Verdoliva, "Exploiting patch similarity for SAR image processing: The nonlocal paradigm," *IEEE Signal Process. Mag.*, vol. 31, no. 4, pp. 69–78, Jul. 2014.

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

[25] Q. Zhang, Q. Yuan, J. Li, Z. Yang, and X. Ma, "Learning a dilated residual network for SAR image despeckling," *Remote Sens.*, vol. 10, no. 2, pp. 1–18, 2018.

[26] F. Lattari, B. Gonzalez Leon, F. Asaro, A. Rucci, C. Prati, and M. Matteucci, "Deep learning for SAR image despeckling," *Remote Sens.*, vol. 11, no. 13, p. 1532, Jun. 2019.

[27] X. Yang, L. Denis, F. Tupin, and W. Yang, "SAR image despeckling using pre-trained convolutional neural network models," in *Proc. Joint Urban Remote Sens. Event (JURSE)*, May 2019, pp. 1–4.

[28] L. Denis, C.-A. Deledalle, and F. Tupin, "From patches to deep learning: Combining self-similarity and neural networks for SAR image despeckling," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 5113–5116.

[29] 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, p. 4389–4403, Sep. 2017.

[30] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva, "SAR image despeckling through convolutional neural networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2017, pp. 5438–5441.

[31] D. Cozzolino, L. Verdoliva, G. Scarpa, and G. Poggi, "Nonlocal CNN SAR image despeckling," *Remote Sens.*, vol. 12, no. 6, p. 1006, Mar. 2020.

[32] X. Ma, C. Wang, Z. Yin, and P. Wu, "SAR image despeckling by noisy reference-based deep learning method," *IEEE Trans. Geosci. Remote Sens.*, early access, May 11, 2020, doi: 10.1109/TGRS.2020.2990978.

[33] J. Lehtinen *et al.*, "Noise2Noise: Learning image restoration without clean data," in *Proc. 35th Int. Conf. Mach. Learn.*, J. Dy and A. Krause, Eds., Stockholm, Sweden, vol. 80, Jul. 2018, pp. 2965–2974.

[34] S. Vitale, G. Ferraioli, and V. Pascazio, "A new ratio image based CNN algorithm for SAR despeckling," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 9494–9497.

[35] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone, "A tutorial on speckle reduction in synthetic aperture radar images," *IEEE Geosci. Remote Sens. Mag.*, vol. 1, no. 3, pp. 6–35, Sep. 2013.

[36] 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.

[37] D.-X. Yue, F. Xu, A. C. Frery, and Y.-Q. Jin, "A generalized Gaussian coherent scatterer model for correlated SAR texture," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2947–2964, Apr. 2020.

[38] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proc. 18th Int. Conf. Adv. Geographic Inf. Syst. (SIGSPATIAL GIS)*, 2010, pp. 270–279.

[39] S. Vitale, G. Ferraioli, and V. Pascazio, "Edge preserving CNN SAR despeckling algorithm," in *Proc. IEEE Latin Amer. GRSS ISPRS Remote Sens. Conf. (LAGIRS)*, Mar. 2020, pp. 12–15.

[40] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.

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

[42] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, pp. 770–778, Jun. 2015.

[43] L. Xu, J. Ren, Q. Yan, R. Liao, and J. Jia, "Deep edge-aware filters," in *Proc. 32nd Int. Conf. Mach. Learn. (PMLR)*, F. Bach and D. Blei, Eds., Lille, France, vol. 37, Jul. 2015, pp. 1669–1678.

[44] A. Arienzo, F. Argenti, L. Alparone, and M. Gherardelli, "Accurate despeckling and estimation of polarimetric features by means of a spatial decorrelation of the noise in complex PolSAR data," *Remote Sens.*, vol. 12, no. 2, p. 331, Jan. 2020.

[45] S. Vitale, D. Cozzolino, G. Scarpa, L. Verdoliva, and G. Poggi, "Guided patchwise nonlocal SAR despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 9, pp. 6484–6498, Sep. 2019.

[46] S. Vitale, G. Ferraioli, and V. Pascazio, "Complexity analysis of an edge preserving CNN SAR despeckling algorithm," Apr. 2020, *arXiv:2004.08345*. [Online]. Available: http://arxiv.org/abs/2004.08345

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

[48] L. Gomez, R. Ospina, and A. Frery, "Unassisted quantitative evaluation of despeckling filters," *Remote Sens.*, vol. 9, no. 4, p. 389, Apr. 2017.

[49] R. M. Haralick, K. Shanmugam, and I. Dinstein, "Textural features for image classification," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, no. 6, pp. 610–621, Nov. 1973.

Sergio Vitale (Graduate Student Member, IEEE) received the M.S. degree *(summa cum laude)* in telecommunication engineering from the University of Napoli Federico II, Naples, Italy, in 2017. He is pursuing the Ph.D. degree in information and communication technology and engineering with University Parthenope, Naples.

His research interests include synthetic aperture radar imaging, image enhancement, super-resolution, and deep learning.

Giampaolo Ferraioli (Senior Member, IEEE) was born in Lagonegro, Italy, in 1982. He received the B.S. and M.S. degrees and the Ph.D. degree in telecommunication engineering from the University Parthenope, Naples, Italy, in 2003, 2005, and 2008, respectively.

He was a Visiting Scientist with the Département Traitement du Signal et des Images (TSI), Télécom ParisTech, Paris, France. He is an Assistant Professor with University Parthenope. His main research interests deal with statistical signal and image processing, radar systems, synthetic aperture radar interferometry, image restoration, and magnetic resonance imaging.

Dr. Ferraioli is on the Editorial Board of *MDPI Remote Sensing*. He is a member of the Technical Liaison Committee for the IEEE TRANSACTIONS ON COMPUTATIONAL IMAGING. He received the IEEE 2009 Best European Ph.D. Thesis in Remote Sensing prize from the IEEE Geoscience and Remote Sensing Society in 2009. He serves as an Associate Editor for the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS.

Vito Pascazio (Senior Member, IEEE) received the M.S. degree *(summa cum laude)* in electronic engineering from the University of Bari Aldo Moro, Bari, Italy, in 1986, and the Ph.D. degree in electronic engineering and computer science from the Department of Electronic Engineering, University of Napoli Federico II, Naples, Italy, in 1990.

In 1990, he was with the Research Institute on Electromagnetics and Electronic Devices, Italian National Council of Research, Naples. He is a Full Professor and the Chair of the Department of Engineering, University Parthenope, Naples. He was a Visiting Scientist with the Laboratoire des Signaux et Systemes, Ecole Superieure d' Electricite (Supelec), Gif sur Yvette, France, from 1994 to 1995, and the Nice Sophia Antipolis University, Nice, France, from 1998 to 1999. He was also the Director of the National Laboratory of Multimedia Communications, Italian National Consortium of Telecommunications, Naples. He has authored or coauthored more than 180 technical articles. His research interests include remote sensing, image processing, and linear and nonlinear estimation, with an emphasis on image computing and processing and reconstruction of microwave and radar images.

Dr. Pascazio was the General Co-Chairman of the International Geoscience and Remote Sensing Symposium (IGARSS)-2015 conference. He was a recipient of the Philip Morris Prize for Scientific and Technological Research in 1990.