As if by magic: self-supervised training of deep despeckling networks with MERLIN

Emanuele Dalsasso, Loïc Denis, Florence Tupin

As if by magic: self-supervised training of deep despeckling networks with MERLIN

Emanuele Dalsasso, Loïc Denis, Florence Tupin

***Abstract*—Speckle fluctuations seriously limit the interpretability of synthetic aperture radar (SAR) images. Speckle reduction has thus been the subject of numerous works spanning at least four decades. Techniques based on deep neural networks have recently achieved a new level of performance in terms of SAR image restoration quality.**

Beyond the design of suitable network architectures or the selection of adequate loss functions, the construction of training sets is of uttermost importance. So far, most approaches have considered a supervised training strategy: the networks are trained to produce outputs as close as possible to speckle-free reference images. Speckle-free images are generally not available, which requires resorting to speckle simulations or the selection of stable areas in long time series to circumvent the lack of ground truth. Self-supervision, on the other hand, avoids the use of speckle-free images.

We introduce a self-supervised strategy based on the separation of the real and imaginary parts of single-look complex SAR images, called MERLIN (coMplex sElf-supeRvised despeckLINg), and show that it offers a straightforward way to train all kinds of deep despeckling networks. Networks trained with MERLIN take into account the spatial correlations due to the SAR transfer function specific to a given sensor and imaging mode. By requiring only a single image, and possibly exploiting large archives, MERLIN opens the door to hassle-free as well as large-scale training of despeckling networks. The code of the trained models is made freely available at https://gitlab.telecom-paris.fr/RING/MERLIN.

***Index Terms*—SAR, image despeckling, deep learning, self-supervised training.**

I. Introduction

THE speckle phenomenon occurs due to the coherent summation of many elementary echoes within a radar resolution cell. It is responsible for the strong fluctuations that dramatically degrade the quality of SAR images. Speckle suppression has been the subject of many research works, from the pioneering works of Lee [1] to the most recent techniques based on deep neural networks [2]–[4].

The first approaches developed to reduce the speckle fluctuations were based on a local averaging of the pixel intensities within a small window. To prevent the spreading of bright targets over the whole window, selection techniques were introduced to exclude pixels with values too different from the value at the reference pixel [1], [5], or to select an oriented window with a homogeneous content [6]. Regularization techniques formulate the restoration problem as the minimization of the sum of a data fidelity term, favoring restored images statistically close to the speckled one, and a regularization term that encourages the solution to be smooth. To prevent the apparition of a blur, edge-preserving terms such as the total-variation have been widely applied [7]–[10]. To preserve both point-like structures and smooth areas with sharp boundaries, image decomposition models were introduced [11], [12]. Wavelets [13] and curvelets [14] were also applied in several works.

The large success of patch-based methods in image denoising [15], [16] fueled a large body of research [17], [18], from intensity-image restoration [19]–[21] to interferometric [22], polarimetric [23], [24] or polarimetric and interferometric [25] modalities.

There has been a revival of interest in the restoration of SAR intensity images with the advent of deep learning. The most distinctive feature of deep despeckling networks is the training approach [4]: (i) the first generation of methods used supervised training techniques where pairs of speckled/speckle-free images were formed to train the network; (ii) then, self-supervised techniques used pairs of images of the same area, acquired at different times; (iii) self-supervised training with a single image represents the ultimate goal.

A fundamental limitation of supervised training (i) is the difficulty to obtain the speckle-free image associated with a speckled image. Most techniques from this category, therefore, rely on synthetic speckle, i.e., the simulation of speckle corruption, starting from an optical image [26]–[28] or from the temporal mean of a long time series of SAR images [29], [30]. In the speckle simulations, however, spatial correlations are typically ignored, which requires that actual SAR images be pre-processed before applying the network (by resampling and inversion of the SAR transfer function [31], [32], or down-sampling [33]). A way to circumvent this problem is to use actual SAR images as input and a temporal average as reference [34], [35]. Any change that occurs during the time series is a potential source for bias. Temporally stable areas must then be selected, which is challenging, in particular with cultivated lands, since a trade-off between sufficient speckle reduction (hence, large temporal series) and limited changes must be found.

Self-supervised training (ii) uses pairs formed by two images of the same area, acquired at separate dates so that the speckle is temporally decorrelated. That way, the second image, though noisy, can be used as a reference for the network. For this to work in practice, changes between the two images must be compensated [36], which requires in turn resorting to a despeckling technique. Compared to supervised

E. Dalsasso and F. Tupin are with LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France, e-mail: forename.name@telecom-paris.fr.

L. Denis is the Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France, e-mail: loic.denis@univ-st-etienne.fr.

techniques, this approach benefits from training with actual SAR images and it is thus robust to spatial correlations of speckle due to the SAR transfer function.

Finally, self-supervised training based on a single-image (iii) uses networks with a specific architecture [37], [38] so that the receptive field does not include the central area, forming a blind-spot [39] that can be used to supervise the estimation by the network. The very specific receptive field of such networks strongly constrains their architecture, which can limit their performance. Moreover, for the self-supervision to succeed, the speckle must be spatially decorrelated, which requires the same pre-processing techniques as described previously. Several variants of the concepts of blind-spot have been recently developed in the literature of image denoising, using various masking strategies [40], [41]. The spatial correlation of the speckle represents, however, a severe limitation to the potential application of these approaches to SAR imaging.

Our contributions: This paper introduces a new training strategy, applicable to all kinds of network architectures. This strategy is fully unsupervised: it only requires single-look complex (SLC) images to perform the training or to process new data once the network is trained. In contrast to other existing works, it does not require additional hypotheses like the absence of spatial correlations of the speckle, or temporal stability throughout a time series. The phase information of single-look complex images is often considered irrelevant when only the intensity is of interest (i.e., apart from the context of SAR interferometry). Finding an interest in the real and imaginary components for the restoration of intensity images may even seem disconcerting. In section II, we derive a statistical model of SLC images showing that two independent and identically distributed images can be extracted from an SLC image. This paves the way to the application of a self-supervised training strategy inspired by noise2noise [42]: MERLIN, see section III. Results obtained in section IV on images at different spatial resolutions confirm the ability of MERLIN to produce high-quality restoration results at medium, high, and very high spatial resolutions.

II. Statistical model of SAR images

Goodman's fully developed speckle model [43] gives the statistical distribution followed by the complex amplitude $z = a\exp(j\varphi)$ resulting from the coherent summation of many elementary echoes $a_n \exp(j\varphi_n)$. Under the hypothesis of a large surface roughness compared to the radar wavelength, each elementary echo has a phase that is independent and uniformly distributed in the range $[-\pi, \pi]$. If the area in the radar resolution cell is homogeneous, elementary amplitudes $a_n > 0$ are also independent and identically distributed (i.i.d.) as well as independent from the phases φ_n. It follows from the central limit theorem, in the limit of a large number of elementary echoes, that the distribution of $z = \sum_n a_n \exp(j\varphi_n)$ converges to a circular complex Gaussian distribution [43]:

$$\mathrm{p}_Z(z) = \frac{1}{\pi r}\exp(-|z|^2/r)\,, \tag{1}$$

where z is the complex amplitude at a given pixel and $r > 0$ is the SAR reflectivity at that pixel.

The multiplicative nature of speckle phenomenon becomes clear by writing z under the form $z = s\sqrt{r}$, with r the reflectivity of the homogeneous area and s a complex random variable distributed according to:

$$\mathrm{p}_S(s) = \frac{1}{\pi}\exp(-|s|^2)\,. \tag{2}$$

The decomposition of z into its real and imaginary parts, $z = a + jb$, leads to:

$$\begin{aligned}\mathrm{p}_Z(z) = \mathrm{p}_Z(a+jb) &= \frac{1}{\pi r}\exp(-(a^2+b^2)/r) \\ &= \underbrace{\frac{1}{\sqrt{2\pi}\sqrt{r/2}}\exp(-a^2/r)}_{\mathcal{N}(0,r/2)}\underbrace{\frac{1}{\sqrt{2\pi}\sqrt{r/2}}\exp(-b^2/r)}_{\mathcal{N}(0,r/2)}\,,\end{aligned} \tag{3}$$

which shows that the real and imaginary parts of the complex amplitude are i.i.d. according to a Gaussian distribution with variance $r/2$, or equivalently, that the real and imaginary parts of the complex-speckle component s in the multiplicative model are i.i.d. and Gaussian-distributed with a variance equal to 1/2.

This multiplicative stochastic model is illustrated in the left part of figure 1. From one pixel to the next, the realization of speckle is different and the random field $\boldsymbol{s} \in \mathbb{C}^K$ of a K-pixels image is a white Gaussian field.

Depending on the acquisition mode, the chosen pixel size, and the spectral apodization applied to reduce sidelobes around bright targets, a specific SAR transfer function then transforms the spatially uncorrelated field $\boldsymbol{z}$ into a spatially correlated field $\tilde{\boldsymbol{z}}$, see Fig.1 (center):

$$\tilde{\boldsymbol{z}} = \mathbf{H}\boldsymbol{z}\,, \tag{4}$$

with $\mathbf{H}$ the spatial-domain operator associated to the SAR transfer function. By linearity of $\mathbf{H}$, we get:

$$\tilde{\boldsymbol{a}} = \mathbf{H}\boldsymbol{a} \quad \text{and} \quad \tilde{\boldsymbol{b}} = \mathbf{H}\boldsymbol{b}\,, \tag{5}$$

which shows that $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{b}}$ are spatially correlated but *mutually independent* random fields. The element-wise multiplication by $\sqrt{\boldsymbol{r}}$ and the linear operation $\mathbf{H}$ transform the white Gaussian fields corresponding to the real and imaginary parts of $\boldsymbol{s}$, distributed according to $\mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{I})$, into two i.i.d. Gaussian fields $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{b}}$ distributed according to $\mathcal{N}(\mathbf{0}, \frac{1}{2}\mathbf{H}\mathrm{diag}(\boldsymbol{r})\mathbf{H}^\top)$, with $\mathrm{diag}(\boldsymbol{r})$ the $K \times K$ diagonal matrix whose diagonal is equal to the vector $\boldsymbol{r} \in \mathbb{R}^K_{+*}$.

Finally, the intensity image in SAR imaging is obtained by summing the squared real and imaginary parts, see the bottom right of Fig.1. The square is applied separately to the real and imaginary components. Random fields $\tilde{\boldsymbol{a}}^2$ and $\tilde{\boldsymbol{b}}^2$ are thus still i.i.d.

In summary, as depicted in Fig.1, the statistical model of SAR image formation shows that the two components $\tilde{\boldsymbol{a}}^2$ and $\tilde{\boldsymbol{b}}^2$, corresponding to the squared real and imaginary part of an SLC image that are added to form the SAR intensity image, are *independent and identically distributed*. Each component contains half of the information, or, in other words, has a signal-to-noise ratio (SNR) that is $1/\sqrt{2}$ times the SNR of the

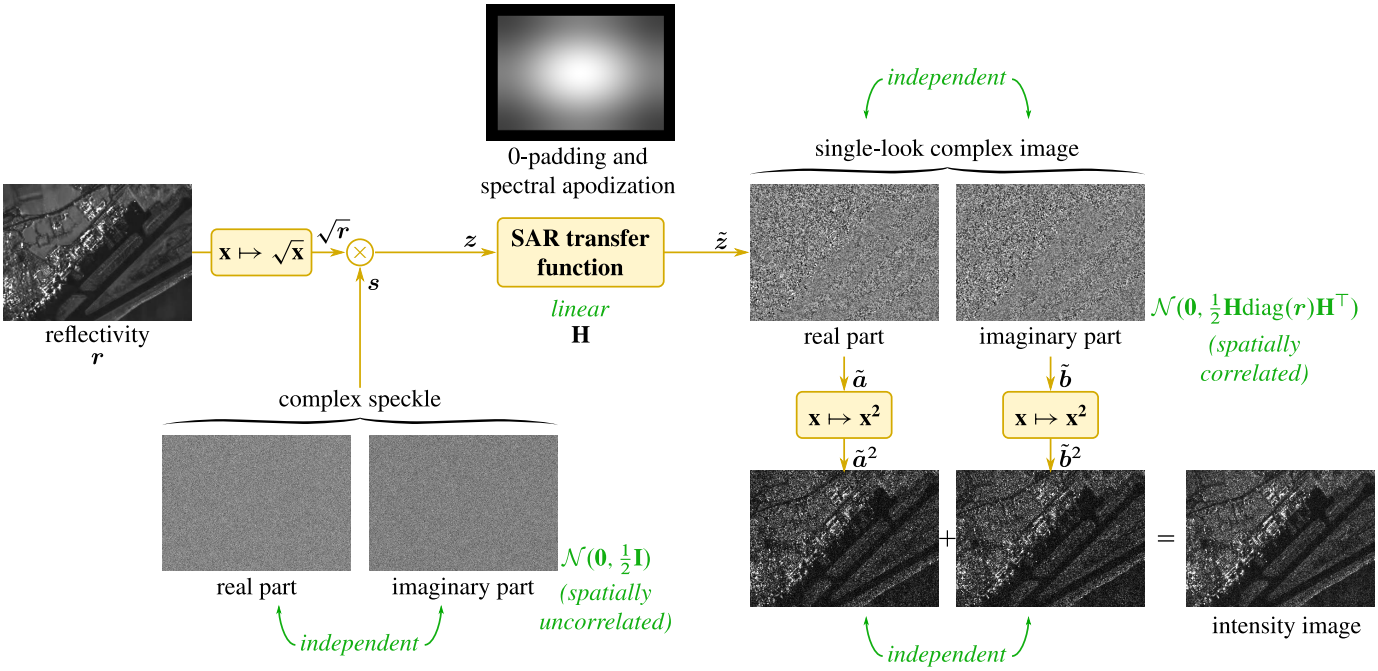

Figure 1. A statistical model of speckle in SAR image: the intensity image on the right is a corrupted version of the reflectivity image shown on the left. The single-look complex image contains spatially-correlated speckle components that are independent in the real and imaginary parts. The SAR transfer function shown here corresponds to Sentinel-1 stripmap mode. For visualization purposes, a non-linear look-up table is used to display intensity images.

intensity image (the intensity corresponds to the sum of these two independent components, its variance is halved, which corresponds to a SNR improvement by a factor $\sqrt{2}$).

III. MERLIN: COMPLEX SELF-SUPERVISED DESPECKLING

Since SLC images provide two i.i.d. components that contain half the information[1] from the intensity image, a self-supervised training strategy can be built by processing one component (e.g., the real part) and evaluating the restoration quality on the other component (e.g., the imaginary part). This corresponds to an ideal application case of the noise2noise principle [42] in which a deep neural network is trained to predict a noisy image from another independent noisy realization. Since the realization-specific random perturbation can not be guessed by the network, it tends to remove the noise in the input image even if no noiseless image is provided to the loss function.

We follow a similar approach to train networks by performing a coMplex sElf-supeRvised despeckLINg (MERLIN). Our approach is graphically summarized in figure 2: during the training phase (step A of the figure), the network is trained to process only the real part and a loss function is evaluated to measure how close the estimated reflectivity is to the imaginary component. Once the network is trained, it can be applied to reduce speckle noise in SLC images (step B of the figure). This time, both the real and imaginary parts are independently processed using the same network weights (i.e., a single network is trained in step A). The two estimations are then combined to produce the final estimation (by averaging). When despeckling SLC images, all the information is used, i.e., both the real and imaginary parts.

In order to define the loss function used during the training phase (step A), it is necessary to decide which parameters should be estimated. The SAR transfer function has a significant impact on the image appearance: the 0-padding controls the pixel size while the spectral apodization sets the height of the sidelobes. Rather than inverting the SAR transfer function, we consider producing an image with the same characteristics (pixel size and bright point signature). With an ideal SAR transfer function $\mathbf{H} = \mathbf{I}$ (the identity matrix of dimension $K \times K$), the real and imaginary parts $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{b}}$ have a variance equal to $\boldsymbol{r}/2$. With a non-ideal transfer function, the variance corresponds to the diagonal of matrix $\frac{1}{2}\mathbf{H}\mathrm{diag}(\boldsymbol{r})\mathbf{H}^\top$, i.e., the variance of the k-th pixel is $\tilde{r}_k/2$ with $\tilde{r}_k = \sum_\ell \mathrm{H}_{k\ell}^2 r_\ell$. With MERLIN, we aim at estimating the values $\tilde{r}_k$ for each pixel. The marginal distribution of $\tilde{a}_k$ and $\tilde{b}_k$ (the real and imaginary parts at pixel k of the SLC image) is a centered Gaussian with variance $\tilde{r}_k/2$. We thus define the following loss function $\mathcal{L}$:

$$\mathcal{L}(\tilde{\boldsymbol{r}}, \tilde{\boldsymbol{b}}) = \sum_k \frac{1}{2} \log\left(\tilde{r}_k\right) + \frac{\tilde{b}_k^2}{\tilde{r}_k}, \tag{6}$$

which corresponds, up to an additive constant, to the sum over all pixels of the opposite of the log-likelihood of the marginal distribution. To reduce the dynamic range, it is beneficial that the network inputs and outputs be expressed in log-scale. We introduce $\check{r}_k = \log \tilde{r}_k$, $\check{a}_k = \log|\tilde{a}_k|$, and $\check{b}_k = \log|\tilde{b}_k|$ to define the equivalent loss function expressed with log-scale

[1] we consider here that the raw phase angle($\boldsymbol{z}$) is non-informative, which of course is no longer true when considering multiple SLC images in interferometric configuration

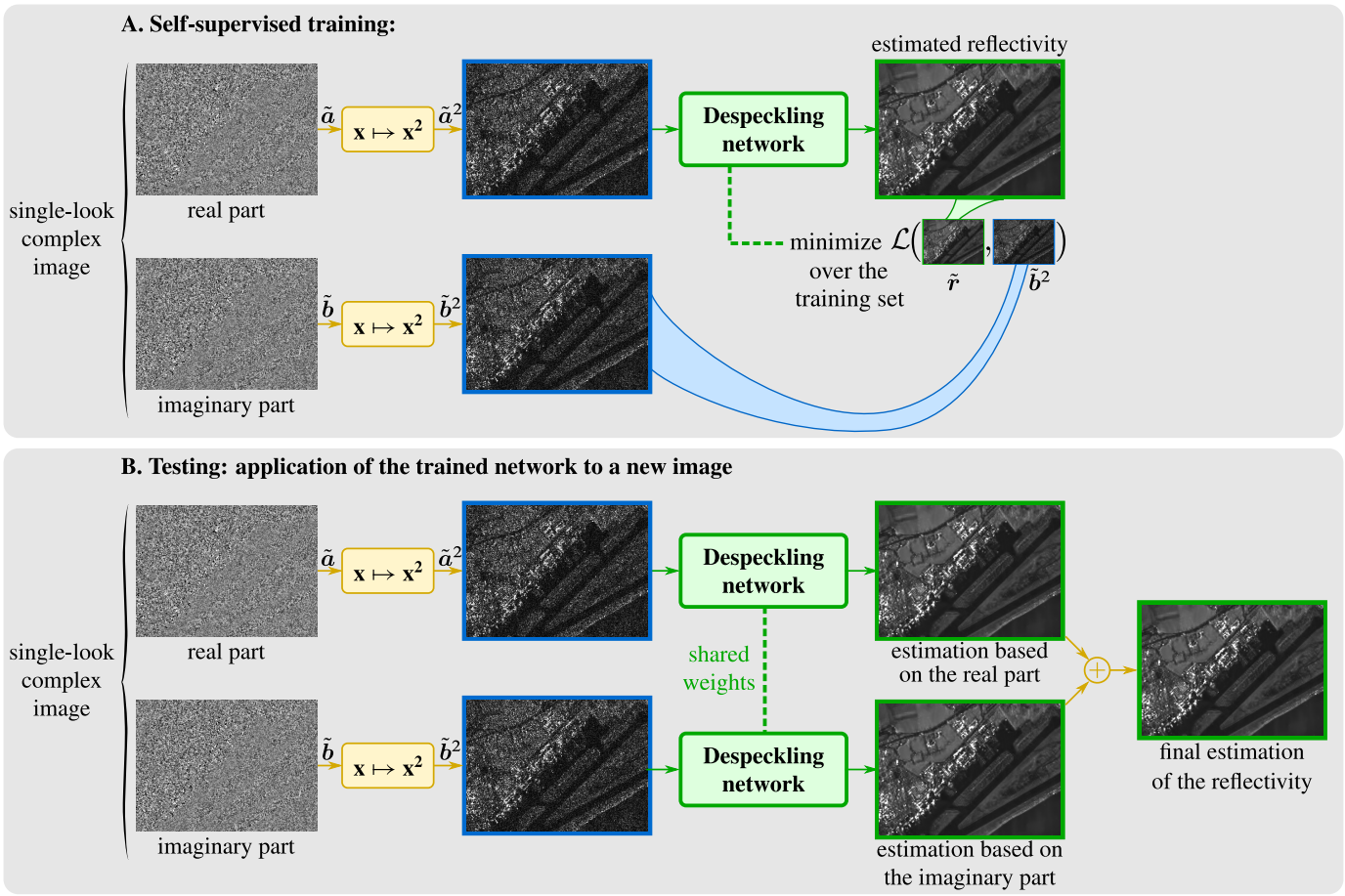

Figure 2. The principle of MERLIN: during step A, the despeckling network is trained to estimate the reflectivity based solely on the real part. The loss function evaluates the likelihood of the predictions according to the imaginary part. Once the network is trained, it can be used as shown in B: the real and imaginary parts are processed separately using networks with the same weights. The outputs are combined to form the final estimation. Note that, to simplify the figure, step A is illustrated only with the real part as input but real and imaginary parts can be swapped during training to increase the number of training samples, see text.

images:

$$\mathcal{L}_{\log}(\check{\boldsymbol{r}}, \check{\boldsymbol{b}}) = \sum_k \frac{1}{2}\check{r}_k + \exp\left(2\check{b}_k - \check{r}_k\right) . \tag{7}$$

Note that since the real and imaginary parts are i.i.d. and given that we use the same network weights in step B to process both the real and imaginary parts, the training phase (step A) can not only be performed with the real part as input and the loss $\mathcal{L}(\tilde{\boldsymbol{r}}, \tilde{\boldsymbol{b}})$, but also with the imaginary part as input and the loss $\mathcal{L}(\tilde{\boldsymbol{r}}, \tilde{\boldsymbol{a}})$ (only the former case is represented in Fig.2 for simplicity reasons while both are applied in practice).

IV. Experimental validation

In contrast to the family of self-supervised methods derived from the concept of blind-spot [39] that require the receptive field of the network to exclude the central pixel(s), MERLIN imposes no constraint on the type of neural network used to perform the estimation. In the following experiments, we used a simple U-Net architecture [44]. This network, originally developed for semantic segmentation, performs very well on image denoising tasks and can be trained quickly [29], [36], [42].

To limit the dynamic range of the images at the input of the network, images $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{b}}$ are log-transformed and normalized using a fixed affine transform.

In the following set of images, a residual U-Net (based on the network described in [42]) has been trained on SLC SAR images. The experiments have been conducted both on images with synthetic speckle noise and on real SAR images. The hyper-parameters used to train the network for each imaging modality are listed in table I. The weights of the trained models are made available for testing at https://gitlab.telecom-paris.fr/RING/MERLIN.

A. Evaluation of MERLIN on images with synthetic speckle

We first evaluate the capability of MERLIN to train a network to restore images synthetically corrupted by speckle. The training set of speckle-free images has been built according to [30]. We consider an ideal SAR transfer function: $\mathbf{H} = \mathbf{I}$ so that many different despeckling techniques can be applied and compared. Figure 3 compares restoration results on 3 different images with simulated speckle, for the same network architecture but two different training strategies: (i) a supervised training where the network has access to the full

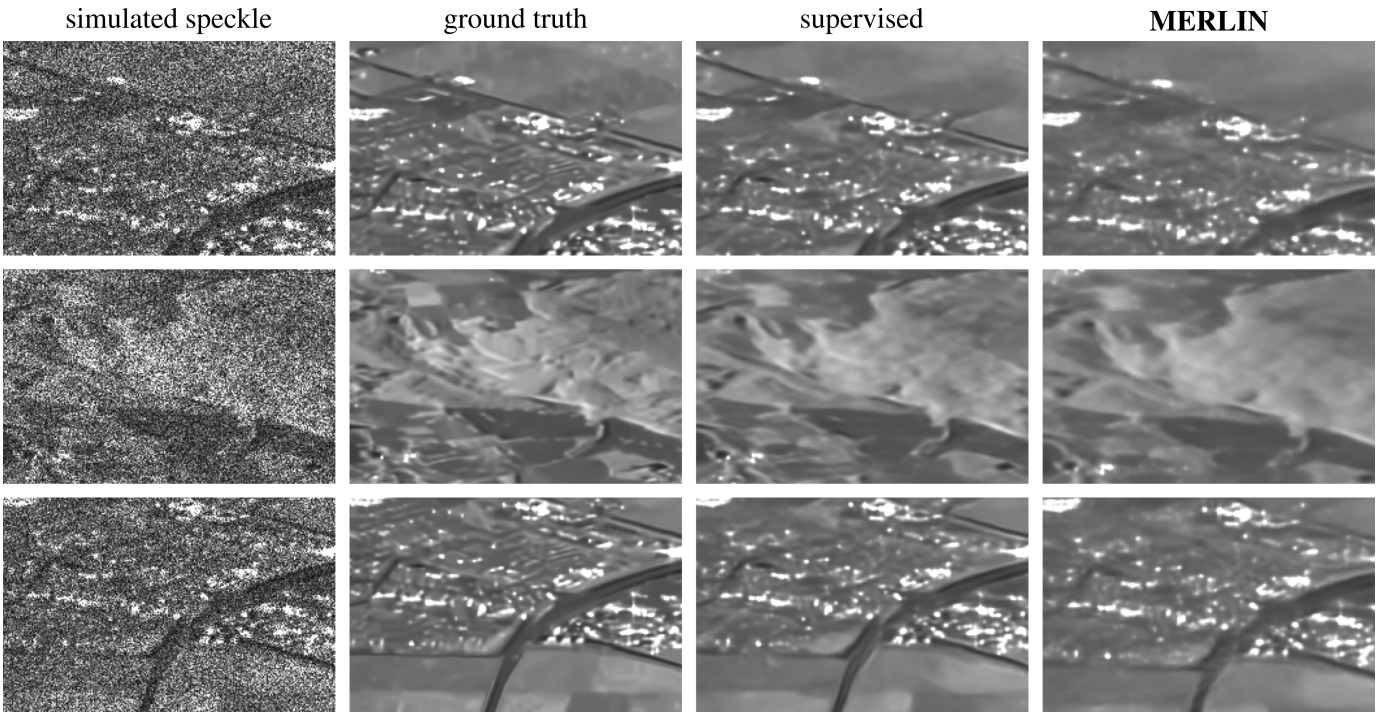

Figure 3. Despeckling results on images corrupted by synthetic speckle: the first column shows the noisy images obtained by multiplying the ground truth images $\sqrt{\boldsymbol{r}}$ shown in the second column by a white speckle field. The third column gives despeckling results obtained with a U-Net trained in a supervised fashion (SAR2SAR, step A [36]). The last column gives despeckling results obtained with the same network trained with MERLIN. Compared to the supervised training, MERLIN is penalized because it only has access to a single speckle realization. The images shown in each row are regions of interest extracted from the images Lely, Limagne, and Marais1 that appear in table II. Additional images can be seen on https://gitlab.telecom-paris.fr/RING/MERLIN.

Table I
DESCRIPTION OF THE TRAINING PARAMETERS FOR ALL EXPERIMENTS CARRIED OUT WITH A RESIDUAL U-NET TRAINED WITH MERLIN.

	Synthetic speckle	**TerraSAR-X stripmap**	**TerraSAR-X spotlight**	**SETHI**
# images	7	3	4	1
patch size	256×256	256×256	256×256	256×256
batch size	12	12	12	12
# batches	1035	4217	2254	6355
# epochs	30	30	30	30
gradient norm [45]	1.0	1.0	0.5	1.0
learning rate	10^{-2}	10^{-2}	10^{-2}	10^{-2}
	10^{-3} after 6 epochs	10^{-3} after 4 epochs	10^{-3} after 4 epochs	10^{-3} after 4 epochs
	10^{-4} after 20 epochs	10^{-4} after 20 epochs	10^{-4} after 20 epochs	10^{-4} after 20 epochs

intensity image (i.e., $\tilde{\boldsymbol{a}}^2+\tilde{\boldsymbol{b}}^2$) and the loss function is evaluated on a different, independent, speckle realization drawn from the same ground-truth image (i.e., step A of SAR2SAR algorithm [36]); (ii) a self-supervised training with MERLIN where the network only has access to either $\tilde{\boldsymbol{a}}^2$ or $\tilde{\boldsymbol{b}}^2$ and the loss function is $\mathcal{L}(\check{\boldsymbol{r}}, \check{\boldsymbol{b}})$ or $\mathcal{L}(\check{\boldsymbol{r}}, \check{\boldsymbol{a}})$, respectively. Note that in the approach (i) the loss function used during training corresponds to:

$$\mathcal{L}_{\log}(\check{\boldsymbol{r}}, \check{\boldsymbol{a}}') + \mathcal{L}_{\log}(\check{\boldsymbol{r}}, \check{\boldsymbol{b}}') = \sum_k \check{r}_k + \exp\left(\check{i}'_k - \check{r}_k\right) , \qquad (8)$$

with $2\check{\boldsymbol{a}}' = \log \boldsymbol{a}'^2$, $2\check{\boldsymbol{b}}' = \log \boldsymbol{b}'^2$, and $\check{\boldsymbol{i}}' = \log(\boldsymbol{a}'^2 + \boldsymbol{b}'^2)$ the log-transformed versions of the square of the real and imaginary parts, and the intensity of the *second* noisy realization. Obviously, with half the information available, the same network trained with MERLIN can not perform as well as when trained with the supervised training (i). The degradation in image quality remains limited, however, as seen on figure 3.

Table II gives PSNR values, expressed on amplitude images $\sqrt{\boldsymbol{r}}$, for several despeckling methods. Depending on the image, the U-Net trained with MERLIN performs at least as well or better than methods like SAR-BM3D [20] or NL-SAR [46] that are not based on deep neural networks. The performance seems comparable on average to that of SAR-CNN [34] (this network has been retrained on the same simulated speckle noise images as MERLIN or SAR2SAR_A, providing the ground truth image to compute the loss function). Numerical values confirm our analysis of figure 3: when trained with MERLIN, the U-Net produces results that are slightly worse than when real and imaginary parts are processed jointly and the network is trained in a supervised fashion. Compared to the self-supervised method Speckle2Void [38] which uses a

Table II
COMPARISON OF DENOISING QUALITY IN TERMS OF PSNR ON AMPLITUDE IMAGES. FOR EACH GROUND TRUTH IMAGE, 20 NOISY INSTANCES ARE GENERATED. 1σ CONFIDENCE INTERVALS ARE GIVEN. PER-METHOD AVERAGES ARE INDICATED AT THE BOTTOM.

Images	Noisy	SAR-BM3D (patch-based)	NL-SAR (patch-based)	MuLoG+BM3D (patch-based)	SAR-CNN (deep network) (supervised)	SAR2SAR_A (deep network) (supervised)	Speckle2Void (deep network) (self-supervised)	**MERLIN** (deep network) (self-supervised)
Marais 1	10.05±0.014	23.56±0.134	21.71±0.126	23.46±0.079	24.65±0.086	**25.73**±0.125	24.89±0.102	25.25±0.113
Limagne	10.87±0.047	21.47±0.309	20.25±0.196	21.47±0.218	22.65±0.291	**24.45**±0.119	23.40±0.121	23.86±0.111
Saclay	15.57±0.134	21.49±0.368	20.40±0.270	21.67±0.244	23.47±0.228	**23.60**±0.437	19.00±0.481	22.34±0.484
Lely	11.45±0.005	21.66±0.445	20.54±0.330	22.25±0.437	**23.79**±0.491	23.67±0.542	19.28±0.575	22.85±0.467
Rambouillet	8.81±0.069	23.78±0.198	22.28±0.113	23.88±0.169	**24.73**±0.080	24.16±0.385	21.47±0.488	23.53±0.316
Risoul	17.59±0.036	29.98±0.264	28.69±0.201	30.99±0.376	**31.69**±0.283	30.68±0.230	21.42±0.119	29.80±0.138
Marais 2	9.70±0.093	20.31±0.783	20.07±0.755	21.59±0.757	23.36±0.807	**26.63**±0.215	25.04±0.222	26.10±0.193
Average	12.00	23.17	21.99	23.62	24.91	**25.56**	22.07	25.13

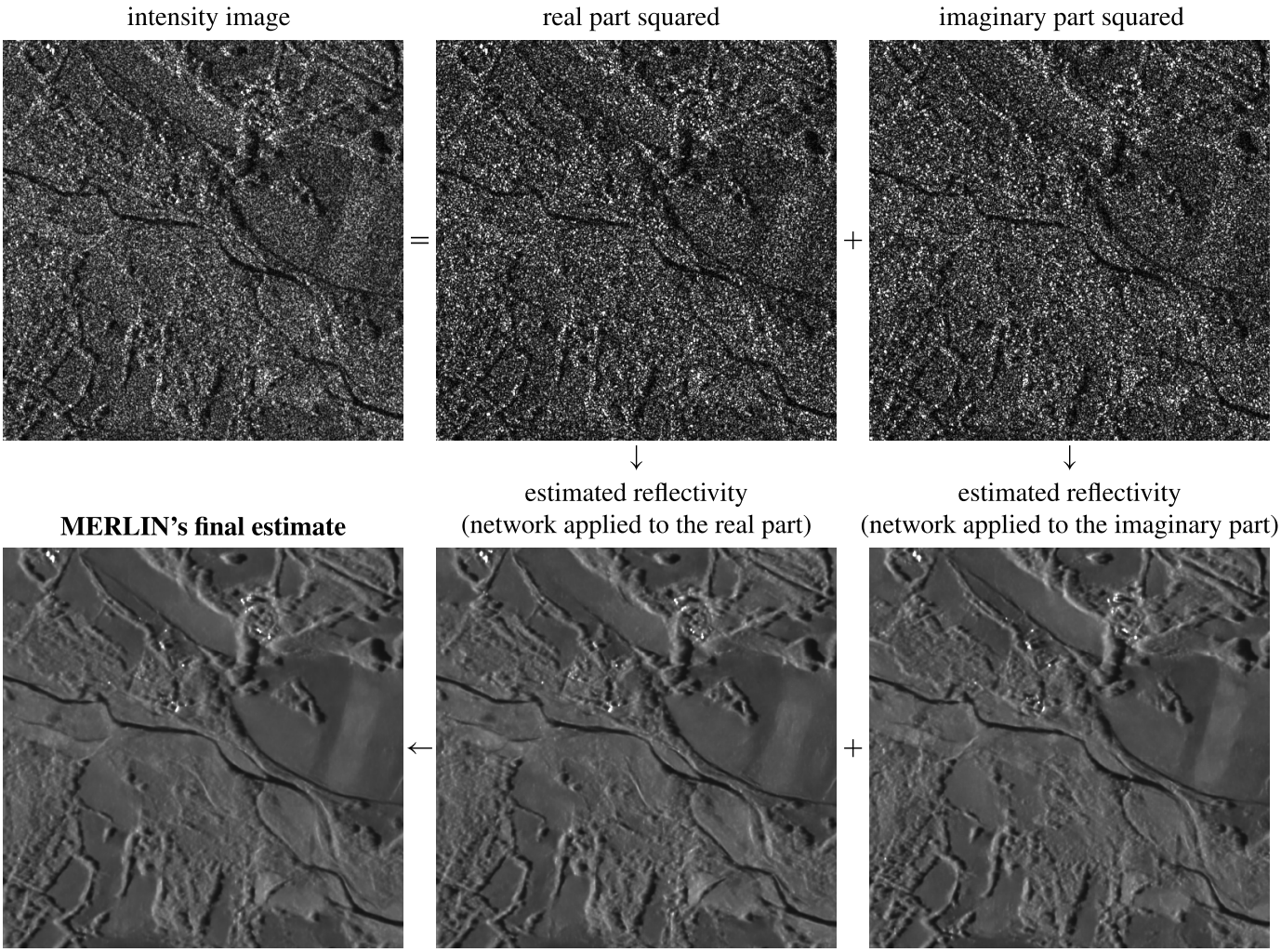

Figure 4. Application of a U-Net trained with MERLIN on a TerraSAR-X image near Serre-Ponçon dam, in the French Alps, acquired in stripmap mode. Additional despeckling results on TerraSAR-X images in stripmap mode can be seen on https://gitlab.telecom-paris.fr/RING/MERLIN.

specific network architecture to obtain a receptive field with a central blind spot, the performance of the U-Net network trained with MERLIN is notably better. We show in the next paragraph that, when applied to actual SAR images, the gain brought by self-supervision with MERLIN becomes very appealing.

B. Restoration of actual SAR images

When actual SAR images are considered, it is beneficial to train a network specifically for a given sensor and a particular imaging mode. The SAR transfer function varies from one imaging mode to the other, as well as the spatial resolution and, thus, the structures that can be resolved. In this paragraph, we show results obtained with the same network architecture

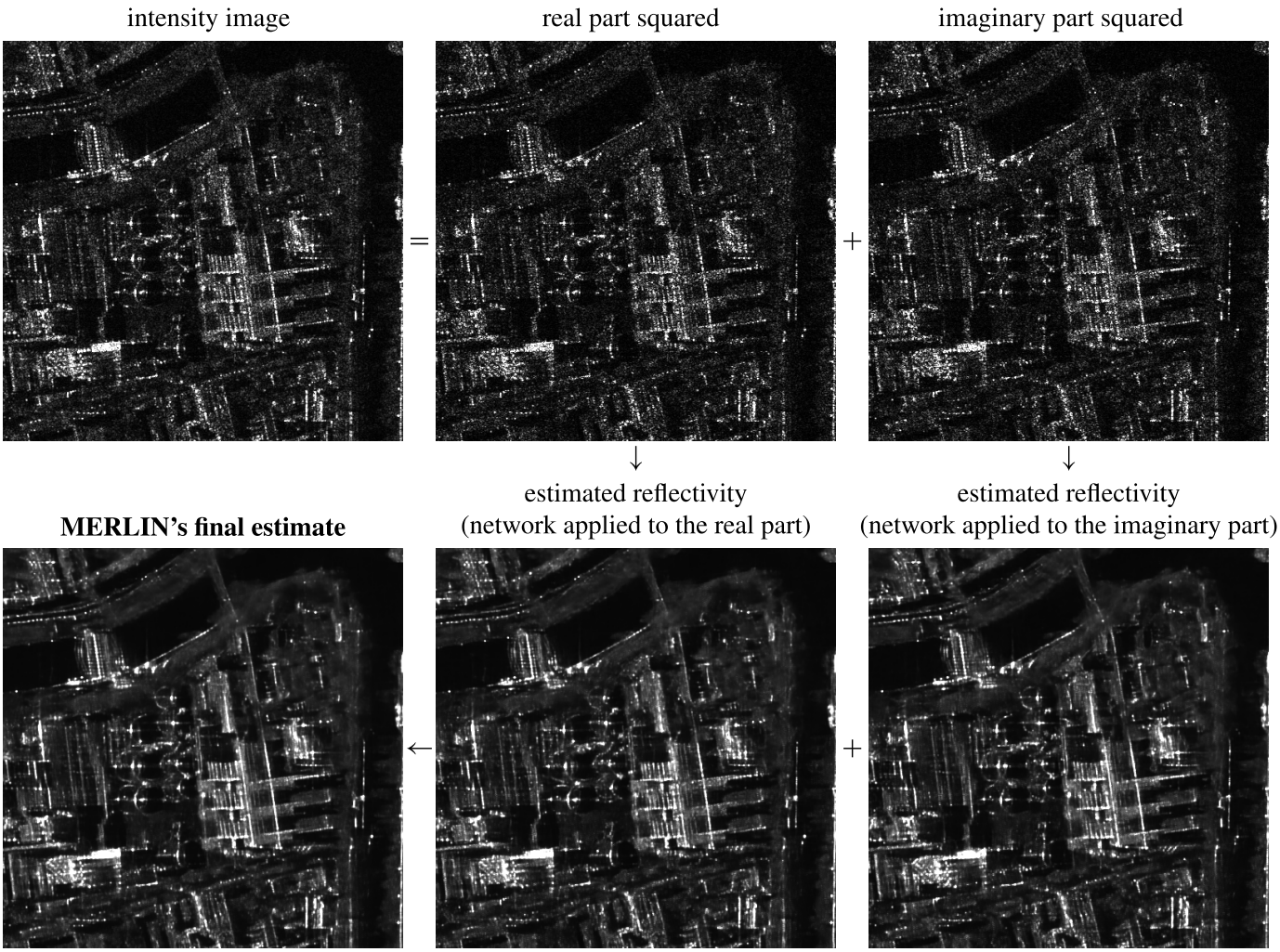

Figure 5. Application of a U-Net trained with MERLIN on a TerraSAR-X image of Berlin, Germany, acquired in spotlight mode. Additional despeckling results on TerraSAR-X images in SpotLight mode can be seen on https://gitlab.telecom-paris.fr/RING/MERLIN.

but different trainings each performed on images of the same type.

Figure 4 shows results obtained on a TerraSAR-X image acquired in stripmap mode over an agricultural area in the French Alps. The first row of the figure illustrates the decomposition of the intensity image into its squared real $\tilde{\boldsymbol{a}}^2$ and imaginary $\tilde{\boldsymbol{b}}^2$ components. The second row shows the estimations produced by the network trained by MERLIN from each component and the final estimate. A qualitative analysis of the results shows a very good restoration of fine details and textures as well as bright targets.

Figure 5 gives results obtained on a TerraSAR-X image acquired in SpotLight mode over an urban area: a small area of the city of Berlin, Germany. The content in this area is very different from the previous region shown in Fig.4: there are many bright targets and the images have a higher spatial resolution. Bright targets are preserved while homogeneous areas are smoothed. When many point-like targets are aligned horizontally or vertically (i.e., in the direction of the sidelobes), the network tends to merge the targets into a line. This phenomenon could possibly be reduced by considering a larger training set and/or a different network architecture.

Figure 6 shows how MERLIN performs in very-high-resolution airborne imaging. The same U-Net network as previously is trained *on a single image* ($9\,130 \times 10\,000$ pixels) captured with SETHI [47] by the French aerospace laboratory ONERA in 2014. The image has a pixel size of 13cm in range and 19cm in azimuth (the spatial resolution is about 35cm). Two regions of interest are displayed together with optical images at 20cm resolution (orthorectified image by the French geographic institute, IGN). Processing such an image is challenging for a despeckling algorithm because of the spatial correlations of speckle and the strong sidelobes around bright targets. Vegetation seems to be well restored: the stripes visible both in optical and SAR images are preserved and the low-contrasted tree response in the bottom row of Fig.6 is recovered. Few distortions seem to be applied to bright targets in the top row of Fig.6.

Figure 7 compares the results produced by MERLIN with two other despeckling filters: SAR-BM3D [21] and Speckle2Void [38]. For each restoration result, the residual image (ratio between the noisy and the denoised image) is also shown. To account for speckle spatial correlations,

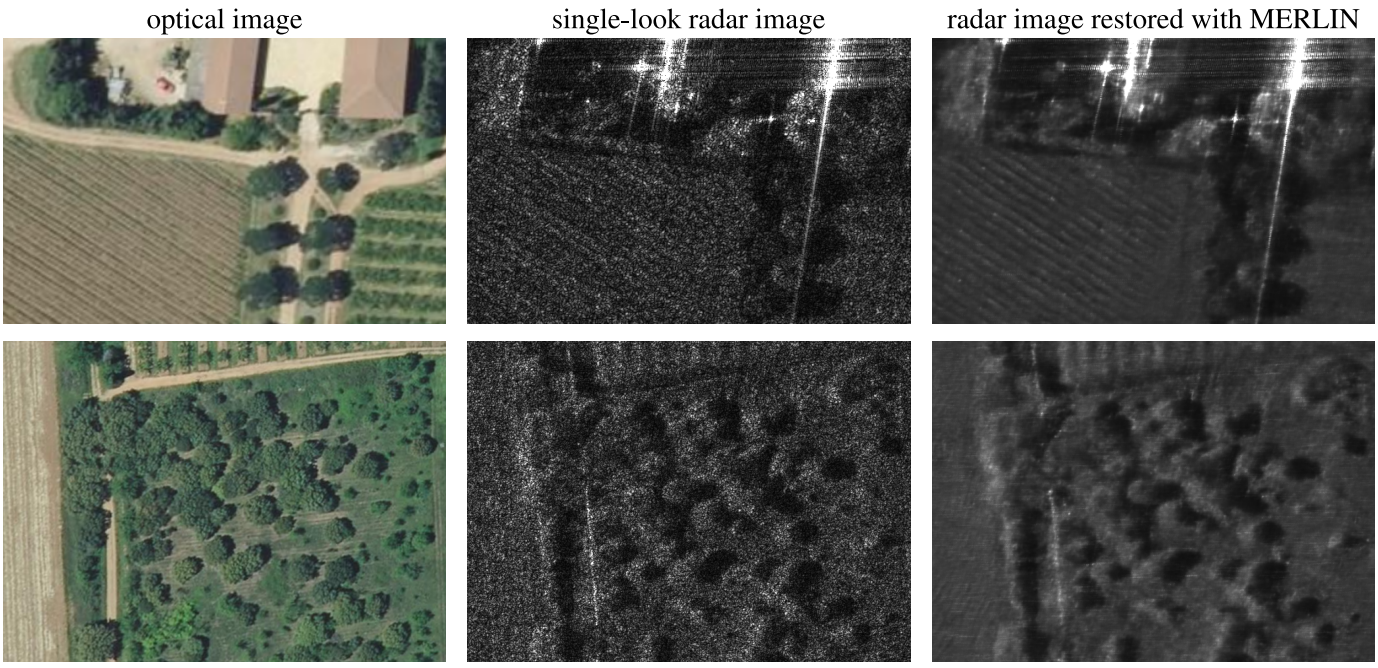

Figure 6. Application of a U-Net trained with MERLIN on a 35cm spatial resolution airborne image on an agricultural area near Nîmes, France, acquired in 2014 by SETHI sensor (©ONERA). The corresponding 20cm resolution optical images (source: Geoportail ©IGN), shown on the left, date back from 2018. Additional despeckling results on airborne radar images can be seen on https://gitlab.telecom-paris.fr/RING/MERLIN.

before applying SAR-BM3D and Speckle2Void, images have been decorrelated. The blind speckle decorrelator proposed by Lapini *et al.* [31] has been used, as suggested by the authors of Speckle2Void [38]. Visual inspection of the residual images suggests that some structures have been attenuated by SAR-BM3D, with the edges appearing a bit fuzzy. Moreover, some artifacts arise in homogeneous areas.

The blind-spot structure of the network employed in Speckle2Void makes it hard to recover isolated bright points, as it is difficult to predict their existence from the neighboring pixels. Textures are well recovered but a slight bias can be observed in the lake of Fig.7.b. Almost no structures can be identified in the residual images of MERLIN. Being robust to speckle spatial correlations and relying on all the pixels in the receptive field of the CNN, MERLIN produces a pleasant result with a good preservation of both geometrical structures and detailed textures, whilst strongly suppressing speckle noise.

To allow a more extensive evaluations of MERLIN, additional restoration results are provided at https://gitlab.telecom-paris.fr/RING/MERLIN.

V. Discussion

The statistical model for SAR image formation presented in section II and figure 1 served as a basis to derive the loss function used to train networks with MERLIN, Eqs.(6) and (7). It is based on Goodman's speckle model which is well-established for homogeneous areas imaged at a medium to high resolution. It is known to be less relevant when very high-resolution images are considered, especially in urban areas due to the presence of strong scatterers that dominate the responses within a resolution cell. Many alternative statistical models were proposed in the literature [48] [49] [50] [51]. A major difficulty to incorporate such models within MERLIN's loss function is that they depend on additional parameters that would require to be locally set to account for the content of each resolution cell (level of heterogeneity in the cell). From a pragmatic point of view, the qualitative analysis of the results produced by MERLIN on high-resolution images (Fig.5) seems to indicate that the behavior of the network is satisfactory even in the very-high-resolution regime. Performing a more in-depth analysis would probably require using high-quality SAR simulators to provide ground truths for quantitative validation.

Note that, due to the phase modulation applied to perform Terrain Observation with Progressive Scans SAR (TOPSAR) in most of Sentinel-1 acquisition modes (in particular, Interferometric Wide swath, IW), a direct application of MERLIN is *not possible*. It is mandatory that these SLC images be deramped before processing, see [52].

One may wonder if MERLIN could possibly work on intensity-only images, by generating fake phase information (a phase could be drawn at each pixel according to a uniform distribution in $[-\pi, \pi]$). This would work perfectly well in the case of an ideal SAR transfer function: the results presented in section IV-A would not change if only the intensity was provided to MERLIN and a random phase was generated afterward. When real images are considered, the SAR transfer function is no longer ideal and the statistical distribution of the actual phase differs from a random white field. Figure 8.c shows the restored images obtained by a network trained on TerraSAR-X Stripmap intensity images, with a phase generated randomly. There are remaining speckle fluctuations and artifacts in the form of a high-frequency texture due to the

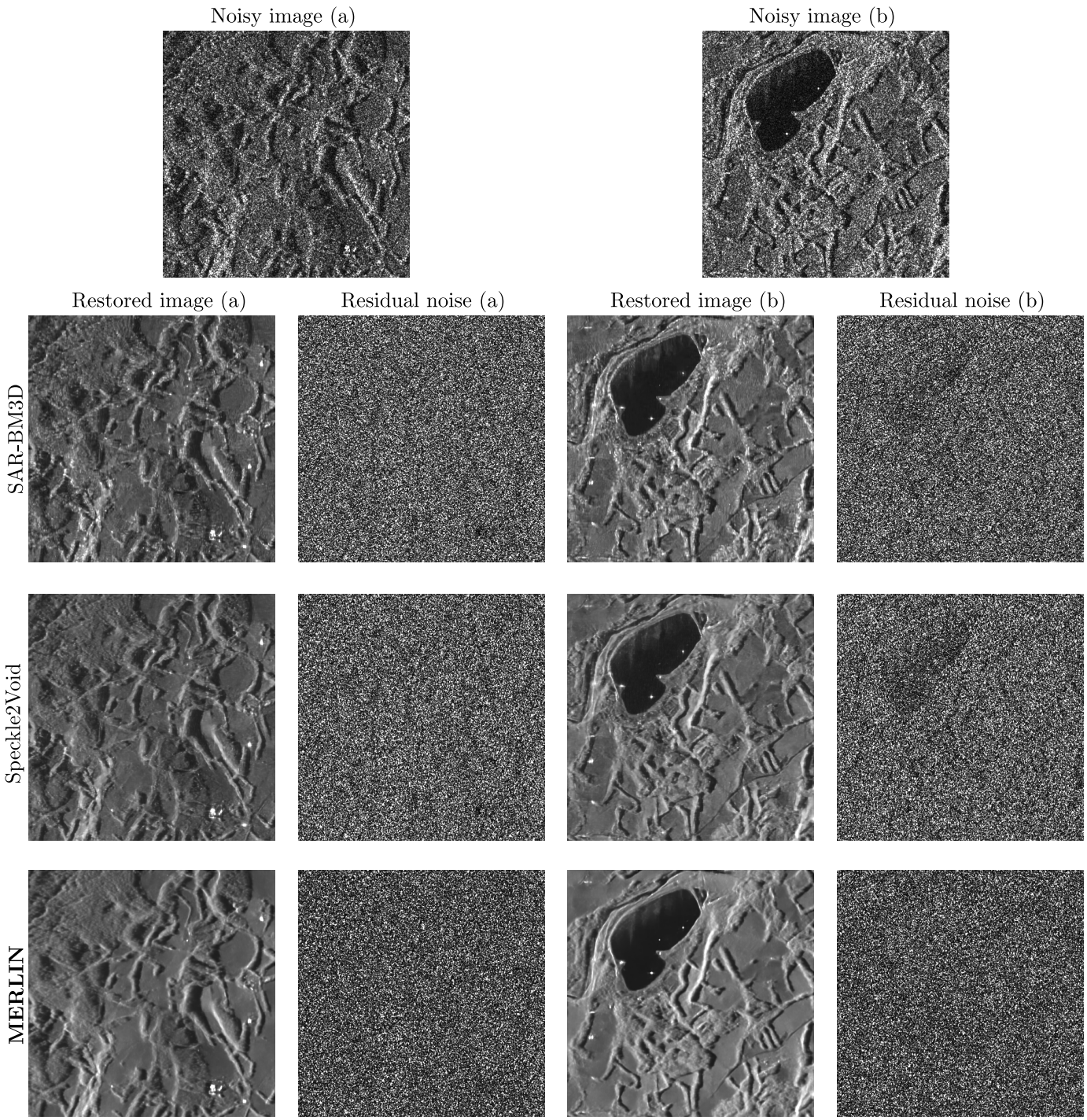

Figure 7. Restoration results of three despeckling filters on two TerraSAR-X images near Serre-Ponçon dam, in the French Alps, acquired in stripmap mode. The residual intensity images (*i.e.* the ratio noisy/denoised) is provided to assist in the visual analysis.

mismatch between the spatial correlations of the intensity and the phase. Those are absent from the results produced when applying MERLIN on the SLC image (Fig.8.b). A possible way to circumvent this problem would consist in subsampling the images to decrease the speckle spatial correlations, both at training and at testing time, and draw random phases. This leads to images free of correlation artifacts, see Fig.8.d, at the cost of a degradation of the highest-frequency details (such as thin lines). In conclusion, it is preferable to use SLC images when available to directly apply MERLIN. If only the intensity is available, then the speckle should first be spatially whitened (by inverting the SAR transfer function [31], [32] or by subsampling [33]) before applying MERLIN on the pseudo-SLC image obtained by drawing random phases.

Another point worth discussion is the inversion of the SAR transfer function. The loss functions in Eqs.(6) and (7) were derived from the *marginal distributions*. Starting from the full

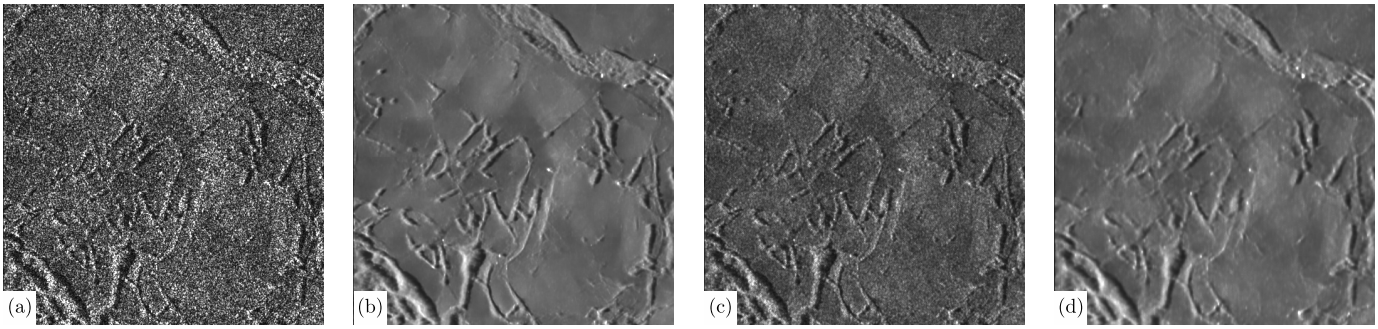

Figure 8. (a) A TerraSAR-X image in stripmap mode. (b) Image restored with MERLIN. (c)-(d) Images restored by a U-Net trained with self-supervision by replacing the real phase with a random phase sampled uniformly in $[-\pi, \pi]$. To produce the result shown in (d) input images are subsampled by a factor of two both at training time and at test time; pixels of the filtered image are then interpolated to recover the original image size

distribution, a different loss function would be obtained:

$$\mathcal{L}_{\text{full}}(\boldsymbol{r}, \tilde{\boldsymbol{b}}) = \sum_k \left(\frac{1}{2} \log r_k \right) + \tilde{\boldsymbol{b}}^\top \left[\mathbf{H} \text{diag}(\boldsymbol{r}) \mathbf{H}^\top \right]^{-1} \tilde{\boldsymbol{b}}, \quad (9)$$

which boils down to Eq.(6) when $\mathbf{H} = \mathbf{I}$. The loss function of Eq.(9) is much more costly to evaluate since $\mathbf{H}$ corresponds to a Toeplitz-block Toeplitz matrix (a 2D convolution in the direct domain, or a product in the Fourier domain). The inversion $\left[\mathbf{H} \text{diag}(\boldsymbol{r}) \mathbf{H}^\top \right]^{-1}$ can not be derived in closed form, which makes the training much more costly (several iterations of a conjugate gradients algorithm would typically be necessary for each evaluation of the loss function). Moreover, some form of regularization on $\boldsymbol{r}$ would be necessary since the data $\tilde{\boldsymbol{a}}$ and $\tilde{\boldsymbol{b}}$ are not sufficient to constrain $\boldsymbol{r}$ beyond the cutoff frequency of the SAR system.

As illustrated in Fig.3 and commented in section IV-A, the self-supervision with MERLIN comes at a cost: the real and imaginary components are processed separately. This limits the performance compared to joint processing of both components since the network is forced to handle images with a worse SNR (by a factor $1/\sqrt{2}$). This is partially compensated when the restorations computed separately on each component are finally combined. We think that this drawback is amply compensated by the good adaptation of the network to the SAR transfer function and the ability with MERLIN to use very large training sets (possibly, entire archives from a sensor).

Finally, compared to other self-supervision approaches, MERLIN imposes no limitation to the architecture of the network and does not assume a spatially decorrelated speckle. Together with the possibility to straightforwardly include huge archives of images in the training set, this opens new possibilities to consider highly expressive network architectures (e.g., very deep) for despeckling.

VI. Conclusion

We have shown that single-look complex images offer an ideal framework to self-train despeckling networks. The proposed generic training approach, called MERLIN, imposes no constrain on the architecture of the network. It completely suppresses the hassle of building training sets with reference images. We believe that this will dramatically change the way deep despeckling networks are used: networks specific to a sensor/acquisition mode can be easily trained, reaching a higher performance than general-purpose networks. Relieved from the worry of building training sets, future work can focus on designing clever network architectures. With MERLIN, very large scale training using entire archives of SAR images produced with a specific sensor mode can be contemplated, which could potentially lead to unprecedented despeckling performances.

Acknowledgments

The airborne SAR images processed in this paper were provided by ONERA, the French aerospace lab, within the project ALYS ANR-15-ASTR-0002 funded by the DGA (Direction Générale à l'Armement) and the ANR (Agence Nationale de la Recherche). Some of the TerraSAR-X images have been provided by the German Space Agency DLR for the project DLR-MTH0232.

References

[1] J. Lee, "Digital image smoothing and the sigma filter," *Computer vision, graphics, and image processing*, vol. 24, no. 2, pp. 255–269, 1983.

[2] G. Fracastoro, E. Magli, G. Poggi, G. Scarpa, D. Valsesia, and L. Verdoliva, "Deep learning methods for SAR image despeckling: trends and perspectives," *arXiv preprint arXiv:2012.05508*, 2020.

[3] X. Zhu, S. Montazeri, M. Ali, Y. Hua, Y. Wang, L. Mou, Y. Shi, F. Xu, and R. Bamler, "Deep learning meets SAR: concepts, models, pitfalls, and perspectives," *IEEE Geoscience and Remote Sensing Magazine (GRSM)*, 2021.

[4] L. Denis, E. Dalsasso, and F. Tupin, "A review of deep-learning techniques for SAR image restoration," in *IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2021.

[5] G. Vasile, E. Trouvé, J. Lee, and V. Buzuloiu, "Intensity-driven adaptive-neighborhood technique for polarimetric and interferometric SAR parameters estimation," *IEEE Trans. Geos. Remote Sens.*, vol. 44, no. 6, pp. 1609–1621, 2006.

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

[7] G. Aubert and J.-F. Aujol, "A variational approach to removing multiplicative noise," *SIAM journal on applied mathematics*, vol. 68, no. 4, pp. 925–946, 2008.

[8] L. Denis, F. Tupin, J. Darbon, and M. Sigelle, "SAR image regularization with fast approximate discrete minimization," *IEEE Trans. Image Proc.*, vol. 18, no. 7, pp. 1588–1600, 2009.

[9] J. M. Bioucas-Dias and M. A. Figueiredo, "Multiplicative noise removal using variable splitting and constrained optimization," *IEEE Trans. Image Proc.*, vol. 19, no. 7, pp. 1720–1730, 2010.

[10] G. Steidl and T. Teuber, "Removing multiplicative noise by Douglas-Rachford splitting methods," *Journal of Mathematical Imaging and Vision*, vol. 36, no. 2, pp. 168–184, 2010.

[11] J.-F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle, "Image decomposition application to SAR images," in *International Conference on Scale-Space Theories in Computer Vision*. Springer, 2003, pp. 297–312.

[12] S. Lobry, L. Denis, and F. Tupin, "Multitemporal SAR image decomposition into strong scatterers, background, and speckle," *IEEE Jour. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 9, no. 8, pp. 3419–3429, 2016.

[13] H. Xie, L. E. Pierce, and F. T. Ulaby, "SAR speckle reduction using wavelet denoising and Markov random field modeling," *IEEE Trans. Geos. Remote Sens.*, vol. 40, no. 10, pp. 2196–2212, 2002.

[14] S. Durand, J. Fadili, and M. Nikolova, "Multiplicative noise cleaning via a variational method involving curvelet coefficients," in *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2009, pp. 282–294.

[15] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Simul*, vol. 4, pp. 490–530, 2005.

[16] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering," *IEEE Trans. Imag. Proc.*, vol. 16, no. 8, pp. 2080–2095, Agu.

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

[18] F. Tupin, L. Denis, C.-A. Deledalle, and G. Ferraioli, "Ten Years of Patch-Based Approaches for SAR Imaging: A Review," in *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2019, pp. 5105–5108.

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

[20] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 50, no. 2, pp. 606–616, 2011.

[21] D. Cozzolino, S. Parrilli, G. Scarpa, G. Poggi, and L. Verdoliva, "Fast adaptive nonlocal SAR despeckling," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 2, pp. 524–528, 2013.

[22] C.-A. Deledalle, L. Denis, and F. Tupin, "NL-InSAR: Nonlocal interferogram estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 4, pp. 1441–1452, 2010.

[23] J. Chen, Y. Chen, W. An, Y. Cui, and J. Yang, "Nonlocal filtering for polarimetric SAR data: A pretest approach," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 5, pp. 1744–1754, 2010.

[24] L. Torres, S. J. Sant'Anna, C. da Costa Freitas, and A. C. Frery, "Speckle reduction in polarimetric SAR imagery with stochastic distances and nonlocal means," *Pattern Recognition*, vol. 47, no. 1, pp. 141–157, 2014.

[25] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, "NL-SAR: A unified nonlocal framework for resolution-preserving (Pol)(In) SAR denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 53, no. 4, pp. 2021–2038, 2014.

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

[27] ——, "Generative adversarial network-based restoration of speckled SAR images," in *2017 IEEE CAMSAP*. IEEE, 2017, pp. 1–5.

[28] 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, p. 196, 2018.

[29] 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, 2019.

[30] E. Dalsasso, X. Yang, L. Denis, F. Tupin, and W. Yang, "SAR Image Despeckling by Deep Neural Networks: from a pre-trained model to an end-to-end training strategy," *Remote Sens.*, vol. 12, no. 16, p. 2636, 2020.

[31] A. Lapini, T. Bianchi, F. Argenti, and L. Alparone, "Blind speckle decorrelation for SAR image despeckling," *IEEE Trans. Geos. Remote Sens.*, vol. 52, no. 2, pp. 1044–1058, 2013.

[32] R. Abergel, L. Denis, S. Ladjal, and F. Tupin, "Subpixellic methods for sidelobes suppression and strong targets extraction in single look complex SAR images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 11, no. 3, pp. 759–776, 2018.

[33] E. Dalsasso, L. Denis, and F. Tupin, "How to handle spatial correlations in SAR despeckling? Resampling strategies and deep learning approaches," in *13th European Conference on Synthetic Aperture Radar (EUSAR)*. VDE ITG, 2021, pp. 1233–1238.

[34] G. Chierchia, D. Cozzolino, G. Poggi, and L. Verdoliva, "SAR image despeckling through convolutional neural networks," in *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2017, pp. 5438–5441.

[35] D. Cozzolino, L. Verdoliva, G. Scarpa, and G. Poggi, "Nonlocal CNN SAR Image Despeckling," *Remote Sens.*, vol. 12, no. 6, p. 1006, 2020.

[36] E. Dalsasso, L. Denis, and F. Tupin, "SAR2SAR: a semi-supervised despeckling algorithm for SAR images," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 4321–4329, 2021.

[37] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "Towards Deep Unsupervised Sar Despeckling with Blind-Spot Convolutional Neural Networks," in *IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium*, 2020, pp. 2507–2510.

[38] ——, "Speckle2void: Deep self-supervised sar despeckling with blind-spot convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, 2021.

[39] S. Laine, T. Karras, J. Lehtinen, and T. Aila, "High-quality self-supervised deep image denoising," in *Advances in Neural Information Processing Systems*, 2019, pp. 6970–6980.

[40] K. Lee and W.-K. Jeong, "Noise2Kernel: Adaptive Self-Supervised Blind Denoising using a Dilated Convolutional Kernel Architecture," *arXiv preprint arXiv:2012.03623*, 2020.

[41] Y. Xie, Z. Wang, and S. Ji, "Noise2Same: Optimizing A Self-Supervised Bound for Image Denoising," *arXiv preprint arXiv:2010.11971*, 2020.

[42] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning Image Restoration without Clean Data," in *International Conference on Machine Learning*. PMLR, 2018, pp. 2965–2974.

[43] J. W. Goodman, *Speckle phenomena in optics: theory and applications*. Roberts and Company Publishers, 2007.

[44] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[45] J. Zhang, T. He, S. Sra, and A. Jadbabaie, "Why gradient clipping accelerates training: A theoretical justification for adaptivity," *arXiv preprint arXiv:1905.11881*, 2019.

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

[47] R. Baqué, O. R. du Plessis, N. Castet, P. Fromage, J. Martinot-Lagarde, J.-F. Nouvel, H. Oriot, S. Angelliaume, F. Brigui, H. Cantalloube *et al.*, "SETHI/RAMSES-NG: New performances of the flexible multi-spectral airborne remote sensing research platform," in *2017 European Radar Conference (EURAD)*. IEEE, 2017, pp. 191–194.

[48] J.-P. Ovarlez, F. Pascal, and P. Forster, "Covariance Matrix Estimation in SIRV and Elliptical Processes and Their Applications in Radar Detection," in *Modern Radar Detection*, 2015, pp. 295–332.

[49] H. Sportouche, J.-M. Nicolas, and F. Tupin, "Mimic Capacity Of Fisher And Generalized Gamma Distributions For High Resolution SAR Image Statistical Modeling," *IEEE Jour. Sel. Top. App. Earth Obs. Remote Sens.*, vol. 10, no. 12, pp. 5724–5735, 2017.

[50] J.-M. Nicolas and F. Tupin, "A New Parameterization for the Rician Distribution," *IEEE Geoscience and Remote Sensing Letters*, vol. 17, no. 11, 2020.

[51] D.-X. Yue, F. Xu, A. C. Frery, and Y.-Q. Jin, "Synthetic Aperture Radar Image Statistical Modeling: Part One - Single-Pixel Statistical Models," *IEEE Geoscience and Remote Sensing Magazine*, vol. 9, no. 1, 2021.

[52] N. Miranda, "Definition of the TOPS SLC deramping function for products generated by the S-1 IPF," *Eur. Space Agency, Paris, France, Tech. Rep*, 2014.