Self-supervised training strategies for SAR image despeckling with deep neural networks

Emanuele Dalsasso[a], Loïc Denis[b], Max Muzeau[c,d], and Florence Tupin[a]
[a]LTCI, Télécom Paris, Institut Polytechnique de Paris, France
[b]Univ Lyon, UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT- ETIENNE, France
[c]SONDRA, CentraleSupélec, Université Paris-Saclay, 91192 Gif-sur-Yvette, France
[d]DEMR, ONERA, Université Paris-Saclay, 91120 Palaiseau, France

Abstract

Images acquired by Synthetic Aperture Radar (SAR) are affected by speckle, making their interpretation difficult. Most recently, the rise of deep learning algorithms has led to groundbreaking results. The training of a neural network typically requires matched pairs of speckled / speckle-free images. To account for the speckle present in actual images and simplify the generation of training sets, self-supervision approaches directly train the network on speckled SAR data. Self-supervision exploits a form of diversity, either temporal, spatial, or based on the real/imaginary parts. We compare the requirements in terms of data preprocessing and the performance of three self-supervised strategies.

1 Introduction

The coherent summation of several echoes produced by elementary scatterers located within the radar resolution cell leads to strong fluctuations: the *speckle phenomenon*. These fluctuations are detrimental to the visual or automated analysis of SAR images. In homogeneous areas with rough surfaces or scattering volumes, the speckle is well described by Goodman's model [1]. Numerous approaches have been developed to reduce speckle fluctuations, from selection filters to wavelets and variational techniques [2], patch-based methods [3], and, more recently, deep neural networks [4, 5, 6].

The advent of deep learning techniques has led to results of unprecedented quality in the field of image restoration, in particular in the case of images corrupted by an additive white Gaussian noise. Convolutional neural networks (CNNs) can be trained to learn mapping noisy images to their noise-free counterparts [7]. While pairs of high-quality / low-quality images can easily be obtained for natural images, for example by varying the exposure time, it is much harder in remote sensing and in particular in SAR imaging. Almost speckle-free images can be obtained by averaging long time series of SAR images or by using optical images. The speckled versions of these ground truth images can then be obtained by generating synthetic speckle according to Goodman's model [8]. These simulations generally do not account for the spatial correlations of speckle observed in actual SAR images: this leads to a domain-shift between network training and the application to SAR images and requires an additional image preprocessing to reduce the correlations. Without an adequate preprocessing, networks trained on white noise lead to strong artefacts when applied to correlated speckle [9].

Training despeckling networks directly on SAR images, without ground truth, offers several advantages: (i) building the training set is much easier, it can thus include large amounts of images; (ii) there is no domain-shift between network training and network application since the network is trained directly on actual speckle, provided that images with a similar content are present in the training set. Self-supervised despeckling methods are based on a common principle: if a signal contains a deterministic and a stochastic component, only the deterministic component can be predicted given an independent realization of the signal. Self-supervised training then starts by splitting the data into subsets between which speckle is statistically independent. A network is then trained to predict, from a first subset of the data, the observations of the second subset. As described in more details in Section 3, the splitting of the data can exploit different kinds of information diversity: temporal [10], spatial [11], or even real / imaginary parts of single-look complex (SLC) images [12]. Each splitting approach has different advantages and limitations. The goal of this work is to analyze these differences, compare the performance of three neural networks trained in a self-supervised way on some TerraSAR-X Stripmap images, and draw some conclusions.

2 Goodman's speckle model

Goodman's fully developed speckle model considers rough and homogeneous areas, so that it can be assumed that, for each elementary scatterer within a resolution cell,

- the phase and amplitude are both independent and identically distributed (i.i.d.),
- phase and amplitude are independent from one an-

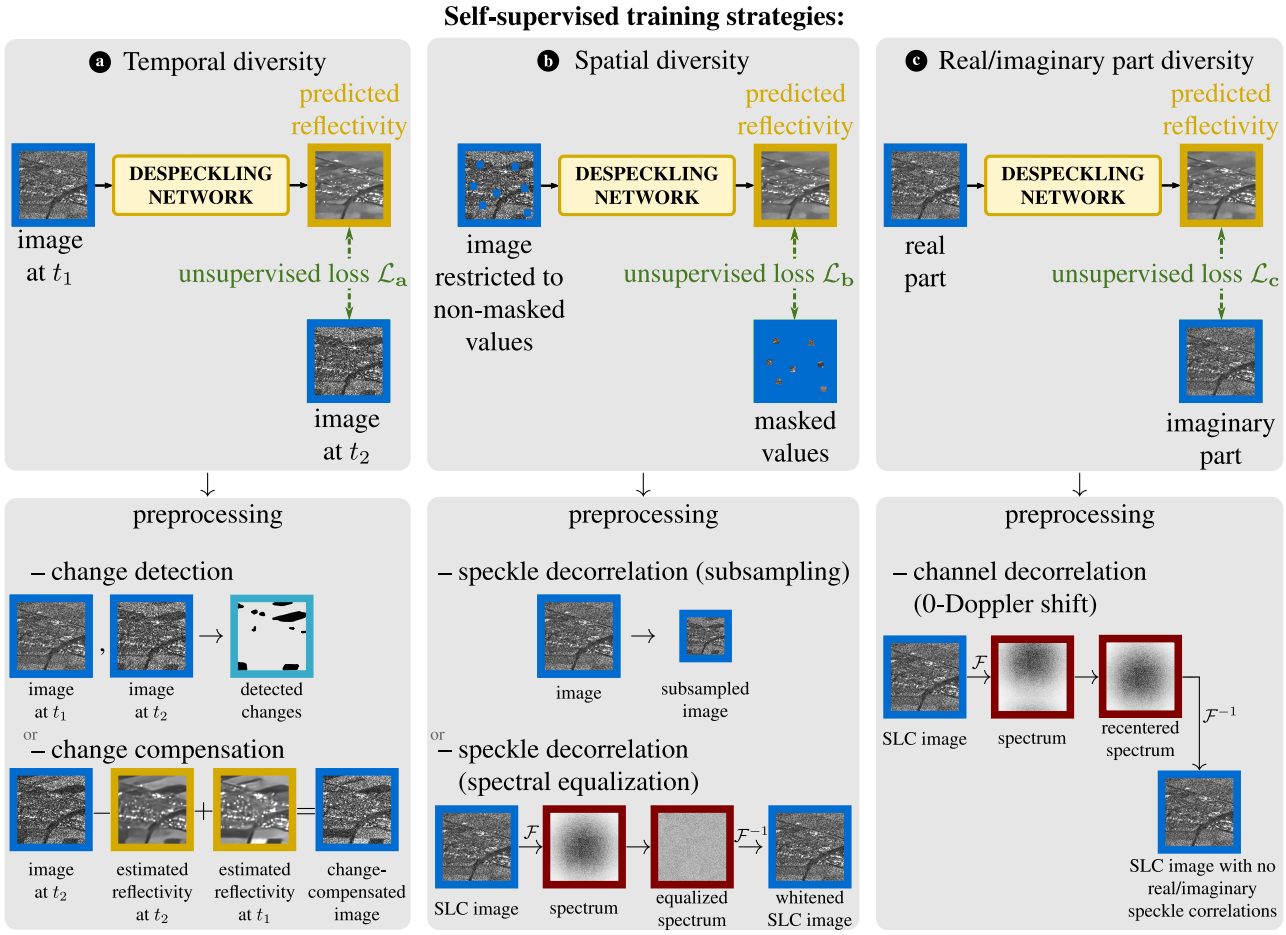

Figure 1 Three approaches for self-supervised training strategies are possible: ⓐ using temporal diversity (pairs of images captured at different dates); ⓑ using spatial diversity (by masking part of the spatial information); and ⓒ using the real / imaginary part diversity (by feeding only the real or imaginary part to the network and supervising the training with the other component). Each strategy requires a different preprocessing of the data to ensure that the statistical guarantees for self-supervised training are fulfilled.

other,

- phases are uniformly distributed between $-\pi$ and π ,
- the number of elementary scatterers within a resolution cell is large.

By application of the central limit theorem, the real and imaginary parts of the complex amplitude $z = a + jb$ that results from the coherent summation of all elementary complex phasors follow independent Gaussian distributions:

$$\mathrm{p}_{A,B}(a,b) = \underbrace{\frac{1}{\sqrt{\pi r}} \exp\left(-\frac{a^2}{r}\right)}_{\mathcal{N}(0,r/2)} \underbrace{\frac{1}{\sqrt{\pi r}} \exp\left(-\frac{b^2}{r}\right)}_{\mathcal{N}(0,r/2)}, \quad (1)$$

where r represents the reflectivity of the resolution cell. The intensity $i = |z|^2 = a^2 + b^2$ thus follows an exponential distribution:

$$\mathrm{p}_I(i) = \frac{1}{r} \exp\left(-\frac{i}{r}\right) \quad (2)$$

and the mean intensity over a homogeneous area is $\mathbb{E}[I] = r$ while the variance is $\mathrm{Var}[I] = r^2$. In the case of multi-looked intensities, the distribution becomes a gamma law:

$$\mathrm{p}_I(i) = \frac{L^L i^{L-1}}{\Gamma(L) r^L} \exp\left(-L\frac{i}{r}\right), \quad (3)$$

with L the number of looks and $\Gamma(.)$ the gamma function.

Goodman's speckle model provides the ground for deriving the self-supervised training loss functions described next.

3 Self-supervised training strategies

3.1 Principle

Deep neural networks map an input $\boldsymbol{y}$ (the speckled image) to an output $\widehat{\boldsymbol{x}}$ (the image of estimated reflectivities) through a non-linear function $f_{\boldsymbol{\theta}}(\cdot)$ parameterized by a vector of weights $\boldsymbol{\theta}$ (the network coefficients). During training, weights $\boldsymbol{\theta}$ are optimized with respect to a criterion: the training loss function $\mathcal{L}$.

In a supervised training setting, the vector of weights is

obtained by solving a minimization problem of the form:

$$\arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \mathcal{L}\left[f_{\boldsymbol{\theta}}(\boldsymbol{y}_k), \boldsymbol{x}_k^\star\right], \tag{4}$$

where the set of all pairs $\{(\boldsymbol{y}_k, \boldsymbol{x}_k^\star)\}_{k=1..K}$ forms the training set and $\mathcal{L}$ measures the proximity between the estimate $\widehat{\boldsymbol{x}}_k = f_{\boldsymbol{\theta}}(\boldsymbol{y}_k)$, computed based on the noisy image $\boldsymbol{y}_k$, and the corresponding ground truth $\boldsymbol{x}_k^\star$. The loss functions $\mathcal{L}$ typically chosen are the squared L2 norm, the L1 norm, or perceptual distances (e.g., SSIM or L2 distances computed between low-level features of a pretrained network [13]).

Self-supervised training approaches do not require access to the ground truth images $\boldsymbol{x}_k^\star$. Instead, each noisy sample $\boldsymbol{y}_k$ of the training set is split into two parts: $\boldsymbol{y}_k^{\text{in}}$ and $\boldsymbol{y}_k^{\text{val}}$. Only the first part $\boldsymbol{y}_k^{\text{in}}$ is fed to the network and the other part $\boldsymbol{y}_k^{\text{val}}$ is used to evaluate the loss function:

$$\arg\min_{\boldsymbol{\theta}} \sum_{k=1}^{K} \mathcal{L}\left[f_{\boldsymbol{\theta}}(\boldsymbol{y}_k^{\text{in}}), \boldsymbol{y}_k^{\text{val}}\right]. \tag{5}$$

In contrast to the supervised case (4), the loss function now evaluates the proximity between the estimated image $\widehat{\boldsymbol{x}}_k = f_{\boldsymbol{\theta}}(\boldsymbol{y}_k^{\text{in}})$ and the other part of the noisy data $\boldsymbol{y}_k^{\text{val}}$. For the training to work, images $\boldsymbol{y}_k^{\text{in}}$ and $\boldsymbol{y}_k^{\text{val}}$ must correspond to a common (yet unknown) ground truth image $\boldsymbol{x}_k^\star$ and also be statistically independent (the speckle component of each image is independent). Under these conditions, a natural choice for the loss function $\mathcal{L}$ is the neg-log-likelihood:

$$\mathcal{L}\left[f_{\boldsymbol{\theta}}(\boldsymbol{y}_k^{\text{in}}), \boldsymbol{y}_k^{\text{val}}\right] = -\log \mathrm{p}_Y\left(\boldsymbol{y}_k^{\text{val}} | \boldsymbol{x}_k^\star = f_{\boldsymbol{\theta}}(\boldsymbol{y}_k^{\text{in}})\right). \tag{6}$$

3.2 Variants of self-supervised training for despeckling

The way each training sample is split into $\boldsymbol{y}_k^{\text{in}}$ and $\boldsymbol{y}_k^{\text{val}}$ leads to quite different self-supervision methods. We cover in the following three approaches exploiting successively the temporal, spatial, or real /imaginary part diversities. These approaches are summarized in figure 1.

ⓐ Temporal diversity:

A natural way to obtain two speckle realizations of the same radar scene consists of collecting two images at two dates with a temporal separation sufficient to ensure the temporal decorrelation of the speckle. With increasing temporal separations, more changes might also occur within the scene. We postpone the discussion on how to handle those changes and consider for now that the radar scene remains basically unchanged except for the small changes that cause the speckle decorrelation. A training sample then corresponds to a pair of noisy intensities $(\boldsymbol{i}_k^{t_1}, \boldsymbol{i}_k^{t_2})$ of speckled images corresponding to the same (unavailable) ground truth reflectivity $\boldsymbol{r}_k^\star$. The splitting of a training sample can either be $\boldsymbol{y}_k^{\text{in}} = \boldsymbol{i}_k^{t_1}$ and $\boldsymbol{y}_k^{\text{val}} = \boldsymbol{i}_k^{t_2}$ or the converse. The unsupervised loss function can be derived from equations (6) and (3), under the approximation of spatially uncorrelated speckle:

$$\begin{aligned}\mathcal{L}_{\mathbf{a}}\left[f_{\boldsymbol{\theta}}(\boldsymbol{i}_k^{t_1}), \boldsymbol{i}_k^{t_2}\right] &= -\log \mathrm{p}_I\left(\boldsymbol{i}_k^{t_2} | \boldsymbol{r}_k^\star = f_{\boldsymbol{\theta}}(\boldsymbol{i}_k^{t_1})\right) \\ &\propto \sum_{\ell} \log\left[f_{\boldsymbol{\theta}}(\boldsymbol{i}_k^{t_1})\right]_\ell + \frac{\left[\boldsymbol{i}_k^{t_2}\right]_\ell}{\left[f_{\boldsymbol{\theta}}(\boldsymbol{i}_k^{t_1})\right]_\ell} + \text{cst.},\end{aligned} \tag{7}$$

where cst. represents a term that does not change when the network parameters $\boldsymbol{\theta}$ vary and the notation $[\cdot]_\ell$ represents the ℓ-th pixel of an image.

Note that, given the high dynamic range of SAR images, it is beneficial to provide log-transformed intensities to the network and to apply an exponential transform to the outputs of the network to compress the dynamic range.

In practice, it is necessary to handle the changes in the radar scene that occurred between times t_1 and t_2. Two approaches are possible: if these changes are of limited spatial expansion, they can be detected (using a change-detection method) and the pixels in the changed area discarded from the computation of the loss function [8]. A more refined way to account for changes is to apply a change compensation operation to image $\boldsymbol{i}_k^{t_2}$. This way, all changes, even the smallest, are accounted for. Since change compensation requires a method to estimate the reflectivities $\widehat{\boldsymbol{r}}^{t_1}$ and $\widehat{\boldsymbol{r}}^{t_2}$, the algorithm SAR2SAR [10] performs several training steps to progressively refine the despeckling network to benefit from improved change compensations.

ⓑ Spatial diversity:

In the absence of additional images of the same scene at other dates, data splitting can be spatial. The idea is to mask some observed values and to train the network to predict the values that are masked by computing the unsupervised loss on those masked values. Several techniques can be used to perform the masking: masked values can be replaced by a local average, a purely random value, zero, or the network architecture can be designed in order to achieve a "blind spot", i.e., the central area of the patch is unconnected to the output (the receptive field of the neural net does not contain this area). The latter approach leads to effective training because all observed pixels appear in turn in the loss function. Direct application of a network with a blind spot would be disappointing, especially in SAR imaging where point-like structures are common, since such punctual elements would be lost when falling within the blind spot. In order to also account for the observations inside the blind spot in the reflectivity estimate, the algorithm Speckle2Void [11] follows the Bayesian approach of [14]: the network predicts a parametric distribution $g_R(r; \boldsymbol{\alpha}_{k,\ell})$ (where $\boldsymbol{\alpha}_{k,\ell} = [f_{\boldsymbol{\theta}}(\mathcal{V}_{\boldsymbol{i}_k}(\ell))]_\ell$ is the vector of parameters output by the network, defining the function $r \mapsto g_R(r; \boldsymbol{\alpha}_{k,\ell})$). This parametric distribution characterizes the reflectivity $[\boldsymbol{r}_k]_\ell$ at pixel ℓ, based on the observed intensities $\mathcal{V}_{\boldsymbol{i}_k}(\ell)$ in the neighborhood of pixel ℓ (the neighborhood excludes pixel ℓ which represents the blind spot). It defines a prior that can then be combined with the actual observed intensity $[\boldsymbol{i}_k]_\ell$, for example using the posterior

mean estimator:

$$[\widehat{\boldsymbol{r}}_k]_\ell = \int r \frac{\mathrm{p}([\boldsymbol{i}_k]_\ell|r) g_R(r;\boldsymbol{\alpha}_{k,\ell})}{\int \mathrm{p}([\boldsymbol{i}_k]_\ell|r') g_R(r';\boldsymbol{\alpha}_{k,\ell})\,\mathrm{d}r'}\,\mathrm{d}r\,. \quad (8)$$

The training loss function is obtained from equation (6) by setting $\boldsymbol{y}_k^{\text{in}} = \mathcal{V}_{\boldsymbol{i}_k}(\ell)$ and $\boldsymbol{y}_k^{\text{val}} = [\boldsymbol{i}_k]_\ell$:

$$\mathcal{L}_{\mathbf{b}}\left[f_{\boldsymbol{\theta}}(\mathcal{V}_{\boldsymbol{i}_k}(\ell)), [\boldsymbol{i}_k]_\ell\right] = -\log \int \mathrm{p}([\boldsymbol{i}_k]_\ell|r) g_R(r;\boldsymbol{\alpha}_{k,\ell})\,\mathrm{d}r \quad (9)$$

with $\boldsymbol{\alpha}_{k,\ell} = [f_{\boldsymbol{\theta}}(\mathcal{V}_{\boldsymbol{i}_k}(\ell))]_\ell$.
In practice, it is much more convenient if the integrals in equations (8) and (9) are known in closed form. This is the case if the parametric distribution g_R is a conjugate prior of the likelihood $\mathrm{p}([\boldsymbol{i}_k]_\ell|r)$. In Speckle2Void [11], Molini *et al.* therefore use an inverse gamma distribution for g_R.

An important requirement to apply a spatial splitting is that the speckle be spatially uncorrelated. If it is correlated, then the network will be able to recover the masked noisy intensity $[\boldsymbol{i}_k]_\ell$ using only the neighborhood $\mathcal{V}_{\boldsymbol{i}_k}(\ell)$ and the resulting estimation will still be corrupted by speckle. To prevent this, speckle correlations must be reduced either by subsampling the image or by applying a spectral equalization (and possibly a resampling if the image has been oversampled), see [15, 16, 9]. Moreover, to obtain a blind spot centered on pixel ℓ, the architecture of the neural network must be carefully designed, which severely constrains the choice of the network.

ⓒ Real/imaginary part diversity:
In a recent paper [12], we have suggested another possible splitting of SLC training data for unsupervised training: the decomposition into the real and imaginary parts. Under reasonable assumptions (a SAR system with a real-valued response, i.e., with a transfer function with Hermitian symmetry), the real and imaginary parts of an SLC image are indeed statistically independent. In contrast to the approach ⓐ based on temporal diversity, there is no issue with possible changes between the two components since they are acquired simultaneously.
The splitting of a sample from the training set takes either the form $\boldsymbol{y}_k^{\text{in}} = \boldsymbol{a}_k$ and $\boldsymbol{y}_k^{\text{val}} = \boldsymbol{b}_k$ or the converse (real and imaginary parts can be swapped randomly during training), where we recall that $\boldsymbol{a}_k$ and $\boldsymbol{b}_k$ are the real and imaginary parts of the SLC image $\boldsymbol{z}_k$. From equations (6) and (1), the unsupervised training loss takes the form:

$$\begin{aligned}\mathcal{L}_{\mathbf{c}}\left[f_{\boldsymbol{\theta}}(\boldsymbol{a}_k), \boldsymbol{b}_k\right] &= -\log \mathrm{p}_B(\boldsymbol{b}_k|\boldsymbol{r}_k^\star = f_{\boldsymbol{\theta}}(\boldsymbol{a}_k)) \\ &= \sum_\ell \tfrac{1}{2}\log\left[f_{\boldsymbol{\theta}}(\boldsymbol{a}_k)\right]_\ell + \frac{[\boldsymbol{b}_k]_\ell^2}{[f_{\boldsymbol{\theta}}(\boldsymbol{a}_k)]_\ell} + \text{cst.}\end{aligned} \quad (10)$$

and

$$\mathcal{L}_{\mathbf{c}}\left[f_{\boldsymbol{\theta}}(\boldsymbol{b}_k), \boldsymbol{a}_k\right] = \sum_\ell \tfrac{1}{2}\log\left[f_{\boldsymbol{\theta}}(\boldsymbol{b}_k)\right]_\ell + \frac{[\boldsymbol{a}_k]_\ell^2}{[f_{\boldsymbol{\theta}}(\boldsymbol{b}_k)]_\ell} + \text{cst.} \quad (11)$$

In order to ensure that the real and imaginary parts $\boldsymbol{a}_k$ and $\boldsymbol{b}_k$ are independent, a preprocessing may be necessary to achieve a 0-Doppler shift and possibly to correct for spectrum asymmetries, as depicted at the bottom right of figure 1 and in [12].

4 Experimental comparison and discussion

We illustrate each self-supervised training strategy by despeckling TerraSAR-X Stripmap images with the following algorithms:

- SAR2SAR [10], trained using the temporal diversity (strategy ⓐ);
- Speckle2Void [11], trained using the spatial diversity (strategy ⓑ);
- MERLIN [12], trained using the real / imaginary part diversity (strategy ⓒ).

The networks in SAR2SAR and MERLIN share the same architecture derived from the U-Net of [17]. The network used in Speckle2Void has a specific architecture in order to enforce a blind spot at the center of the receptive field. To ensure a fair comparison, all three training strategies are applied on a dataset of TerraSAR-X Stripmap images. As a baseline, we also include the results of SAR-BM3D [18] which is not based on a deep neural network.
Figure 2 displays restoration results for each method as well as the residual noise, i.e., the ratio image $\boldsymbol{i}_k/\widehat{\boldsymbol{r}}_k$. Upon successful estimation of the reflectivity image $\boldsymbol{r}_k$, the residual $\boldsymbol{i}_k/\widehat{\boldsymbol{r}}_k$ should correspond solely to the speckle component and contain no structure from the original intensity image.
By close visual inspection, it appears that the baseline method, SAR-BM3D, produces some artifacts in smooth areas (fields, high vegetation) or around strong scatterers in the form of streaks or oscillations. SAR2SAR provides the sharpest restorations but suppresses less strongly the speckle fluctuations: remaining fluctuations can still be observed in some areas. Speckle2Void restores satisfyingly smooth areas but has difficulties to handle strong punctual scatterers. Finally, MERLIN seems to offer the smoothest estimates for vegetated areas and produces an estimation that is slightly more blurry than SAR2SAR on urban areas. The residual noise images display more structure in the images produced by SAR-BM3D and Speckle2Void than with MERLIN or SAR2SAR. It seems that the latter removes almost no significant structure from the original image: no extended pattern can be identified in the residual image.
Beyond the analysis of restoration quality, there are pros and cons for each approach that are summarized in Table 1. Exploiting temporal diversity requires pairs of registered intensity images with sufficient temporal separation to ensure speckle decorrelation. Such image pairs may not be available when considering airborne SAR systems, which limits the applicability of the approach. Moreover, the change detection and/or compensation must be reliable

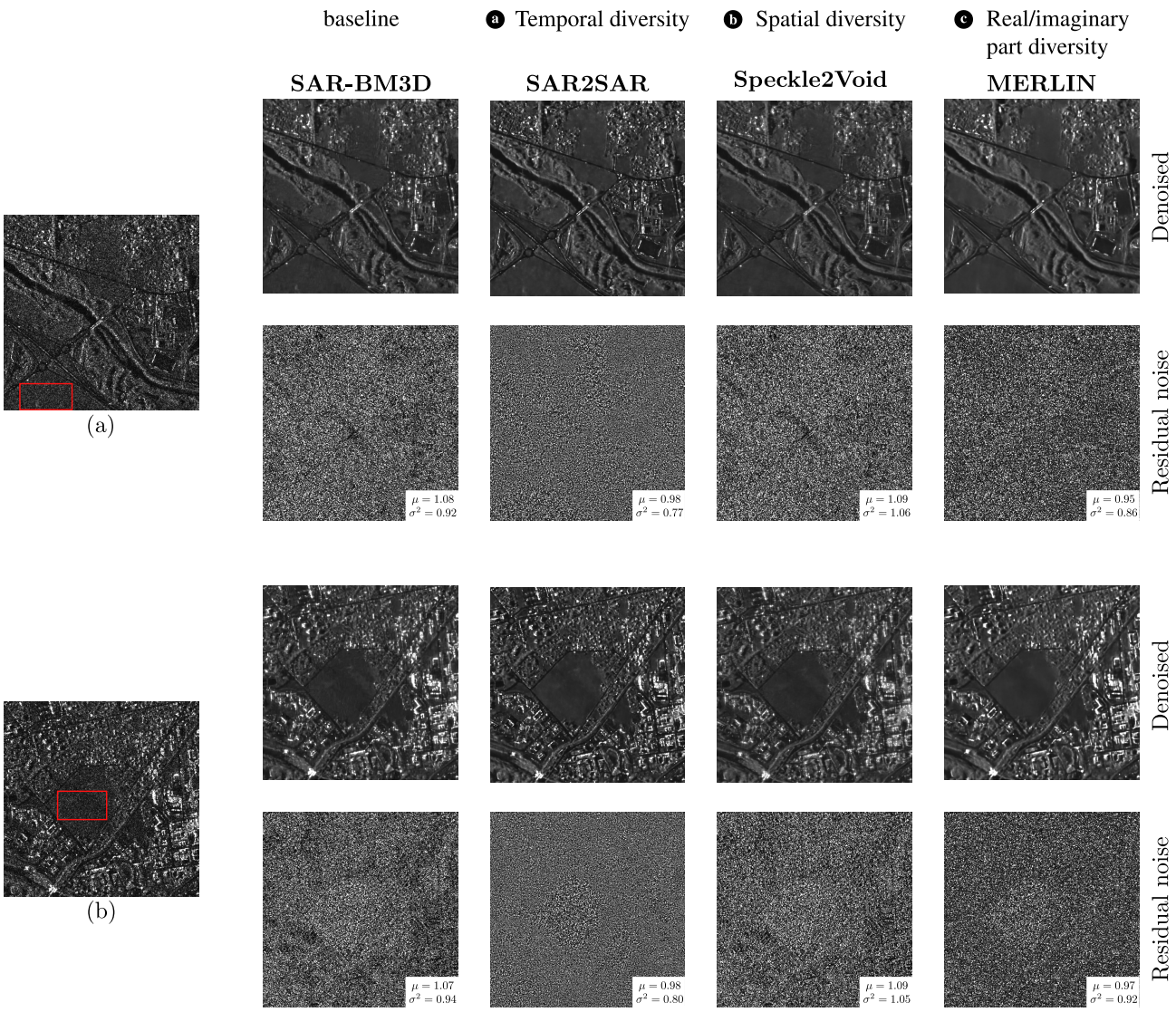

Figure 2 Results on two TerraSAR-X Stripmap images. For each method, the ratio noisy/denoised is shown to check if some structures are removed from the original image by the despeckling step. The statistics (mean and variance) estimated on the ratio image in the area surrounded by the red rectangle indicates that all algorithms are effective at removing speckle in physically homogeneous areas (the closer to 1, the better).

	a Temporal diversity SAR2SAR [10]	**b Spatial diversity** Speckle2Void [11]	**c Real/imaginary diversity** MERLIN [12]
training data requirement	pairs of registered intensity images	intensity images	SLC images
preprocessing impact	bias if imperfect change detection/compensation	spectral equalization changes image appearance	limited impact
training complexity	+++	++	+
flexibility of net architecture	+++	+	+++
fine tuning to new images	+	+++	+++

Table 1 Summary of the advantages and limitations of each self-supervised training strategy.

otherwise a bias may appear in the estimates. Updating the change compensation method with the despeckling network leads to a training in several steps which makes it heavier than other self-supervised approaches. Since training requires pairs of images at different dates, fine tuning the network on a new image (with a content that dif-

fers from the training set) may not be possible if only a single image is available. Speckle2Void, based on spatial diversity, requires only intensity images, which simplifies the building of the training set and the fine tuning to new images. The requirement of spatially uncorrelated speckle leads to a preprocessing that changes the image appearance (a resampling may be necessary and the spectral apodization is removed). Only very specific network architectures lead to a blind spot in the receptive field, this limits the design choices and our experiments indicate that the network requires more care to reach convergence than a simpler U-Net network. MERLIN self-supervised training strategy only requires SLC images. The preprocessing step ensuring that the spectrum is symmetrical has a very limited impact on the image appearance and on the computational burden. Training is easy and in practice can be performed even on a single image. Unlike Speckle2Void, any network architecture can be considered, which allows the straightforward transfer of evolved network architectures developed in the field of natural image restoration.

5 Conclusions

Self-supervised training of deep neural networks offers many advantages over conventional supervised training techniques: building the training set is greatly simplified and generalization to unseen data is largely improved given that actual SAR images are used in the training phase rather than images with synthetic speckle. There now exists several approaches to perform self-supervision, based on common ideas of splitting each training sample into two parts: one part fed to the neural network and the other used to compute the self-supervised training loss. Depending on whether a temporal, spatial or real/imaginary part separation is used, various methods have been proposed in the literature. Each offers distinct advantages and drawbacks both in terms of restoration quality and practicality.

As the reference image is not needed, with self-supervised methods it is possible to fine-tune the network on a given area of interest leading to improved performances at test time on that specific area. This is particularly true for strategies requiring spatial diversity or real/imaginary part diversity, where a single image is sufficient to supervise the training.

6 Acknowledgments

This project has been funded by ANR (the French National Research Agency) and DGA (Direction Générale de l'Armement) under ASTRAL project ANR-21-ASTR-0011.

7 Literature

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

[2] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone, "A tutorial on speckle reduction in synthetic aperture radar images," *IEEE GRSM*, 2013.

[3] C.-A. Deledalle, L. Denis, G. Poggi, F. Tupin, and L. Verdoliva, "Exploiting patch similarity for SAR image processing: The nonlocal paradigm," *IEEE Signal Processing Magazine*, 2014.

[4] 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 GRSM*, 2021.

[5] G. Fracastoro, E. Magli, G. Poggi, G. Scarpa, D. Valsesia, and L. Verdoliva, "Deep learning methods for SAR image despeckling: trends and perspectives," *GRSM*, 2021.

[6] B. Rasti, Y. Chang, E. Dalsasso, L. Denis, and P. Ghamisi, "Image restoration for remote sensing: Overview and toolbox," *GRSM*, 2021.

[7] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE TIP*, 2017.

[8] S. Vitale, G. Ferraioli, and V. Pascazio, "Analysis on the building of training dataset for deep learning SAR despeckling," *IEEE GRSL*, 2021.

[9] E. Dalsasso, L. Denis, and F. Tupin, "How to handle spatial correlations in SAR despeckling? resampling strategies and deep learning approaches," in *EUSAR*, 2021.

[10] ——, "SAR2SAR: a semi-supervised despeckling algorithm for SAR images," *IEEE JSTARS*, 2021.

[11] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "Speckle2void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks," *IEEE TGRS*, 2021.

[12] E. Dalsasso, L. Denis, and F. Tupin, "As if by magic: self-supervised training of deep despeckling networks with MERLIN," *TGRS*, 2022.

[13] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, "The unreasonable effectiveness of deep features as a perceptual metric," in *IEEE CVPR*, 2018.

[14] S. Laine, T. Karras, J. Lehtinen, and T. Aila, "High-quality self-supervised deep image denoising," in *NIPS*, 2019.

[15] A. Lapini, T. Bianchi, F. Argenti, and L. Alparone, "Blind speckle decorrelation for SAR image despeckling," *IEEE TGRS*, 2013.

[16] 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 JSTARS*, 2018.

[17] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," in *ICML*. PMLR, 2018.

[18] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, "A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage," *IEEE TGRS*, 2011.