Towards Deep Unsupervised SAR Despeckling with Blind-Spot Convolutional Neural Networks

TOWARDS DEEP UNSUPERVISED SAR DESPECKLING WITH BLIND-SPOT CONVOLUTIONAL NEURAL NETWORKS

Andrea Bordone Molini, Diego Valsesia, Giulia Fracastoro, Enrico Magli

Politecnico di Torino, Italy

ABSTRACT

SAR despeckling is a problem of paramount importance in remote sensing, since it represents the first step of many scene analysis algorithms. Recently, deep learning techniques have outperformed classical model-based despeckling algorithms. However, such methods require clean ground truth images for training, thus resorting to synthetically speckled optical images since clean SAR images cannot be acquired. In this paper, inspired by recent works on blind-spot denoising networks, we propose a self-supervised Bayesian despeckling method. The proposed method is trained employing only noisy images and can therefore learn features of real SAR images rather than synthetic data. We show that the performance of the proposed network is very close to the supervised training approach on synthetic data and competitive on real data.

***Index Terms*—** SAR, speckle, convolutional neural networks, unsupervised

1. INTRODUCTION

Synthetic Aperture Radar (SAR) is a coherent imaging system and as such it strongly suffers from the presence of speckle, a signal dependent granular noise. Speckle noise makes SAR images difficult to interpret, preventing the effectiveness of scene analysis algorithms for, e.g., image segmentation, detection and recognition. Several despeckling methods applied to SAR images have been proposed working either in spatial or transform domain. The first attempts at despeckling employed filtering-based techniques operating in spatial domain such as Lee filter [1], Frost filter [2], Kuan filter [3], and Gamma-MAP filter [4]. Wavelet-based methods [5, 6] enabled multi-resolution analysis. More recently, non-local filtering methods attempted to exploit self-similarities and contextual information. A combination of non-local approach, wavelet domain shrinkage and Wiener filtering in a two-step process led to SAR-BM3D [7], a SAR-oriented version of BM3D [8].

In recent years, deep learning techniques have set the benchmark in many image processing tasks, achieving exceptional results in problems such as image restoration [9], super resolution [10], semantic segmentation [11]. Recently, some despeckling methods based on convolutional neural networks (CNNs) have been proposed [12, 13], attempting to leverage the feature learning capabilities of CNNs. Such methods use a supervised training approach where the network weights are optimized by minimizing a distance metric between noisy inputs and clean targets. However, clean SAR images do not exist and supervised training methods resort to synthetic datasets where optical images are used as ground truth and their artificially speckled version as noisy inputs. This creates a domain gap between the features of synthetic training data and those of real SAR images, possibly leading to presence of artifacts or poor preservation of radiometric features. SAR-CNN [13] addressed this problem by averaging multi-temporal SAR data of the same scene to obtain a ground truth. However, acquisition of multi-temporal data, scene registration and robustness to variations can be challenging.

Self-supervised denoising methods represent an alternative to train CNNs without having access to the clean images. Noise2Noise [14] proposed to use pairs of images with the same content but independent noise realizations. This method is not suitable for SAR despeckling due to the difficulty in accessing multiple images of the same scene with independently drawn noise realizations. Noise2void [15] further relaxes the constraints on the dataset, requiring only a single noisy version of the training images, by introducing the concept of blind-spot networks. Assuming spatially uncorrelated noise, and excluding the center pixel from receptive field of the network, the network learns to predict the value of the center pixel from its receptive field by minimizing the ℓ_2 distance between the prediction and the noisy value. The network is prevented from learning the identity mapping because the pixel to be predicted is removed from the receptive field. The blind-spot scheme used in Noise2void [15] is carried out by a simple masking method, keeping a few pixels active in the learning process. Laine et al. [16] devised a novel convolutional blind-spot network architecture capable of processing the entire image at once, increasing the efficiency. They also introduce a Bayesian framework to include noise models and priors on the conditional distribution of the blind spot given the receptive field.

In this paper, we use the self-supervised Bayesian denoising with blind-spot networks proposed in [16], adapting the model to the noise and image statistics of SAR images, thus

This research has been funded by the Smart-Data@PoliTO center for Big Data and Machine Learning technologies.

enabling direct training on real SAR images. Our method bypasses the problem of training a CNN on synthetically-speckled optical images and using it to denoise SAR images, since in general transfer knowledge from optical to SAR images is a very difficult task as imaging geometries and content are quite dissimilar due to the different imaging mechanisms. To the best of our knowledge, this is the first self-supervised method to deal with real SAR images.

2. BACKGROUND

CNN denoising methods estimate the clean image by learning a function that takes each noisy pixel and combines its value with the local neighboring pixel values (receptive field) by means of multiple convolutional layers interleaved with non-linearities. Taking this from a statistical inference perspective, a CNN is a point estimator of $p(x_i|y_i, \Omega_{y_i})$, where x_i is the i^{th} clean pixel, y_i is the i^{th} noisy pixel and Ω_{y_i} represents the receptive field composed of the noisy neighboring pixels, excluding y_i itself. Noise2void predicts the clean pixel x_i by relying solely on the neighboring pixels and using y_i as a noisy target. The CNN learns to produce an estimate of $\mathbb{E}_{x_i}[x_i|\Omega_{y_i}]$ using the ℓ_2 loss when in presence of Gaussian noise. The drawback of Noise2void is that the value of the noisy pixel y_i is never used to compute the clean estimate.

The Bayesian framework devised by Laine et al. [16] explicitly introduces the noise model $p(y_i|x_i)$ and conditional pixel prior given the receptive field $p(x_i|\Omega_{y_i})$ as follows:

$$p(x_i|y_i, \Omega_{y_i}) \propto p(y_i|x_i)p(x_i|\Omega_{y_i}).$$

The role of the CNN is to predict the parameters of the chosen prior $p(x_i|\Omega_{y_i})$. The denoised pixel is then obtained as the MMSE estimate, i.e., it seeks to find $\mathbb{E}_{x_i}[x_i|y_i, \Omega_{y_i}]$. Under the assumption that the noise is pixel-wise i.i.d., the CNN is trained so that the data likelihood $p(y_i|\Omega_{y_i})$ for each pixel is maximized. The main difficulty involved with this technique is the definition of a suitable prior distribution that, when combined with the noise model, allows for close-form posterior and likelihood distributions. We also remark that while imposing a handcrafted distribution as $p(x_i|\Omega_{y_i})$ may seem very limiting, it is actually not since i) that is the *conditional* distribution given the receptive field rather than the raw pixel distribution, and ii) its hyperparameters are predicted by a powerful CNN on a pixel-by-pixel basis.

3. PROPOSED METHOD

Following the notation in Sec. 2, this section presents the Bayesian model we adopt for SAR despeckling and the training procedure. A summary is shown in Fig. 1.

3.1. Model

We consider the multiplicative SAR speckle noise model: $y_i = n_i x_i$ where x represents the unobserved clean image and n the uncorrelated multiplicative speckle. Concerning noise modeling, we choose the widely-used $\Gamma(L, L)$ distribution for an L-look image. We model the conditional prior

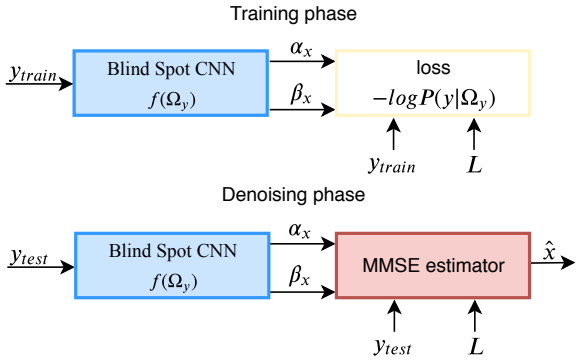

Fig. 1. Scheme depicting the training and the testing phases.

distribution given the receptive field as an inverse Gamma distribution with shape α_{x_i} and scale β_{x_i}:

$$p(x_i|\Omega_{y_i}) = \text{invG}(\alpha_{x_i}, \beta_{x_i}),$$

where α_{x_i} and β_{x_i} depend on Ω_{y_i}, since they are the outputs of the CNN at pixel i. For the chosen prior and noise models, the posterior distribution is also an inverse Gamma:

$$p(x_i|y_i, \Omega_{y_i}) = \text{invG}(L + \alpha_{x_i}, \beta_{x_i} + Ly_i). \quad (1)$$

Finally, the noisy data likelihood $p(y_i|\Omega_{y_i})$ can be obtained in closed form:

$$p(y_i|\Omega_{y_i}) = \frac{L^L y_i^{L-1}}{\beta_{x_i}^{-\alpha_{x_i}} Beta(L, \alpha_{x_i})(\beta_{x_i} + Ly_i)^{L+\alpha_{x_i}}},$$

with the Beta function defined as $Beta(L, \alpha_{x_i}) = \frac{\Gamma(L)\Gamma(\alpha_{x_i})}{\Gamma(L+\alpha_{x_i})}$. This distribution is also known as the G_I^0 distribution introduced in [17]. It has been observed that it is a good model of highly heterogeneous SAR data in intensity format like urban areas, primary forests and a deforested area.

3.2. Training

The training procedure learns the weights of the blind-spot CNN, which is used to produce the estimates for parameters α_{x_i} and β_{x_i} of the inverse gamma distribution $p(x_i|\Omega_{y_i})$. We refer the reader to [16] on how to implement a CNN so that it has a central blind spot. The blind-spot CNN is trained to minimize the negative log likelihood $p(y_i|\Omega_{y_i})$ for each pixel, so that the estimates of α_{x_i} and β_{x_i} fit the noisy observations. Our loss function is as follows:

$$l = -\sum_i \log p(y_i|\Omega_{y_i}).$$

3.3. Testing

In testing, the blind-spot CNN processes the SAR image to estimate α_{x_i} and β_{x_i} for each pixel. The despeckled image is then obtained through the MMSE estimator, i.e., the expected value of the posterior distribution in Eq. (1):

$$\hat{x}_i = \mathbb{E}[x_i|y_i, \Omega_{y_i}] = \frac{\beta_{x_i} + Ly_i}{L + \alpha_{x_i} - 1}.$$

Table 1. Synthetic images - PSNR (dB)

Image	PPB [18]	SAR-BM3D [7]	SAR-CNN [13]	Proposed
Cameraman	23.02	24.76	26.15	25.90
House	25.51	27.55	28.60	27.96
Peppers	23.85	24.92	26.02	25.99
Starfish	21.13	22.71	23.37	23.32
Butterfly	22.76	24.48	26.05	25.82
Airplane	21.22	22.71	23.93	23.67
Parrot	21.88	24.17	25.92	25.44
Lena	26.64	27.85	28.70	28.54
Barbara	24.08	25.37	24.70	24.36
Boat	24.22	25.43	26.05	26.02
Average	*23.43*	*24.99*	*25.95*	*25.70*

Table 2. Quantitative results on SAR real images

Metrics	PPB [18]	SAR-BM3D [7]	SAR-CNN [13]	Proposed
μ_r	0.9103	0.9398	0.9845	1.0271
σ_r	0.8715	0.6834	0.8458	0.9837
ENL	44.56	22.80	29.98	8.91

Notice that this estimator combines both the per-pixel prior estimated by the CNN and the noisy realization.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section we describe the results of our method through a two-step validation analysis. First, we train and test the network on a synthetic dataset where the availability of ground truth images allows to compute objective performance metrics. We compare our method with the following despeckling algorithms: PPB [18], SAR-BM3D [7] and SAR-CNN [13]. This allows to understand the denoising capability of our self-supervised method in comparison with both traditional methods and a CNN-based one with supervised training. In the second experiment, training is conducted directly on real SAR images. To compare the despeckling methods, we rely on some no-reference performance metrics such as equivalent number of looks (ENL), and moments of the ratio image (μ_r, σ_r), and on visual inspection.

The network architecture we use in the experiments is composed of four branches with shared parameters (handling the four directions of the blind-spot receptive field, see [16]) in a first part with 17 blocks composed of 2D convolution with 3×3 kernel, batch normalization and Leaky ReLU nonlinearity. After that, the branches are merged with a series of three 1×1 convolutions.

4.1. Synthetic dataset

In this experiment we employ natural images to construct a synthetic SAR-like dataset. Pairs of noisy and clean images are built by generating speckle to simulate a single-look intensity image ($L = 1$). During training patches are extracted from 450 different images of the Berkeley Segmentation Dataset (BSD) [19]. The network has been trained for around 400 epochs with a batch size of 16 and learning rate equal to 10^{-5} with the Adam optimizer. Table 1 shows performance results on a set of well-known testing images in terms of PSNR. It can be noticed that our self-supervised method outperforms PPB and SAR-BM3D. Moreover, it is interesting to notice that while the proposed approach does not use the clean data for training, it achieves comparable results with respect to the supervised SAR-CNN method. Fig. 2 shows that also from a qualitative perspective. Despite the absence of the true clean images during training, our method produces images as visually pleasing as those produced by SAR-CNN with comparable edge-preservation capabilities.

4.2. TerraSAR-X dataset

In this experiment we employ single-look TerraSAR-X images[1]. Most of the despeckling works in literature assume the multiplicative speckle noise to be a white process. However, the transfer function of SAR acquisition systems can introduce a statistical correlation across pixels. One of the assumption for the blind-spot network training to work is that the noise has to be pixel-wise independent so that the network cannot predict the noise component from the receptive field. Hence, both training and testing images are pre-processed through a blind speckle decorrelator [20] to whiten them. During training patches are extracted from 16000 256×256 whitened SAR images. The network has been trained for around 100 epochs with a batch size of 16 and learning rate of 10^{-5} with the Adam optimizer.

Table 2 and Fig. 3 show the results obtained on three 1000×1000 test images disjoint from the training ones. ENL is computed over manually-selected homogeneous areas. It can be noticed that the proposed method is very close to the desired statistics of the ratio image, showing that indeed it removes a significant noise component, and that it better preserves edges and fine textures. It also does not hallucinate artifacts over homogeneous regions, while SAR-CNN tends to oversmooth and produce cartoon-like edges. However, the degree of smoothing over homogeneous areas is somewhat limited as confirmed by the ENL values and deserves further investigation. We conjecture that residual spatial correlation in the speckle may affect the network on real images, since excellent performance is observed on synthetic speckle.

5. CONCLUSION

In this paper we introduced the first self-supervised deep learning SAR despeckling method which only requires real single look complex images. Learning directly from the true SAR data rather than simulated imagery avoids transfering between domains for improved fidelity.

6. REFERENCES

[1] Jong-Sen Lee, "Speckle analysis and smoothing of synthetic aperture radar images," *Computer Graphics and Image Processing*, vol. 17, no. 1, pp. 24 – 32, 1981.

[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 Transactions on*

[1] https://tpm-ds.eo.esa.int/oads/access/collection/TerraSAR-X/tree

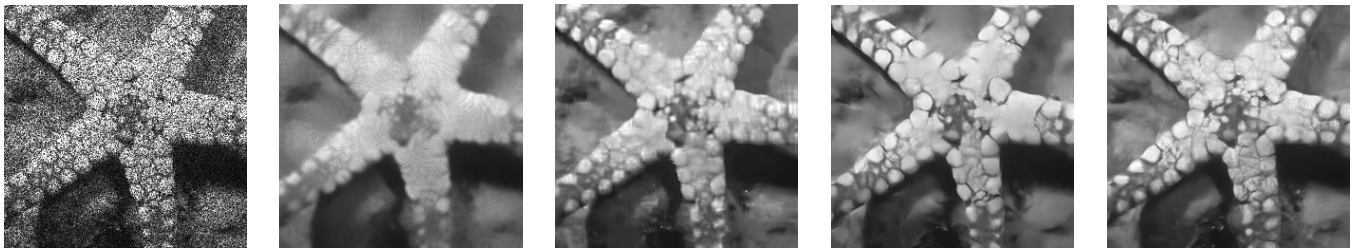

Fig. 2. Synthetic images: Noisy, PPB (21.13 dB), SAR-BM3D (22.71 dB), SAR-CNN (23.37 dB), our method (23.32 dB).

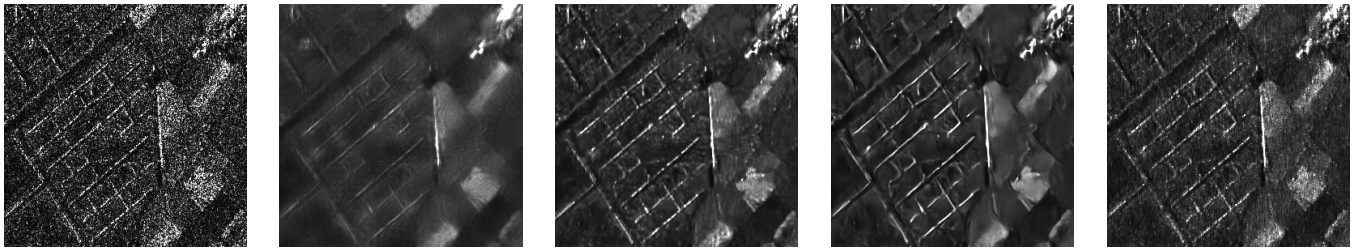

Fig. 3. Real SAR images: Noisy, PPB, SAR-BM3D, SAR-CNN, our method.

Pattern Analysis and Machine Intelligence, vol. PAMI-4, no. 2, pp. 157–166, March 1982.

[3] D. Kuan, A. Sawchuk, T. Strand, and P. Chavel, "Adaptive restoration of images with speckle," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 35, no. 3, pp. 373–383, March 1987.

[4] A. Lopes, E. Nezry, R. Touzi, and H. Laur, "Structure detection and statistical adaptive speckle filtering in SAR images," *International Journal of Remote Sensing*, vol. 14, no. 9, pp. 1735–1758, 1993.

[5] Hua Xie, L. E. Pierce, and F. T. Ulaby, "SAR speckle reduction using wavelet denoising and Markov random field modeling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 40, no. 10, pp. 2196–2212, Oct 2002.

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

[7] 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, Feb 2012.

[8] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Image denoising by sparse 3-D transform-domain collaborative filtering," *IEEE Transactions on Image Processing*, vol. 16, no. 8, pp. 2080–2095, Aug 2007.

[9] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, July 2017.

[10] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "DeepSUM: Deep Neural Network for Super-Resolution of Unregistered Multitemporal Images," *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–13, 2019.

[11] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2015, pp. 3431–3440.

[12] P. Wang, H. Zhang, and V. M. Patel, "SAR Image Despeckling Using a Convolutional Neural Network," *IEEE Signal Processing Letters*, vol. 24, no. 12, pp. 1763–1767, Dec 2017.

[13] 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)*, July 2017, pp. 5438–5441.

[14] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning image restoration without clean data," in *Proceedings of the 35th International Conference on Machine Learning*. 2018, Proceedings of Machine Learning Research, pp. 2965–2974, PMLR.

[15] A. Krull, T-O. Buchholz, and F. Jug, "Noise2Void - Learning Denoising from Single Noisy Images," in *CVPR*, 2018.

[16] 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. 6968–6978.

[17] A. C. Frery, H. . Muller, C. C. F. Yanasse, and S. J. S. Sant'Anna, "A model for extremely heterogeneous clutter," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 3, pp. 648–659, May 1997.

[18] C. 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, Dec 2009.

[19] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. 8th Int'l Conf. Computer Vision*, July 2001, vol. 2, pp. 416–423.

[20] A. Lapini, T. Bianchi, F. Argenti, and L. Alparone, "Blind speckle decorrelation for SAR image despeckling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 2, pp. 1044–1058, Feb 2014.