Enhanced Deep Learning SAR Despeckling Networks Based on SAR Assessing Metrics

Sergio Vitale, *Member, IEEE*, Giampaolo Ferraioli, *Senior Member, IEEE*, Vito Pascazio, *Fellow, IEEE*, and Luis Gomez Deniz, *Senior Member, IEEE*

***Abstract*—The proposal of deep learning (DL) solutions for synthetic aperture radar (SAR) image despeckling has recently widespread. Such solutions have been mainly designed from a DL perspective by leveraging the training and validation stage on the use of typical norm-based cost functions. For going beyond the DL perspective, in this letter, we propose an SAR-based validation stage by using SAR assessing metrics in the design and hyperparameter selection of neural networks. In the first phase, SAR assessing metrics may be used only as validation metrics to highlight critical issues that cannot be spotted with standard image-processing quality metrics. In a second phase, the same SAR assessing metrics may be used directly for enhancing the DL solution by addressing specific issues that arose during the previous SAR-based validation stage. To this aim, three different DL SAR despeckling solutions and four different SAR assessing metrics have been considered. The outcome of this analysis shows the importance of including SAR knowledge in the training and validation stages of the design of a DL solution for SAR image despeckling.**

***Index Terms*—Assessment, convolutional neural networks (CNNs), deep learning (DL), despeckling, image restoration, synthetic aperture radar (SAR).**

I. Introduction

SYNTHETIC aperture radar (SAR) imaging is a powerful and fundamental method for Earth observation. The coherent nature of SAR acquisition systems leads to the generation of complex images affected by a noise called speckle impairing their interpretation and hindering the performance of SAR-based applications [1]. To solve such issues, several methods for despeckling have been proposed in the last decades, spanning from local to nonlocal (NL) approaches [2] till the increasing focus on deep learning (DL)-based solutions [3]. Numerous solutions have been proposed in the last years, exploiting various key aspects and offering increasingly sophisticated methodologies. Indeed, the definition of DL solutions for SAR image despeckling has evolved through the design of different architectures, from simple convolutional neural networks (CNNs) [4] to complex NL-based architectures [5] or multiscaling autoencoders [6], the construction of different training approaches, from supervised approaches using both synthetic data and real multitemporal data [4], [7], [8] to unsupervised and self-supervised strategies leveraging on real data only [6], [9], and the definition of different cost functions, from single Euclidean norms to multiobjective ones [10], [11].

Commonly, the design, training, and hyperparameter settings of all DL-based methods pass through the observation of cost functions and validation metrics during the training. The observation of validation metrics is of crucial importance for the definition of DL solution. As a matter of fact, the training of DL solution requires a validation stage where the network is tested on a disjoint subset, and validation performance is monitored for assessing the behavior of the defined solution on data different from trained ones. This procedure indicates the generalization ability of the proposed solution and provides hints for its improvement. In this letter, the use of SAR-based metrics is proposed as validation metrics to be used in the validation stage of the DL method definition. The aim is to provide a SAR relevance to the validation stage going beyond the simple validation assessment and providing a wider view and more relevant information of the specific method. The idea is to select and observe the evolution of the performance of SAR-based metrics on the validation dataset with a twofold scope: 1) providing SAR-specific evaluation of the training of a specific DL method and 2) exploiting these outcomes for enhancing the defined DL solution by including the SAR-based metrics itself. The rest of the letter is organized as follows. Section II describes the proposed validation assessment. Section III reports the results and discussion of the proposed analysis. Section IV presents the conclusion.

II. Methodology

The definition of the DL method requires a training stage and a validation stage performed on two disjoint subsets. The validation stage is responsible for assessing performances during the training and the hyperparameter selection (learning rate, epochs, stopping strategy, etc.). In this letter, we propose a validation stage relying on SAR assessing metrics to provide

Received 7 April 2025; accepted 5 May 2025. Date of publication 9 June 2025; date of current version 18 June 2025. *(Corresponding author: Sergio Vitale.)*

Sergio Vitale and Vito Pascazio are with the Dipartimento di Ingegneria, University of Naples Parthenope, 80143 Naples, Italy, and also with the National Inter-University Consortium for Telecommunications, 43124 Parma, Italy (e-mail: sergio.vitale@uniparthenope.it; vito.pascazio@uniparthenope.it).

Giampaolo Ferraioli is with the Dipartimento di Scienze e Tecnologie, University of Naples Parthenope, 80143 Naples, Italy, and also with the National Inter-University Consortium for Telecommunications, 43124 Parma, Italy (e-mail: giampaolo.ferraioli@uniparthenope.it).

Luis Gomez Deniz is with the Centro de Tecnologías de la Imagen (CTIM), Departamento de Ingeniería Electrónica y Automática, University of Las Palmas de Gran Canaria, 35001 Las Palmas de Gran Canaria, Spain (e-mail: luis.gomez@ulpgc.es).

Digital Object Identifier 10.1109/LGRS.2025.3577907

SAR insights into the validation outcome and, extracting useful information for the improvement and enhancement of the specific DL method. In particular, we propose the selection of SAR-based metrics to be evaluated during the validation stage to have a SAR-based perspective of the training procedure.

As a matter of fact, most of the DL methods for SAR image despeckling are trained by using Euclidean norms as the cost function, thanks to their robustness and stability in the optimization process but exclude any SAR physical meaning and do not allow any insight from SAR perspectives. Thus, we propose to select well-known SAR-based metrics to use as validation metrics for evaluating the performances from a SAR perspective.

The lack of real ground truth in the SAR despeckling framework has given rise to different training strategies that can be broadly categorized into two classes: supervised and unsupervised training. This motivates the choice toward no-reference-based metrics. Specifically, four SAR-based metrics, relying on the multiplicative model of the speckle and the fully developed hypothesis, widely used for the SAR science community have been considered [10], [12]. Nevertheless, such a procedure may be enlarged to other ones. The multiplicative model of the speckle leads to the definition of the intensity of the SAR image given in

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

where Y is the intensity of the SAR image, N is the speckle noise, and X is the noise-free texture. According to the fully developed hypothesis, the scatterers within a resolution are independent and identically distributed, and, thus, the speckle results in Gamma distributed with unitary mean and variance $1/L$, where L is the number of looks [13].

The chosen metrics are: equivalent number of looks (ENLs), mean of image (MoI), mean of ratio (MoR), and variation of ratio (VoR).

1) *ENL:* It estimates the number of single-look images that should be incoherently averaged to obtain equivalent despeckling performances and it is defined as the ratio between the square of the mean and the variance of the filtered image

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

The higher the ENL is, the better the filter.
2) *MoR:* It estimates the mean of the ratio $\widehat{N} = Y/\widehat{X}$ that, in the case of an ideal filter, should be unitary.
3) *VoR:* It estimates the variance of the ratio $\widehat{N} = Y/\widehat{X}$ that, in the case of an ideal filter, should be unitary.
4) *MoI:* It estimates the mean of the filtered image $\widehat{X}$ that, due to the unitary mean of the noise, an ideal filter should preserve as the mean of the unfiltered image Y. The ideal filter produces MoI = 0.

As for a generic numerical assessment, it is necessary to have a joint observation of the metric to gather all the information that can be retrieved from their individual inspection. In this case, the ENL evaluates the noise suppression ability of the filter, while the other evaluates the radiometric and noise statistical preservation ability. The joint evaluation of these metrics is almost informative for the filter quality. For example, a higher ENL that comes at the cost of poorer performance on radiometric preservation indicates an attitude to over-smoothing. At the same time, good performance on radiometric preservation that comes at the cost of poor ENL indicates a tendency toward under-smoothing.

Thus, we propose showing performances in terms of SAR assessing metrics on the validation set during the training to provide a more exhaustive meaning of the validation stage. Using the above SAR-based metrics as validation accuracy metrics allows for the evaluation of training performances from an SAR perspective and, consequently, analyzes potential critical issues to be solved (i.e., overfitting, instability with respect to some metrics, etc.). As a matter of fact, it is worth noting that, in an initial phase, the selected SAR-based metrics are not involved in the learning process, but they will act only as evaluation indicators of the filtering evolution of the considered DL method. Upon the observation over the training epochs of such indicators, the same SAR assessing metrics may serve as an additional cost function to enhance the designed DL solution. The PyTorch implementation of the considered metrics is available at https://github.com/impress-parthenope

III. Experimental Part

In this section, the SAR assessing metrics described in Section II are used as validation metrics for three well-assessed state-of-the-art DL solutions proposed for SAR image despeckling. In particular, the methods SAR-CNN [4], DeSpeckNet [11], and MONet [8] have been trained on the same dataset, built following the hybrid approach [7]. In the first phase, the three solutions have been trained using their original cost functions and the metrics have been evaluated during the training as validation metrics. According to the results from the observation of this validation performance, the assessing metrics have been included in the training cost function itself to improve specific critical aspects. For a fair analysis, the three solutions have been trained on the same dataset. A stack of 27 real SAR images acquired from TerraSAR-X over the area of Barcelona (Spain) has been considered for the construction of the dataset. The dataset is composed of 100096 training patches, while the validation set is composed of 10222 patches spatially disjoint from the training ones.

A. SAR Assessing Metrics in the Validation Stage

Generally, the validation assessment during the training of DL-based methods passes through the observation of the designed cost function on the validation dataset. Since the usual cost functions are Euclidean-norm based, such validation assessment gives only an indication of the training quality useful for hyperparameter selection. For example, in Fig. 1, the trend of Euclidean norms, L_2 on the left and L_1 on the right, on the validation dataset for the three DL methods is shown. Beyond the acceptable improvement of the cost functions and to some consideration on their convergence, both prove a good training stage. However, nothing more can be concluded.

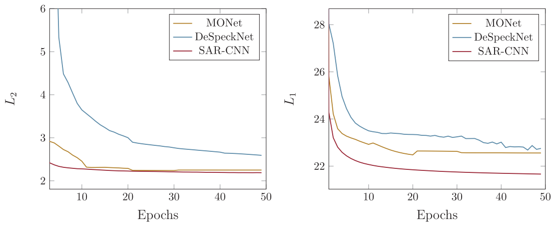

Fig. 1. Euclidean norms computed on the validation set during the training for MONet, SARCNN, and SARDRN. L_2-norm on the left; L_1-norm on the right.

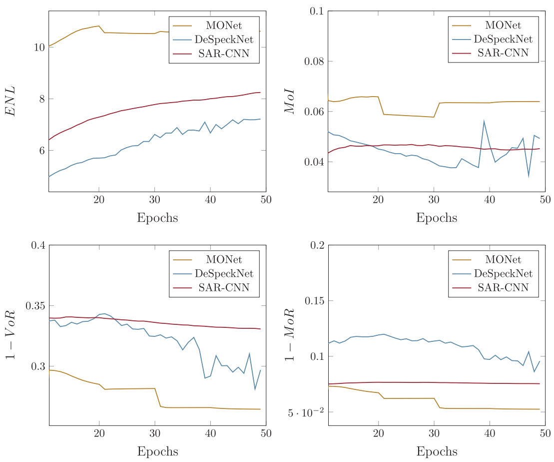

Fig. 2. SAR assessing metrics computed on the validation set during the training for MONet, SARCNN, and DeSpeckNet.

Therefore, it is clear that involving SAR assessing metrics in the validation stage is more informative. The SAR-based metrics performance evaluated on the validation dataset is presented in Fig. 2.

For the sake of visualization, 1-MoR and 1-VoR replace the MoR and VoR, respectively, and the L_1-norm between the MoI and the expected value, normalized to its energy, replaces the simple MoI. Therefore, the closer to zero are the normalized MoI, 1-MoR, and 1-VoR, the better the filter is. These plots allow us to evaluate the average performance of the methods, epoch by epoch, providing a view of performance from different SAR perspectives; moreover, they also indicate the robustness and generalization ability of the networks. As a matter of fact, by simply observing the trend of the Euclidean norms in Fig. 1, the three DL methods show a smooth and gradual improvement during the training. Instead, the observation of the performance of SAR assessing metrics in Fig. 2 provides richer information, highlighting pros and drawbacks. For each method, indeed, some metrics show improvements and others highlight unstable behavior (i.e., overfitting and instability). MoI, VoR, and MoR results are almost constant during the training for SAR-CNN, that is, the training has no impact on them. The DeSpeckNet tends to improve on all the metrics even its behavior on MoI, VoR, and MoR results almost unstable in the last epochs. The MONet shows the best ENL and ratio-based metrics, but the worst MoI. These aspects may show weaknesses and, therefore, suggest new strategies for either limiting or overcoming them. It is worth noting that knowledge could not have been spotted by just observing the L_1-norm validation loss.

TABLE I

Numerical Assessment. Averaged Values of SAR Metrics (Ideal Value) Carried Out Over 100 Patches for Three Different Sensors (Best Values in Bold, Second-Best Underlined)

Sensor	Method	ENL(inf)	MoI (0)	MoR (1)	VoR(1)
COSMO-SkyMed	MONet	11	0.040	0.992	0.612
	MONet-MoI	11	**0.031**	0.975	0.60
	MONet-MoR	11	0.45	**0.995**	**0.62**
	DeSpeckNet	**10**	0.031	**0.96**	**0.59**
	DeSpeckNet-MoI	9	**0.017**	0.93	0.54
	DeSpeckNet-MoR	9	0.017	0.94	0.54
	SAR-CNN	**7**	**0.02**	0.93	0.5
	SAR-CNN-MoI	6	**0.008**	0.91	0.49
	SAR-CNN-MoR	**7**	0.05	**1.01**	**0.65**
RADARSAT-2	MONet	22	0.02	0.95	0.63
	MONet-MoI	22	**0.01**	0.94	0.61
	MONet-MoR	**24**	0.03	**0.97**	**0.67**
	DeSpeckNet	**14**	0.02	0.89	0.54
	DeSpeckNet-MoI	12	0.01	0.91	0.55
	DeSpeckNet-MoR	12	**0.007**	**0.93**	**0.58**
	SAR-CNN	**13**	0.02	0.93	0.57
	SAR-CNN-MoI	11	**0.01**	0.93	0.55
	SAR-CNN-MoR	11	0.04	**1.01**	**0.71**

B. SAR Assessing Metrics in the Training Stage

The previous observation of SAR evaluating metrics during the validation phase reveals interesting insights that offer more precise information on the DL method behavior in learning speckle removal. Leveraging this information can be of fundamental importance for improving the DL-designed filter.

Indeed, the analysis reported in Fig. 3 suggests that the SAR assessing metrics used just for evaluation could be used as a cost function for either limiting or improving each method on a specific issue. To this aim, the three DL methods have been retrained by adding one of the SAR assessing metrics to the original cost function. In particular, besides the original training, the MoI or the MoR metrics have been added to the original cost function for the training of each method. The resulting SAR assessing metrics performance evaluated on the validation dataset for training with the original cost function (yellow), with the original + MoI (blue line), and with original + MoR (red line) have been reported in Fig. 3.

Observing the MONet (first row in Fig. 3), it is evident how the introduction of the MoI as a cost function leads to improvement in the MoI itself, as well as the use of the MoR tends to improve the ratio metrics.

Similar considerations can be applied to SAR-CNN. It is interesting to note how the ratio metrics for SAR-CNN have strongly improved with the use of MoR in the cost function. It looks like the original training has no effect on them and, therefore, using a specific ratio-based metric pushes the network in that direction. This is not the case for the DeSpeckNet. Its unstable behavior on the MoI is improved by introducing the MoI as a cost function: it tends to make all the SAR performance more gradual, improving the MOI itself but slightly worsening the others. Instead, the introduction of the MoR and VoR as a cost function does not lead to an

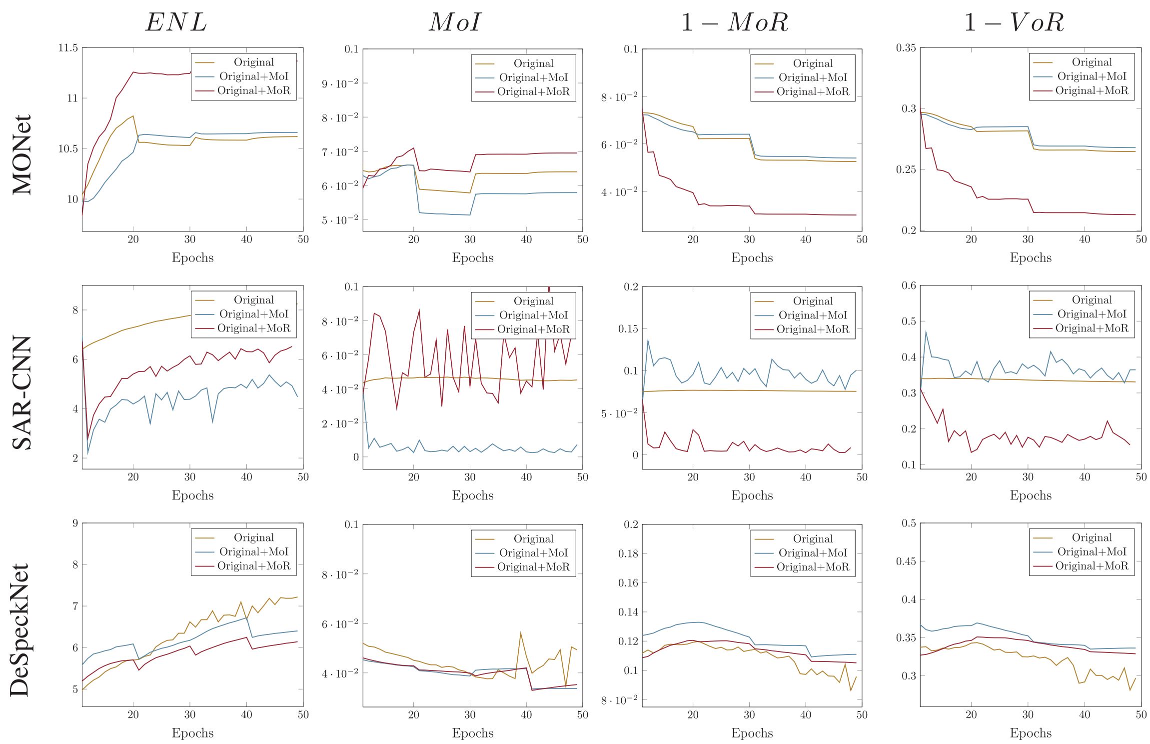

Fig. 3. SAR assessing metrics computed on the validation set during the training for original (yellow), Moi version (blue), and MoR version (red) of the MONet, SARCNN, and DeSpeckNet.

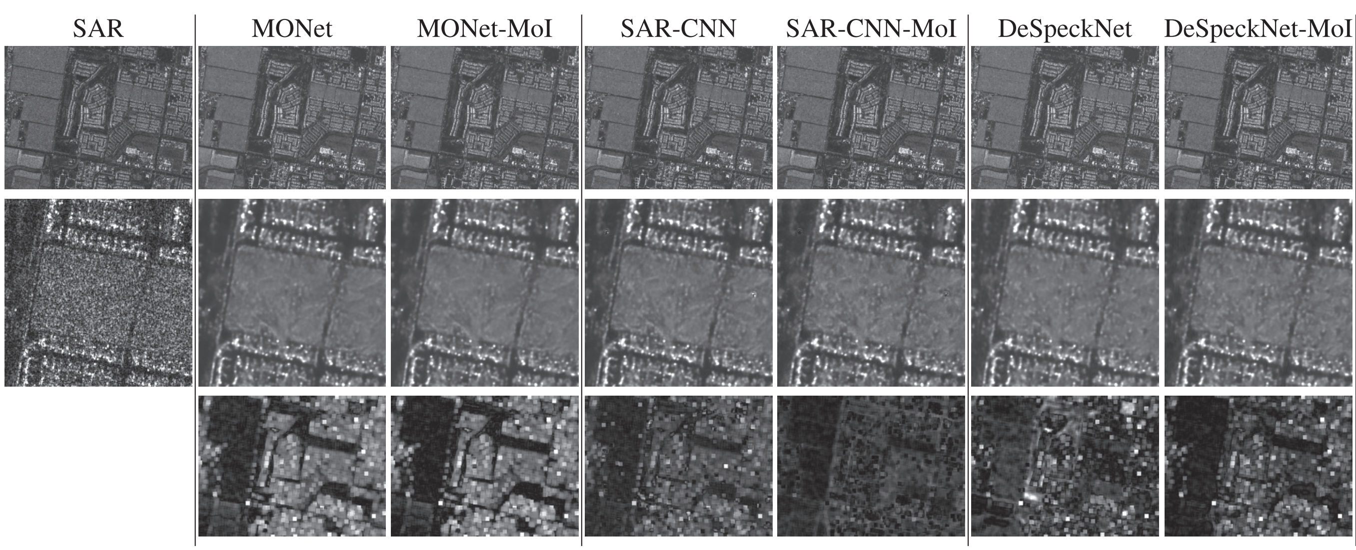

Fig. 4. Results on the RADARSAT-2 image. The SAR image is in the top left. Filtered results of the original version and MoI version of the MONet, SAR-CNN, and DeSpeckNet in the first row. Close up on the homogeneous area in the second row. Normalized difference between MoI of the SAR image and filtered result on the last row. The ideal difference should be zero (set as black).

improvement, which implies that the effect of original training is already effective.

An interesting outcome comes out from Table I, where the numerical assessment of the three DL methods and their three training variations on two real datasets is reported. In particular, the SAR-based metrics are evaluated and averaged on 100 homogeneous patches extracted from a RADARSAT-2 and COSMO-SkyMed image, respectively. This numerical assessment follows quite faithfully the results shown in Fig. 3. In most cases, each method improves on the SAR metric specifically used in the training. Observing the MoI, in almost all cases, all the methods trained with the MoI achieve the best performance. A similar conclusion can be said for ratio-based metrics, which achieved the best results when methods were trained with MoR.

To give a visual interpretation, some results of the effect of the introduction of the MoI as a cost function on the two real datasets, RADARSAT-2 and COSMO-SkyMed, are

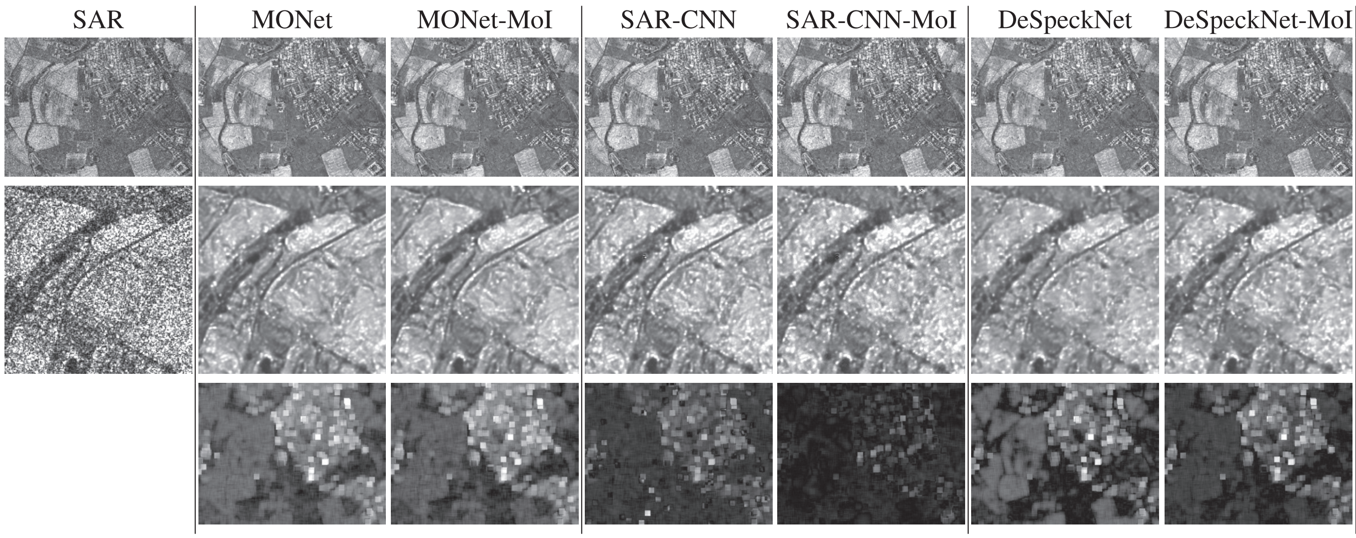

Fig. 5. Results on the COSMO-SkyMed image. The SAR image is in the top left. Filtered results of the original version and the MoI version of the MONet, SAR-CNN, and DeSpeckNet in the first row. Close up on the homogeneous area in the second row. Normalized difference between the MoI of the SAR image and filtered result on the last row. The ideal difference should be zero (set as black).

depicted in Figs. 4 and 5, respectively. In the first row, the SAR image and the results of each method and its MoI version are provided. In the second row, a close-up of a homogeneous area is presented. In the last row, the normalized difference between the MoI computed on the SAR image and the MoI computed on the filtered image is reported. In both cases, the difference between the visual results of the original method and its MoI version is barely visible by observing the results either on the whole area (first row) or on the zoom on the homogeneous one (second row). Only by observing the last row, the impact of the MoI is evident (it is worth reminding that in homogeneous areas, the difference between the MoI of the noisy image and the MoI of the ideal filter should be zero). The normalized difference between MoIs makes evident the effect of the introduction of the MoI as a cost function: the MoI performance is systematically improved on homogeneous areas (as expected) for MONet, SAR-CNN, and DeSpeckNet.

From the numerical and visual assessment, it is clear that the inclusion of the SAR metrics in the training provides a different perspective with respect to the classical cost functions.

IV. Conclusion

In this work, the involvement of SAR assessing metrics in the validation and definition of DL solutions for SAR image despeckling is exploited. The aim of including SAR assessing metrics in the validation, first, and in the training, later, is to provide more detailed insights specific to the SAR despeckling issue of the DL solution itself. Four SAR assessing metrics and three supervised DL solutions have been considered. No-reference metrics have been selected to easily extend such study to unsupervised methods. The experimental results show that using the SAR assessing metrics as a validation metric allows us to identify critical issues in the training from an SAR despeckling perspective that could not be spotted with classical metrics. Moreover, the results show that directly including the SAR assessing metrics may help in overcoming or limiting the spotted issues. The work outlines new research paths to the intrinsic factors contributing to such improvements in terms of model architecture, training approach, or data characteristics which could be addressed in the future.

References

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

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

[3] G. Fracastoro, E. Magli, G. Poggi, G. Scarpa, D. Valsesia, and L. Verdoliva, "Deep learning methods for synthetic aperture radar image despeckling: An overview of trends and perspectives," *IEEE Geosci. Remote Sens. Mag.*, vol. 9, no. 2, pp. 29–51, Jun. 2021.

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

[5] D. Cozzolino, L. Verdoliva, G. Scarpa, and G. Poggi, "Nonlocal CNN SAR image despeckling," *Remote Sens.*, vol. 12, no. 6, p. 1006, Mar. 2020. [Online]. Available: https://www.mdpi.com/2072-4292/12/6/1006

[6] E. Dalsasso, L. Denis, and F. Tupin, "As if by magic: Self-supervised training of deep despeckling networks with MERLIN," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 4704713.

[7] S. Vitale, G. Ferraioli, and V. Pascazio, "Analysis on the building of training dataset for deep learning SAR despeckling," *IEEE Geosci. Remote Sens. Lett.*, vol. 19, pp. 1–5, 2022.

[8] S. Vitale, G. Ferraioli, A. C. Frery, V. Pascazio, D.-X. Yue, and F. Xu, "SAR despeckling using multiobjective neural network trained with generic statistical samples," *IEEE Trans. Geosci. Remote Sens.*, vol. 61, 2023, Art. no. 5216812.

[9] X. Ma, C. Wang, Z. Yin, and P. Wu, "SAR image despeckling by noisy reference-based deep learning method," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 12, pp. 8807–8818, Dec. 2020.

[10] S. Vitale, G. Ferraioli, and V. Pascazio, "Multi-objective CNN-based algorithm for SAR despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 11, pp. 9336–9349, Nov. 2021.

[11] A. G. Mullissa, D. Marcos, D. Tuia, M. Herold, and J. Reiche, "DeSpeckNet: Generalizing deep learning-based SAR image despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5200315.

[12] G. Di Martino, M. Poderico, G. Poggi, D. Riccio, and L. Verdoliva, "Benchmarking framework for SAR despeckling," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1596–1615, Mar. 2014.

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