Multi-Objective CNN Based Algorithm for SAR Despeckling

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

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

Index Terms—Image Restoration, Despeckling, SAR, Statistical Distribution, CNN, Deep Learning.

I. INTRODUCTION

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

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

The first solutions work in the spatial domain, such as [11], [12], [13], [14] and [15], [16]: The first ones are based on a minimum-mean-square-error (MMSE) while the second ones propose a maximum a posteriori (MAP) filter. These methods produce intense smoothing for reducing speckle in homogeneous areas that can not be applied on the edges.

Since the early 1990s, despeckling techniques operating in a transformed domain have been proposed [8], [9], [10], [11]. Filters based on such approach often operate an homomorphic transformation in order to work with additive noise. These solutions embody a strong spatial adaptability in order to better preserve edges, which is a crucial issue in SAR despeckling.

A new research line in the despeckling domain has been drawn by the non local methods, that have shown very effective performances in preserving details, while removing noise [12]. Such methods look for similar patches in the image and merge them in order to produce targets pixels. Usually, differently from the previous solution, statistics of the speckle and of the SAR backscattering [13], [14] are taken into account for the definition of patch similarity. Several algorithms have been defined within the non local paradigm, mainly by differentiating the choice of the similarity criterion or the merging function. For example, the non local paradigm based on different SAR similarity distances is applied by the methods proposed in [15], [16]. Whereas, a ratio-based metric is used in [17], [18].

Hybrid approaches arose like [19], [20], [21] that join the non local paradigm with the wavelet transform.

A detailed review of the aforementioned despeckling filters can be found in [22], [23].

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

Recently, several DL solutions have been proposed for SAR despeckling. Such methods are data driven: differently from the previous classical approaches, it is mandatory to have a dataset composed of many couples of noisy inputs and noise-free images (references). Since for SAR despeckling, a noise-free reference is not available, the first issue for such methods is the construction of simulated dataset.

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

Moreover, instead of using synthetic approach, in other techniques such as [30], [31], the multitemporal average version of SAR acquisition serves as reference. Always for sake of simplicity, this approach is referred as multitemporal approach in the following of the paper. Most of these proposals focus only on the definition of the architecture and use very similar cost functions not taking into account statistical properties of the SAR image and the presence of strong scatterers. In [25] the mean-square-error (MSE) is used as cost function. In [24] and [26] the MSE is combined with a total variation regularization. A smoothed $L_1$ loss adapted to the speckle noise case has been considered in [30]. The first attempt to include first order statistics of the speckle was proposed in [32], whereas in [31] a cost function based on statistic similarity is used.

Moreover, most of the methods do not provide neither code or training dataset for reproducible research.

In this paper a CNN for SAR despeckling that takes into account statistical properties of the SAR image has been proposed. The network is a seventeen layer CNN with skip connection trained with the synthetic approach. Beyond the proposed architecture, the main contribution is in the definition of a multi-objective cost function given by combination of the proposed architecture and of the multi-objective cost function. Experimental results and discussion are presented in Section IV. An ablation study of the cost function has been carried out in Appendix A.

II. METHODOLOGY

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

A. Acquisition Model

The interpretation of SAR image is challenging due to the geometrical properties of SAR imaging system and to the presence of speckle. Indeed, speckle is a multiplicative noise produced by interference among the backscatterings of the objects inside a resolution cell of the sensor [33]. The generic SAR image can be expressed like in Eq. [1]

$$Y = X \cdot N \quad (1)$$

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

The statistical distribution of the speckle is well known under certain conditions. Three main cases can be considered: homogeneous, heterogeneous and extremely heterogeneous areas. Homogeneous areas (such as fields, roads, etc..) are characterized by the lack of dominant scatterers and the surface $X$ can be considered stationary.

This is the case of the Fully Developed hypothesis for the speckle $N$, whose intensity follows the Gamma distribution [33]:

$$p_N(n, L) = \frac{1}{\Gamma(L)} L^L n^{L-1} e^{-nL} \quad (2)$$

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

Heterogeneous (tree and forest) and extremely heterogeneous areas (urban areas), are characterized by objects with shape and dimension that produce geometrical distortions and strong backscattering (e.g. multiple bounces, layover and shadowing). In heterogeneous areas, the speckle can be still considered Gamma distributed but the surface is not stationary anymore. In extremely heterogeneous area, the hypothesis of distributed scatterers is not valid anymore due to the presence of dominant ones. Indeed, the speckle does not follow anymore the fully developed hypothesis [34].

Despite the statistical description of the speckle, it is interesting to consider the statistical distribution of SAR backscattering $Y$ in the three previously presented cases, provided in [14]. Frery et al. proposed the use of the square root of generalized inverse Gaussian distribution $G_A(\alpha, \gamma, \lambda, L)$ as general model for the amplitude return of SAR backscattering, proving that the distribution of the SAR return of homogeneous, heterogeneous and extremely heterogeneous areas are particular case of this distribution depending on the parameter subspace. An extension of this classification considering several possible scenarios has been recently proposed in [35].

According to [14], in the subspace $(\alpha > 0, \gamma = 0, \lambda > 0, L > 0)$, when $\alpha$ and $\lambda$ tends to infinite, the distribution tends to a square root gamma $\Gamma^{1/2}(L, L/\beta)$. In case of a single look ($L=$1), it becomes a Rayleigh distribution with parameter $\beta$ with $\beta$ being the estimation $E[Y^2]$ of the second order statistic. Such distribution describes the return from homogeneous areas.

Moreover, the authors have proved that the SAR return $Y$ in heterogeneous area follows the $K_A(\alpha, \lambda, L)$ distribution (Eq. [3]). This is the distribution the $G_A(\alpha, \gamma, \lambda, L)$ tends to, when the parameter subspace is always $(\alpha > 0, \gamma = 0, \lambda > 0, L > 0)$.

$$p_Y(y, L) = \frac{4\lambda L y}{\Gamma(\alpha)\Gamma(L)} (\lambda L y)^{(\alpha+1)-1} K_{\alpha-L}(2y\sqrt{\lambda L}) \quad (3)$$

where $K_\alpha$ denotes the modified Bessel function of third kind and order $n$.

Finally, the authors have proved that for extremely heterogeneous areas the amplitude distribution of the SAR image...
belongs to the $G_A^\alpha(\alpha, \gamma, L)$ (Eq. [3]), that is the distribution
the $G_A(\alpha, \gamma, L)$ tends to, when the parameter subspace is
$(\alpha < 0, \gamma > 0, \lambda = 0, L > 0)$,

$$p_Y(y, L) = \frac{2L^L \Gamma(L - \alpha) \gamma^{-\alpha} y^{2L-1}}{\Gamma(L) \Gamma(-\alpha) (\gamma + Ly^2)^{L-\alpha}} \quad (4)$$

### B. Data Simulation

In this section the data simulation process adopted for the
training of the proposed CNN based de-speckling algorithm.

Thousands of noise-free images from the optical UC Merced
Land Use dataset [36] have been considered. This dataset
is typically considered for classification purposes thanks to
the presence hundreds images belonging to different classes.
Samples of this dataset are shown in Fig. 1.

![Fig. 1. RGB samples of Merced Land Use dataset](image)

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

![Fig. 2. Simulation process, from left to right: noise-free reference, simulated
noise, simulated SAR image](image)

From the whole dataset, $57526 \times 64 \times 64$ amplitude patches
for the training and $14336 \times 64 \times 64$ for the validation have
been extracted.

### C. Network Architecture

The design of the proposed network architecture comes from
the results achieved in our previous works where, first a
ten layers CNN [32] and later a seventeen layers CNN [37]
have been proposed.

Starting from the result of [37], the proposed neural network
is composed of seventeen convolutional layers. For each layer
we consider ReLU as activation function [38], but for the last.
In all the layers batch normalization [39] is performed except
for the first and the last ones. In addition, skip connections,
that have shown great utility in training deep networks [40],
are introduced in the inner layers.

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

$$z_k = f_k(\Phi_k, z_{k-1}) = \begin{cases}
\sigma(w_k * Y + b_k) & \text{k=1} \\
BN [\sigma (w_k * z_{k-1} + b_k)] + \alpha \cdot f_{k-3}(\Phi_{k-3}, z_{k-4}) & 1 < k < D \\
(w_k * z_{k-1} + b_k) & \text{k=D}
\end{cases} \quad (5)$$

where

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

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

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

For each layer 64 features maps are extracted except for
the last one that has to fit the single channel output. All the
convolitional kernel have dimension $3 \times 3$. In Fig. 3 a scheme
of the network is depicted. The scheme of residual block is
defined in Fig. 3.

### D. Cost Function

The cost function adopted in the proposed algorithm is a
linear combination of three terms, each of them specifically
dedicated to catch and to preserve information from the SAR
image.

The aim is to propose a cost function that take care both
of spatial and statistical properties of the SAR images. The
defined cost function is a combination of three terms (Eq. [6]):
$L_2$ is the mean square error between the reference $X$
filterred image $\tilde{X}$; $L_{KL}$ is the Kullback-Leibler divergence
between the distribution of estimated noise $\tilde{N} = Y/\tilde{X}$
and that of the theoretical one $N_{teo}$; $L_\nabla$ is the MSE between the gradient of the reference $X$
and gradient of the filtered image $\tilde{X}$.

$$L = L_2 + \lambda_{KL} L_{KL} + \lambda_\nabla L_\nabla \quad (6)$$

$$L_2 = L_{MSE} = ||\tilde{X} - X||^2$$

$$L_{KL} = D_{KL}(\tilde{N}, N_{teo})$$

$$L_\nabla = ||\nabla X - \nabla \tilde{X}||^2$$

Let us consider each of the three terms separately.

Naturally, the goal is to train the network to generate an
output as similar as possible to the reference. To this aim the
$L_2$ term directly compares the output $\tilde{X}$ with the reference $X$
and it is responsible of spatial reconstruction.

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

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

The introduction of the $\mathcal{L}_\nabla$ term is two fold: improving the edge preservation \cite{31} and dealing with dominant scatterers in real images. $\mathcal{L}_\nabla$ compares the gradients of $\hat{X}$ with the gradient of $X$. The gradient gives information on the edges but, obviously, is not exactly an edge detector. It highlights transitions in images and so tends to identify the presence of structures. So, if from one side it trains the network in preserving edges, on the other it helps the network in identifying and isolating strong scatterers.

Summing up, the proposed cost function is a linear combination of three terms and both estimated noise-free image $\hat{X}$ and estimated ratio image $\hat{N}$ are involved:

- $\mathcal{L}_2$: this term is responsible of spatial details preservation by directly comparing $X$ and $\hat{X}$.
- $\mathcal{L}_{KL}$: this term is responsible of statistical properties preservation by comparison of statistical distribution of ratio image and theoretical noise.
- $\mathcal{L}_\nabla$: this term is responsible of edge and dominant scatterers preservation by comparing the gradient along the horizontal and vertical direction of $\hat{X}$ and $X$.

In Appendix \ref{ablation} an ablation study on the effects of these three terms has been proposed.

\section{E. Contribution}

In this section, the contribution of the proposed method is described and innovative issues are highlighted. The proposed solution shares some points with IDCNN \cite{24}, SAR-CNN \cite{30} and SAR-DRN \cite{25}. Indeed, the proposed CNN has seventeen layers like SAR-CNN, and also skip connections are added in the inner layers, like in SAR-DRN. Differently from SAR-DRN, a deeper network has been preferred to dilated convolutions. Deeper networks allow to extract more features and to add more abstractions, facilitating the exploitation of the data and the network generalization. The depth has been set experimentally: in \cite{42} it has been proved that deeper network gives better results.

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

\begin{table}[h]
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
Method & Depth & Skip Connection & Spatial Loss & Statistical Loss \\
\hline
ID-CNN & 10 & $\times$ & \& & $\times$ \\
SAR-CNN & 17 & $\times$ & smoothed $\mathcal{L}_3$ & $\times$ \\
SAR-DRN & 7 & $\checkmark$ & $\mathcal{L}_2$ & $\times$ \\
Proposed & 17 & $\checkmark$ & $\mathcal{L}_2 + \mathcal{L}_\nabla$ & $\mathcal{L}_{KL}$ \\
\hline
\end{tabular}
\caption{Main differences among compared DL methods}
\end{table}

\section{III. Experimental Results}

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

For comparison, two different families of despeckling algorithms have been considered: Non Local and Deep Learning based ones. In particular, NL algorithms have been addressed since they are often considered in literature as a benchmark for evaluating achievable performances. Between the available
NL algorithms we considered FANS [21], SAR-BM3D [19] and NOLAND [18].

While the DL based algorithms have been considered in order to compare the performances of the proposed algorithm with methods sharing the same philosophy. In particular, ID-CNN and SAR-DRN as deep learning methods have been used. Given that the DL solutions are data driven, in order to have a fair comparison we re-train from scratch the CNN based solutions on our same dataset following the description of the authors. For this reason, we decide to not compare with SAR-CNN and the proposal of Cozzolino et al in [31]: in fact a fair comparison is not possible because they trained with multitemporal approach and re-training them on our simulated data is reductive. Moreover, as for the rest of DL papers, the authors did not make available their code and training dataset.

The proposed network is trained with mini batch of 128 samples, using the Adam optimizer [43] with parameter $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The learning rate is set to $\eta = 0.0001$ for the first 87 epochs, and after the training is refined for other 35 epochs with a learning rate scaled by 10. The lambdas parameter for the cost function have been set for balancing their effects: $\lambda_{KL} = 10^4$ and $\lambda_{KL} = 1$

A. Metrics

For numerical evaluation both reference and no-reference metrics have been considered. As reference metrics the Structural Similarity (SSIM) index, the Mean Squared Error (MSE) and the Signal to Noise Ratio (SNR) have been used for evaluating results on the simulated dataset, where a reference is available.

- SSIM measures the similarity between $\hat{X}$ and $X$ from a perceptual point of view. The ideal filter would produce SSIM=1
- MSE measures the average similarity between $\hat{X}$ and $X$. The ideal value is zero.
- SNR measures the signal to noise ratio and give us information about the capability of the noise suppression. The higher SNR, better the filter.

Regarding no reference metrics the M-index, the Haralick homogeneity $\delta h$, the residual ENL $r_{ENL}$, the mean of the ratio $\mu_N$ and the Kullback Leibler divergence $D_{KL}$ are considered.

- M-index [44] is a combination of three factors $\delta h$, $r_{ENL}$ and $\mu_N$:
  - $\delta h$ is based on the Haralick homogeneity texture [45] and it is the distance between the homogeneity $h_0$ of ratio image compared with the homogeneity $h_0$ of the random permuted the ratio image itself. It is compute as $\delta h = \frac{|h_0 - h_0^*|}{h_0}$, with $h_2 = \sum_i \sum_j \left( \frac{1}{1 + (i - j)^2} \cdot p_z(i, j) \right)$ where $p_z(i, j)$ is the gray scale level co-occurrence matrix of the ratio image $z$ at an arbitrary position. $\delta h$ computes a sort of correlation of the ratio image and give us information of remaining structures that should not be present after an ideal filtering. The ideal filter will produce $\delta h = 0$.
  - $r_{ENL}$ is the residual ENL and once $n$ homogeneous patches are selected the ENL computed on ratio and SAR image are compared. $r_{ENL} = \frac{1}{n} \sum_{i=1}^{n} \frac{|ENL_{noisey}(i) - ENL_{ratio}(i)|}{ENL_{noisey}(i)}$
  - $r_{\mu}$ is the function of the mean ratio $\mu_N$ computed on the same patches selected for the $r_{ENL}$ $r_{\mu} = \frac{1}{n} \sum_{i=1}^{n} |1 - \mu_N(i)|$

The ideal filter will produce $r_{ENL}$ equal to 0.

- $\delta h$ is based on the Haralick homogeneity texture [45] and it is the distance between the homogeneity $h_0$ of ratio image compared with the homogeneity $h_0$ of the random permuted the ratio image itself. It is compute as $\delta h = \frac{|h_0 - h_0^*|}{h_0}$, with $h_2 = \sum_i \sum_j \left( \frac{1}{1 + (i - j)^2} \cdot p_z(i, j) \right)$ where $p_z(i, j)$ is the gray scale level co-occurrence matrix of the ratio image $z$ at an arbitrary position. $\delta h$ computes a sort of correlation of the ratio image and give us information of remaining structures that should not be present after an ideal filtering. The ideal filter will produce $\delta h = 0$.

The ideal filter will produce $r_{\mu}$ equal to zero.

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

- $D_{KL}$ computes the distance between the statistical distribution of the ratio image with the theoretical Rayleigh distribution.

$$D_{KL}(N, \tilde{N}_{teo}) = \sum_i P_N(i) \log_2 \left( \frac{P_N(i)}{P_{N_{teo}}(i)} \right)$$

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

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

B. Simulated Results

For the simulation 100 images of size $256 \times 256$ have been selected. These belong to 5 classes (20 for each class) of the Merced Land Use dataset not used during the training phase. In Tab. [I] the numerical evaluation for reference metrics, averaged on the whole dataset, is shown. Best solution is expressed in bold, the second best is underlined.

From this evaluation, it is clear that DL solutions outperform the other given that they are trained on a dataset with same properties of the testing one.

The proposed method outperforms all the DL and NL methods. The best NL solution on simulated experiments is SAR-BM3D.

Numerical assessment is not enough and visual inspection is essential for understanding the performance of a filter. Five different images, with different textures are shown in Fig[5] for a qualitative analysis. Together with the noisy images (first column), the noise free reference images are reported. Columns from 3 to 8 show the filtering results of the different considered approaches.

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

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

### TABLE II

**Numerical Assessment on Simulated Dataset**: the value are averaged on the whole simulated testing dataset composed of 100 images

|       | SSIM | SNR  | MSE  |
|-------|------|------|------|
| FANS  | .7049| 8.0432 | .00482 |
| SAR-BM3D | .7379 | 8.4251 | .00438 |
| NOLAND| .6847 | 7.4712 | .00544 |
| ID-CNN| .7231 | 8.3644 | .00437 |
| SAR-DRN| .7437 | 8.7240 | .00406 |
| Proposed| **.7510** | **8.8555** | **.00395** |

### C. Result on Real SAR Images

Three real SAR images from three different sensors have been considered: Pavia (Italy) acquired by CosmoSKY-MED (CSK), Phoenix acquired by RADAR-SAT and Tehran (Iran) acquired by TerraSAR-X (TSX).

In Figs. [6][8] the noisy images (first column) and the results on the considered algorithms (columns 2-7) are shown. In the first row, the results on the whole image are shown. In the second row the results on a particular patch of the whole image is presented. The corresponding ratio images are in the third row. In the left bottom corner, there is a zoom of the blue square box detail in order to better spots the difference among the DL solutions. Because of the lack of a reference, it is difficult to find a metric that can evaluate fairly the filters given they rely on certain mathematical assumption for the speckle that is not sure are confirmed in the real SAR image under test. For this reason the evaluation of filtering performance mostly relies on visual inspection considering the ability of suppressing noise while preserving objects in the scene. To this aim, also the ratio images produced by each method are shown. As noticed in the simulated results, FANS has a good edges preservation but produces over-smoothed results on homogeneous areas. NOLAND better preserves spatial details than FANS, but it is still smooth. SAR-BM3D has the best edges and objects preservation among the NL filters, but the noise is still present on the filtered images. Generally, the deep learning solutions try to more suppress the noise compared with the NL approaches. The proposed solution shows a good trade-off between noise suppression and edges preservation: in homogeneous areas noise is removed without losing many spatial details. Moreover, at the same time the edges are quite well preserved. Similar considerations can be done for SAR-DRN and IDCNN but both of them produces some artefacts: the former introduces a vertical texture in all the images and produces disturbed edges, generating less clean images; the latter has a good edges preservation but less suppresses the noise with respect the other two and produce some black spots.
These considerations can be appreciated on the details and on the relative ratio images shown in Figs. [6-8].

For example, the boundaries of the road highlighted in the blue box of Fig. 6 are retained quite well from the proposed solution and homogeneous areas are reach of spatial details not deleted by the noise removal.

These spatial details barely appear in the NL approaches (except for SAR-BM3D), while edges are well defined. In the zoomed area in the left bottom corner it can be noted that IDCNN and SAR-DRN show the limitations previously mentioned.

Moving to the ratio image, it must be recalled that an ideal filter should produce an uncorrelated ratio image: more correlation, more structures are visible in the ratio, worse is the filtering effect. From the ratio images, the road is more visible for SAR-DRN respect the others: meaning it is heavily filtered and not well preserved. The ratio images of IDCNN and proposed solution are very similar each other with some emergent structure for the former. The NOLAND ratio image looks almost uncorrelated but it is characterized by a large granularity typical of generalized smoothness. Contrary, SAR-BM3D has a ratio image with a very tiny granularity typical of good object preservation but also of a not perfect noise suppression. From the FANS ratio image, it is easily observable a different behaviour of the filter in different areas: large granularity on homogeneous areas proving its over smoothing effect, and very tiny granularity in correspondence of not homogeneous areas.
Regarding the M-index, NOLAND has always the best value followed by proposed solution, except for the TSX-Tehran image, where the second best is FANS. In order to interpret these results, the three factors $\delta h$, $\hat{r}_{ENL}$ and $\mu_N$, whose M-index is a combination, have been extracted. Lower is $\delta h$, less are the remaining structure and higher is the detail preservation during the noise suppression. Lower is $\hat{r}_{ENL}$.
TABLE V  
NUMERICAL ASSESSMENT TSX - TEHRAN

|       | M-index | δh   | r kne | µ N  | D kl  |
|-------|---------|------|-------|-------|-------|
| FANS  | 21.96   | 0.0140 | 97   | 0.8583 | 0.2376 |
| SAR-BM3D | 56.37  | 0.0882 | 103  | 0.9026 | 0.3975 |
| NOLAND | 14.07   | 0.0460 | 23   | 0.8944 | 0.0193 |
| ID-CNN | 26.79   | 0.0289 | 50   | 0.8855 | 0.0760 |
| SAR-DRN | 23.96  | 0.0006 | 47   | 0.8829 | 0.1348 |
| Proposed | 24.34  | 0.0001 | 48   | 0.9006 | 0.0651 |

the ENLs computed on the ratio image are closer to the ENLs computed to the noisy, meaning a better statistical preservation of the noise. The proposed solution shows always the best or the second best value for δh confirming a better details preservation w.r.t other methods. Indeed, the other methods produce more artefacts and the ratio images highlight more structures. Reverse is the situation for the performance on the $r_{ENL}$: the proposed solution is always surpassed by NOLAND and FANS.

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

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

Regarding the $D_{KL}$, we can see that NOLAND has always the best performance. This thanks to the fact that $D_{KL}$ is included in the similarity research process. Proposed solution has the second best performance on CSK-Pavia and on TSX-Tehran, while on RADAR-SAT-Phoneix it reaches the third one. Naturally the $D_{KL}$ results are affected by presence of not homogeneous areas and so they are rather general. It is worth to notice that among the DL methods, the proposed solution has always the best $D_{KL}$ index. This means that using a statistical term as $L_{KL}$ gives the network an added useful statistical information that can not be acquired only by the data.

It is worth to notice that our network is trained under the fully developed hypothesis and the use of $L_\gamma$ aims in preserving objects, details and strong scatterers where that hypothesis is not valid anymore. These points strongly appear in the ratio images produced by the proposed method. This issue will be addressed in the next section.

D. Detection of Not Fully Developed Areas

The presence of strong scatterers is challenging for all the filters and their filtering policy is still an open issue. Some methods, such as SAR-BM3D, NOLAND and FANS, filter them by aggregation of similar patches selected trough a statistical approach; other methods, such as [46], do not filter them at all. Usually, these points are related to extremely heterogeneous areas (urban) where, as pointed out by Frery et al in [14] and Tison et al. in [34], the speckle is not fully developed anymore.

In our case, the training data are simulated under the fully developed hypothesis and so the network never meets such elements in the training dataset. As matter of fact, the Merced Land Use dataset used for the training is composed of several scenarios such as agricultural field, baseball diamonds, forest, residential areas etc. The simulation process transforms all these data in noisy images whose distribution belongs to the Rayleigh distribution typical of homogeneous areas and the K distribution typical of heterogeneous areas. The only distribution not included is the typical one for the extremely heterogeneous areas where not just the speckle is not fully developed but also geometric distortions arise.

![Image](image_url)

Fig. 9. Comparison between two different samples of training dataset: agricultural sample (magenta solid) and forest (black solid). In dashed the theoretical Rayleigh and K distribution

In Fig. 9 the distributions of two samples of the dataset are depicted. The magenta solid curve represents the distribution of a simulated image taken from the “agricultural” class of the dataset. In this case the surface $X$ is almost homogeneous and the distribution of the resulting simulated $Y$ fits the Rayleigh distribution (magenta solid). At the same time, the black solid curve represents the distribution of a simulated image taken from the “forest” class. In this case, the texture $X$ cannot be considered homogeneous but some fluctuation had to be taken into account. Indeed, the distribution fits quite well the $K_{\lambda}(\alpha, \lambda, L)$. The parameter are estimated empirically as $\alpha = 40, \lambda = 8.5$ (the parameter $\gamma = 0$ is omitted). Different heterogeneous images will fit always the $K$ distribution but with different parameters. As expected, homogeneous and heterogeneous textures $X$ combined with the fully developed speckle lead to the Rayleigh and $K$ distributions [34].

Dominant scatterers are not simulated so extremely heterogeneous scenarios are not involved in the training process.

In order to deal with such areas we introduced the $L_\gamma$ term in the cost function. This term aims to isolate and preserve strong scatterers. In fact, it helps in identifying objects, avoiding smoothing effect and preserving the edges. In fact, the only introduction of the term $L_{KL}$ trains the network in producing a ratio image whose distribution is as close as possible to the fully developed speckle. Indeed, combining only the $L_{KL}$ with $L_2$ tends to destroy the information belonging to dominant scatterers producing a strong smoothing effect on their edges (see Fig. 14). This is the reason why, in correspondence of such points high values arise in the ratio image: strong values in the ratio correspond to an excessive filtering.
So the combination of these two terms with the $L_2$ produces in the filtered image scatterers with well preserved edges, but a bit over filtered.

As a matter of fact, on these points the ratio image of an ideal filter should not show a Rayleigh. Thus, the appearance of such points on the ratio image can be considered as a positive issue. It allows to identify such points, having a different statistical distribution (i.e. it allows to automatically identify points belonging to extremely heterogeneous areas).

In order to prove that these points belong to an urban area and so the fully developed hypothesis is not valid any more, their detection directly from the ratio image has been performed. For this goal, a combination of the ratio edge detector proposed in $[4]$ and a Kolomogorov-Smirnov test on the ratio image produced by our algorithm has been applied.

The former aims to highlight the edges and remaining structures in the ratio image, the latter detects the area where the predicted speckle is not fully developed by comparison through a threshold at patch level of ratio image distribution and the fully developed one. In Fig. $10$, a flowchart of this detection process is depicted. In Fig. $11$ the detection map of not fully developed points are shown for CSK-Pavia. In Fig. $12$, it is shown how the detected points on the SAR image (SAR Extremely Heterogeneous points, SAR-EH) generate a pdf (solid magenta curve) that fits quite well the theoretical distribution of $G_{0}^{0}(a, \gamma, L)$ (dashed magenta) indicated by Frery et al. in $[14]$ as the distribution that better describes such areas. The parameters are experimentally estimated as $(a = -0.5, \gamma = 0.145, \lambda = 0, L = 1)$.

At the same time, we carried out the distribution of the SAR image in the remaining points (SAR Heterogeneous points, SAR-H) and this fits the $K_{A}(a, \lambda, L)$ distribution, meaning that all the remaining part of the image belongs to heterogeneous areas. The parameters are still set experimentally as $(a = 2, \gamma = 0, \lambda = 7.5, L = 1)$.

This confirms the fact that our CNN is able to detect the points belonging to the extremely heterogeneous areas directly from the ratio image.

Naturally, this issue is in common with all the CNN that use training data simulated under the fully developed hypothesis. So this procedure could be extended also to the other methods like ID-CNN and SAR-DRN. In Fig. $13$ a patch of CSK-Pavia is shown with relative detection for the DL methods. First of all, it is important to note the different behavior of three CNNs on strong scatterers: proposed method try to isolate the objects by preserving the edges and at the same time produce a strong structure in the ratio; contrary SAR-DRN try to less filter these elements but some distortion are visible both in the filtered image and in the ratio. ID-CNN produces many artefacts not only in correspondence of the scatterers but also in its neighborhood.

So, SAR-DRN does not allow a complete detection of strong scatterers. While IDCNN includes points that do not belong to the SAR-EH points, indeed the resulting pdf (not shown for sake of simplicity) does not fit the $G_{0}^{0}$ distribution.

IV. CONCLUSION

In this paper a CNN for SAR despeckling trained on simulated data has been proposed. The non linearity introduced by the seventeen layers are crucial for features extraction while skip connections are used for avoiding the vanishing gradient problem. Despite the proposed architecture, the main focus is dedicated to the definition of a multi-objective cost function composed of three terms: $L_2, L_{KL}, L_{\gamma}$. The combination
of these three terms allows the preservation of spatial details, statistical properties and strong scatterers. An ablation study proves how the combination is crucial for taking care of these three aspects simultaneously. Experimental validation, both on simulated and real data, show the advantages on including these SAR image properties in the cost function.

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

Moreover the defined cost function allows to manage a critical issue common of DL methods: the filtering of strong scatterers. The defined cost function allows to identify such elements and to preserve them avoiding an incorrect despeckling. Being a DL based method, once the network training is performed the computational time is very low making the filter suited for almost real time applications. Further works will address the possibility of adapting the filter to multilook, multitemporal and multichannel SAR images.

REFERENCES

[1] J. Lee, “Digital image enhancement and noise filtering by use of local statistics,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-2, no. 2, pp. 165–168, March 1980.

[2] V. S. Frost, J. A. Stiles, K. S. Shannugan, and J. C. Holtzman, “A model for radar images and its application to adaptive digital filtering of multiplicative noise,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-4, no. 2, pp. 157–166, 1982.

[3] D. T. Kuan, A. A. Sawchuk, T. C. Strand, and P. Chavel, “Adaptive noise smoothing filter for images with signal-dependent noise,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. PAMI-7, no. 2, pp. 165–177, March 1985.

[4] R. Touzi, A. Lopes, and P. Bousquet, “A statistical and geometrical edge detector for sar images,” IEEE Transactions on Geoscience and Remote Sensing, vol. 26, no. 6, pp. 564–773, Nov 1988.

[5] A. Lopes, R. Touzi, and E. Nezry, “Adaptive speckle filters and scene heterogeneity,” IEEE Transactions on Geoscience and Remote Sensing, vol. 28, no. 6, pp. 992–1000, Nov 1990.

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

[7] A. Lopes, E. Nezry, R. Touzi, and H. Lau, “Maximum a posteriori speckle filtering and first order texture models in sar images,” in 10th Annual International Symposium on Geoscience and Remote Sensing, 1990, pp. 2409–2412.

[8] H. Guo, J. E. Odgeard, M. Lang, R. A. Gopinath, I. W. Selesnick, and C. S. Burrus, “Wavelet based speckle reduction with application to sar based atdi,” in Proceedings of 1st International Conference on Image Processing, 1994, vol. 1, pp. 75–79 vol.1, Sep 1997.

[9] G. Franceschetti, V. Pascazio, and G. Schirinzi, “Iterative homomorphic technique for speckle reduction in synthetic-aperture radar imaging,” Journal of the Optical Society of America A: Optics and Image Science, and Vision, vol. 12, no. 4, pp. 686–694, 1995.

[10] Langis Gagnon and Alexandre Jouran, “Speckle filtering of SAR images: a comparative study between complex-wavelet-based and standard filters,” in Wavelet Applications in Signal and Image Processing V, Akram Aldroubi, Andrew F. Laine, and Michael A. Unser, Eds. International Society for Optics and Photonics, 1997, p. 3169, pp. 80 – 91, SPIE.

[11] 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, 2002.

[12] 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, Dec 2009.

[13] Goodman J. W., “Statistical properties of laser speckle patterns,” in Laser Speckle and Related Phenomena, Dainty J.C., Ed., chapter 2. Springer, Berlin, Heide, 1984.

[14] 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.

[15] C. Goutte, P. Hellier, C. Delcroix, and C. Barril, “Bayesian non local means-based speckle filtering,” in 2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro, 2008, pp. 1291–1294.

[16] Hua Zhong, Jingjing Xu, and Licheng Jiao, “Classification based non-local means despeckling for SAR image,” in MIIPPR 2009: Automatic Target Recognition and Image Analysis, Tianxu Zhang, Bruce Hirsch, Zhiguo Cao, and Hanqing Lu, Eds. International Society for Optics and Photonics, 2009, vol. 7495, pp. 231 – 238, SPIE.

[17] H. Feng, B. Hou, and M. Gong, “Sar image despeckling based on local homogeneous-region segmentation by using pixel-reliativity measurement,” IEEE Transactions on Geoscience and Remote Sensing, vol. 49, no. 7, pp. 7274–7277, 2011.

[18] G. Ferraoli, V. Pascazio, and G. Schirinzi, “Ratio-based nonlocal anisotropic despeckling approach for sar images,” IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 10, pp. 7785–7798, Oct 2019.

[19] S. Parrilli, M. Poderico, C. V. Angelino, and L. Verdoliva, “A nonlocal sar image denoising algorithm based on llmms wavelet shrinkage,” IEEE Transactions on Image Processing and Remote Sensing, vol. 50, no. 2, pp. 606–616, Feb 2012.

[20] H. Zhong, Y. Li, and L. Jiao, “Sar image despeckling using bayesian nonlocal means filter with sigma preselection,” IEEE Geoscience and Remote Sensing Letters, vol. 8, no. 4, pp. 809–813, July 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, Feb 2014.

[22] R. Touzi, “A review of speckle filtering in the context of estimation theory,” IEEE Transactions on Geoscience and Remote Sensing, vol. 40, no. 11, pp. 2392–2404, Nov 2002.

[23] C. A. Deledalle, L. Denis, G. Poggi, F. Tupin, and L. Verdoliva, “Exploiting patch similarity for sar image processing: The nonlocal paradigm,” IEEE Signal Processing Magazine, vol. 31, no. 4, pp. 69–78, July 2014.

[24] 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.

[25] Qiang Zhang, Qiangqiang Yuan, Jie Li, Zhen Yang, and Xiaoshuang Ma, “Learning a diluted residual network for SAR image despeckling,” Remote Sensing, vol. 10, no. 2, pp. 1–18, 2018.

[26] Francesco Lattari, Borja Gonzalez Leon, Francesco Asaro, Alessio Rucci, Claudio Prati, and Matteo Matteucci, “Deep learning for sar image despeckling,” Remote Sensing, vol. 11, no. 13, 2019.

[27] X. Yang, L. Denis, F. Tupin, and W. Yang, “Sar image despeckling using pre-trained convolutional neural network models,” in 2019 Joint Urban Remote Sensing Event (JURSE), 2019, pp. 1–4.

[28] L. Denis, C. Deledalle, and F. Tupin, “From patches to deep learning: Combining self-similarity and neural networks for sar image despeckling,” in IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, 2019, pp. 5113–5116.

[29] Charles-Alban Deledalle, Loic Denis, Sonia Tabti, and Florence Tupin, “Mulog, or how to apply gaussian denoisers to multi-channel sar speckle reduction?,” IEEE Transactions on Image Processing, vol. 26, no. 9, pp. 4833–4840, Sep 2017.

[30] 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.

[31] Davide Cozzolino, Luisa Verdoliva, Giuseppe Scarpa, and Giovanni Poggi, “Nonlocal cnn sar image despeckling,” Remote Sensing, vol. 12, no. 6, 2020.

[32] S. Vitale, G. Ferrari, and V. Pascazio, “A new ratio image based algorithm for sar speckle reduction,” in IGARSS 2019 - 2019 IEEE International Geoscience and Remote Sensing Symposium, July 2019, pp. 9494–9497.

[33] F. Argenti, A. Lapini, T. Bianchi, and L. Alparone, “A tutorial on speckle reduction in synthetic-aperture radar images,” IEEE Geoscience and Remote Sensing Magazine, vol. 1, no. 3, pp. 6–35, Sept 2013.

[34] Céline Tison, Jean Marie Nicolas, Florence Tupin, and Henri Maitre, “A new statistical model for Markovian classification of Urban areas in
high-Resolution SAR images.” IEEE Transactions on Geoscience and Remote Sensing, vol. 42, no. 10, pp. 2046–2057, 2004.

[35] D. Yue, F. Xu, A. C. Frery, and Y. Jin, “A generalized gaussian coherent scatterer model for correlated sar texture,” IEEE Transactions on Geoscience and Remote Sensing, vol. 58, no. 4, pp. 2947–2964, 2020.

[36] Yi Yang and Shawn Newsam, “Bag-of-visual-words and spatial extensions for land-use classification,” in ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems (ACM GIS), 2010.

[37] Sergio Vitale, Giampaolo Ferraioli, and Vito Pascazio, “Edge preserving cnn sar despeckling algorithm,” arXiv:2001.04716, 2020.

[38] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton, “Imagenet classification with deep convolutional neural networks,” in Advances in Neural Information Processing Systems 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds., pp. 1097–1105. Curran Associates, Inc., 2012.

[39] Sergey Ioffe and Christian Szegedy, “Batch normalization: Accelerating deep network training by reducing internal covariate shift,” in Proceedings of the 32nd International Conference on Machine Learning, Francis Bach and David Blei, Eds., Lille, France, 07–09 Jul 2015, vol. 37 of Proceedings of Machine Learning Research, pp. 448–456, PMLR.

[40] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, “Deep residual learning for image recognition,” CoRR, vol. abs/1512.03385, 2015.

[41] Li Xu, Jimmy Ren, Qiong Yan, Renjie Liao, and Jiaya Jia, “Deep edge-aware filters,” in Proceedings of the 32nd International Conference on Machine Learning, Francis Bach and David Blei, Eds., Lille, France, 07–09 Jul 2015, vol. 37 of Proceedings of Machine Learning Research, pp. 1669–1678, PMLR.

[42] Sergio Vitale, Giampaolo Ferraioli, and Vito Pascazio, “Complexity analysis of an edge preserving cnn sar despeckling algorithm,” arXiv:2004.08345, 2020.

[43] Diederik P. Kingma and Jimmy Ba, “Adam: A method for stochastic optimization,” in 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings, 2015.

[44] Luis Gomez, Raydonal Ospina, and Alejandro C. Frery, “Unassisted filters,” Remote Sensing, vol. 9, no. 4, 2017.

[45] R. M. Haralick, K. Shanmugam, and I. Dinstein, “Textural features for image classification,” IEEE Transactions on Systems, Man, and Cybernetics, vol. SMC-3, no. 6, pp. 610–621, 1973.

[46] S. Vitale, D. Cozzolino, G. Scarpa, L. Verdoliva, and G. Poggi, “Guided patchwise nonlocal sar despeckling,” IEEE Transactions on Geoscience and Remote Sensing, vol. 57, no. 9, pp. 6484–6498, Sep. 2019.

**APPENDIX A**

**ON THE IMPACT OF THE COST FUNCTION**

In this section an ablation study has been carried out in order to assess the impact of the defined cost function. The cost function is given by a combination of the terms in Eq. 6. In order to compare the performance and the impact of these three terms, the same architecture is trained on same dataset with a cost function composed once only of the \( L_2 = \nabla \lambda \) term, once with the combination \( L_{kl} = L_2 + \lambda_{kl} L_{KL} \), and once with the combination \( L_\nabla = L_2 + \lambda_\nabla L_\nabla \). These solutions are compared with the proposed method. In Tab. VI, we summarize the results.

**TABLE VI**

**NUMERICAL ASSESSMENT ON SIMULATED DATASET FOR DIFFERENT COST FUNCTIONS: THE VALUE ARE AVERAGED ON THE WHOLE SIMULATED TESTING DATASET COMPOSED OF 100 IMAGES. FROM TOP TO BOTTOM: NETWORK TRAINED WITH \( L_2, L_{kl}, L_\nabla, \mathcal{L} \)**

|      | SSIM | SNR | MSE  |
|------|------|-----|------|
| \( L_2 \) | 0.7509 | 8.8514 | 0.0040 |
| \( L_{kl} \) | 0.7514 | 8.8543 | 0.0039 |
| \( L_\nabla \) | 0.7512 | 8.8585 | 0.0039 |
| \( \mathcal{L} \) | 0.7510 | 8.8555 | 0.0039 |

The results are almost the same for each solution, like there is no difference in introducing such terms in the cost function. It seems that \( L_2 \) is enough for the despeckling. However, these are average metrics that do not take into account the details that make the difference between one solution and an other. Moving to real data, things largely change.

In Fig. 14 a detail for each dataset CSK-Pavia, RADAR-SAT-Phoneix and TSX-Tehran are shown. It can be noted how important is the impact of the cost function. Starting from \( L_2 \) that try to preserve spatial details, the use of the KL divergence in \( L_{kl} \) helps in filtering the homogeneous areas but we lose information on strong scatterers producing smoothing effect. In addition, the \( L_\nabla \) try to preserve edges but does not consider the speckle properties and tends to create strange artefacts in the neighbourhood of the strong scatterers. The proposed cost function \( \mathcal{L} \) allows the filter the image balancing at the same time the statistical properties of the noise and the preservation of dominant scatterers.