InSAR2InSAR: A Self-Supervised Method for InSAR Parameters Estimation

Carla Geara, Colette Gelas, Louis De Vitry, Elise Colin, Florence Tupin

InSAR2InSAR: A Self-Supervised Method for InSAR Parameters Estimation

Carla Geara*†‡, Colette Gelas‡, Louis De Vitry‡, Elise Colin†, and Florence Tupin*
*LTCI, Télécom Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France
†DTIS, ONERA, Université de Paris Saclay, 91123 Palaiseau, France
‡Kanop (RAIL SAS), 91190 Gif-Sur-Yvette, France

***Abstract*—Interferometric Synthetic Aperture Radar (InSAR) is a remote sensing tool that provides comprehensive information about the Earth's surface. However, InSAR parameters are highly corrupted by speckle, which limits their utility. Deep learning methods have recently achieved promising results in improving the reliability of InSAR parameters. Most of the proposed methods are fully supervised. These methods are usually trained on synthetic data, which are not able to fully take into account all the properties of real images. In this paper, we address this issue by extending the self-supervised denoising approach *Noise2Noise*, previously proposed by Lehtinen et al. in 2018, for the joint estimation of InSAR parameters. Additionally, the proposed method uses a loss function that is adapted to the InSAR noise model, making it well-suited for the problem we are addressing.**

***Index Terms*—Deep learning, despeckling, self-supervision, interferometric Synthetic Aperture Radar, noise statistics.**

I. Introduction

Interferometric Synthetic Aperture Radar (InSAR) is a remote sensing technique that uses two or more SAR acquisitions of the same scene to extract information regarding the Earth's surface topography or deformation. This technique is widely used in various fields such as digital terrain model computation and land movement monitoring. Using two interferometric acquisitions (i.e. two acquisitions of the same area taken from slightly different points in space), we can compute the InSAR interferogram, which provides a measure of coherence or correlation between the two acquisitions, and a measure of the phase difference that is related to the topography or movement of the scene. However, the acquisitions are often corrupted by speckle. This speckle is a random noise that depends on the sensor configuration, the scattering mechanism, atmospheric conditions, and the temporal and spatial differences between the acquisitions. These fluctuations can make the InSAR parameters very difficult to exploit.

In recent years, considerable efforts have been invested in overcoming this challenge. The most traditional family of InSAR denoisers is the family of local filters, that use the local neighborhood to smooth out variations. The most classical and known filter is the boxcar filter [1], which consists of a rectangular averaging filter over the image. This filter is very efficient in terms of simplicity, however it suffers from loss in resolution and artifacts around high topography regions. The Lee filter [2] utilizes directional kernels that better align with phase fringes. Additionally, it analyzes local noise statistics to determine the appropriate filtering strength. This allows the Lee filter to preserve details and fringes more effectively.

Another family of filters is the family of non local filters. These methods overcome the limitations of the restricted neighborhood by searching for similarities outside of the local patch. A prominent example is the NL-InSAR [3], which is a probabilistic patch-based filter that uses maximum likelihood to search for similarities inside the image. Another notable technique is the InSAR-BM3D [4], which extends the BM3D [5] denoising method for InSAR parameters estimation.

Variational methods also offer a powerful approach to InSAR denoising by minimizing a cost function that consists of a data fidelity term and a regularization term. MuLog [6] is a generic variational denoising method that incorporates Gaussian denoisers within a multi-channel framework based on the Plug-and-Play ADMM [7] scheme.

Deep learning has recently shown great success in this domain. Most of the existing InSAR denoising approaches in deep learning are fully supervised, such as ϕ-Net [8] and InSAR-MONet [9]. ϕ-Net [8] uses a UNet architecture for the joint estimation of the coherence and phase. The authors focused on developing a large training synthetic dataset that accounts for different scenarios. InSAR-MONet [9] performs InSAR phase denoising by training the network to minimize a multi-objective lost function. This multi-objective loss function penalizes the recovery of spatial information, and preservation of edges and statistical properties.

As fully supervised methods require the disposition of noise-free samples, which are very difficult to acquire, these methods are usually trained on simulated data. Nevertheless, it is difficult to simulate data faithful to real samples. Therefore, it is crucial to explore self-supervised approaches to be able to train networks on real images. One of the existing self-supervised denoising approaches is *Noise2Noise* [10]. The rationale behind this approach is that a deep learning model can be trained to restore a noisy image using a pair of independent noisy observations of the same scene by mapping one noisy sample to the other, without requiring the clean sample. This is only applicable under the condition that the expectation of the noisy observations matches the clean target and that the training dataset is sufficiently large. Although disposing of two noisy samples of the same scene is not frequent, this situation

This work was supported by Kanop (RAIL SAS) and the ANRT (Association Nationale de la Recherche et de la Technologie).

occurs for the Sentinel 1 sensor with the burst overlap. Thanks to this specificity, two noisy interferometric pairs of the same area are available.

Therefore, we propose in this article a self-supervised method relying on the *Noise2Noise* principle to reduce the speckle in InSAR data. The joint estimation of InSAR parameters is done by denoising the interferometric covariance matrix. Additionally, the proposed approach takes into account the peculiarities of the speckle by considering a loss that is adapted to its distribution. We currently exclusively present results based on simulated data, with a strategic intention to migrate towards real Sentinel 1 data in the future. The use of simulated data is crucial to validate the method before applying it to more complex real-world data.

II. Methodology

A. InSAR Statistics

In order to effectively introduce the despeckling method, it is necessary to take into account the speckle's statistics.

As mentioned in the introduction, the proposed method aims to restore the InSAR parameters by denoising the covariance matrix. In the general case of multi-channel SAR acquisitions with D channels (polarimetric and/or interferometric), the empirical covariance matrix $\boldsymbol{C} \in \mathbb{C}^{D\times D}$ can be computed over a local window as in (1):

$$\boldsymbol{C} = \frac{1}{L}\sum_{k=1}^{L} \vec{v}_k \vec{v}_k^{*} \ , \tag{1}$$

where L is the number of pixels considered, $\vec{v}_k \in \mathbb{C}^D$ is the complex amplitude vector at pixel k within the window, and * denotes the Hermitian transpose. In the case of L$\geq$ D, according to Goodman's model [11], the speckle noise in the covariance matrix $\boldsymbol{C}$ follows the Wishart distribution $\mathcal{W}(\mathbf{\Sigma}, L)$ given by:

$$p_{\mathbf{C}}(\mathbf{C}|\mathbf{\Sigma}) = \frac{L^{LD}|\mathbf{C}|^{L-D}}{\Gamma_D(L)|\mathbf{\Sigma}|^{L}} \exp(-Ltr(\mathbf{\Sigma}^{-1}\mathbf{C})) \ , \tag{2}$$

where $\mathbf{\Sigma}$ is a Hermitian positive definite matrix that represents the true covariance matrix. In the interferometric case where $D = 2$, $\mathbf{\Sigma}$ is expressed in (3):

$$\mathbf{\Sigma} = \begin{bmatrix} A_1^2 & A_1A_2\rho_{12}e^{j\phi_{12}} \\ A_1A_2\rho_{12}e^{-j\phi_{12}} & A_2^2 \end{bmatrix} \ , \tag{3}$$

where A_1 and A_2 denote the amplitudes of each acquisition, and ρ_{12} and ϕ_{12} denote the coherence and the phase difference between both interferometric acquisitions respectively. The Wishart distribution models the noise as $\boldsymbol{C} = \mathbf{\Sigma}^{\frac{1}{2}}\boldsymbol{S}\mathbf{\Sigma}^{\frac{1}{2}}$, where $\boldsymbol{S} \sim \mathcal{W}(\mathbf{I}, L)$, with the following expectation and variance [6]:

$$\mathbb{E}[\boldsymbol{C}] = \mathbf{\Sigma} \ , \tag{4}$$

$$Var[\boldsymbol{C}_{ij}] = \frac{1}{L}\mathbf{\Sigma}_{ii}\mathbf{\Sigma}_{jj} \ . \tag{5}$$

The variance of the off-diagonal elements of $\boldsymbol{C}$ depends on the intensities of the signals as show in (5). This implies that the noise is signal-dependent.

B. Loss

Provided that the expectation of the noisy observations matches the clean target (as shown in (4)), the application of the *Noise2Noise* framework is possible. Given two noisy observations of the covariance matrix $(\boldsymbol{C}_1, \boldsymbol{C}_2)$, we estimate the parameters θ of the estimator $f_\theta : \boldsymbol{C} \mapsto \hat{\mathbf{\Sigma}} = f_\theta(\boldsymbol{C})$, by minimizing a loss function $\boldsymbol{\mathcal{L}}$ that compares $f(\boldsymbol{C}_1)$ with $\boldsymbol{C}_2$:

$$\hat{\theta}_{Noise2Noise} \in argmin_\theta \mathbb{E}[\boldsymbol{\mathcal{L}}(f_\theta(\boldsymbol{C}_1), \boldsymbol{C}_2)] \ . \tag{6}$$

To take into account the statistical properties of the speckle noise in the optimization problem, the minimized loss is the negative log-likelihood of the conditional distribution $p(\mathbf{C}|\mathbf{\Sigma})$ presented in (2). This results, in the *Noise2Noise* context, in the following loss shown in (7):

$$\begin{aligned} \boldsymbol{\mathcal{L}}(\hat{\mathbf{\Sigma}}_1, \boldsymbol{C}_2) &= -\log p(\boldsymbol{C}_2|\hat{\mathbf{\Sigma}}_1) \\ &= L(\log|\hat{\mathbf{\Sigma}}_1| + tr(\hat{\mathbf{\Sigma}}_1^{-1}\boldsymbol{C}_2)) + K \ , \end{aligned} \tag{7}$$

where $\hat{\mathbf{\Sigma}}_1 = f(\boldsymbol{C}_1)$ is the denoised estimation of $\boldsymbol{C}_1$ and K is a constant that does not depend on $\hat{\mathbf{\Sigma}}_1$.

C. Methodology

The full methodology, summarized in Fig. 1, follows the step described below:

1) Two noisy interferometric pairs of the same area are taken as input. The covariance matrices of both pairs $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are first computed with L = 1, in order to preserve the resolution of the images. Given that $L < D$, $\boldsymbol{C}$ is rank-deficient. This issue was thoroughly discussed in [12], and the covariance matrix pre-processing operations proposed by the authors were applied here in order to mitigate the rank-deficiency problem.
2) The matrix logarithm of the input covariance matrix $\tilde{\boldsymbol{C}}_1 = \log \boldsymbol{C}_1$ is then computed. It was shown in [6] that applying the matrix-logarithm stabilizes the noise variance and reduces its dependency on the signal.
3) Given that the covariance matrix is complex, it is transformed into a real-valued tensor, by separating the real and imaginary values of the upper triangle matrix of $\tilde{\boldsymbol{C}}_1$.
4) The denoised output is then converted into a complex covariance matrix $\hat{\mathbf{\Sigma}}_1$ (after converting back to the complex and computing the matrix exponential).
5) Finally, the network parameters are optimized by minimizing the negative log-likelihood as described in Section II-B.

The model architecture adopted is the UNet architecture detailed in Fig. 1.

III. Experiments

To evaluate the performance of the proposed method, the training is first conducted on simulations before moving on to real data. In this section, we show the generation process of the simulated data, the experimental setup and the results.

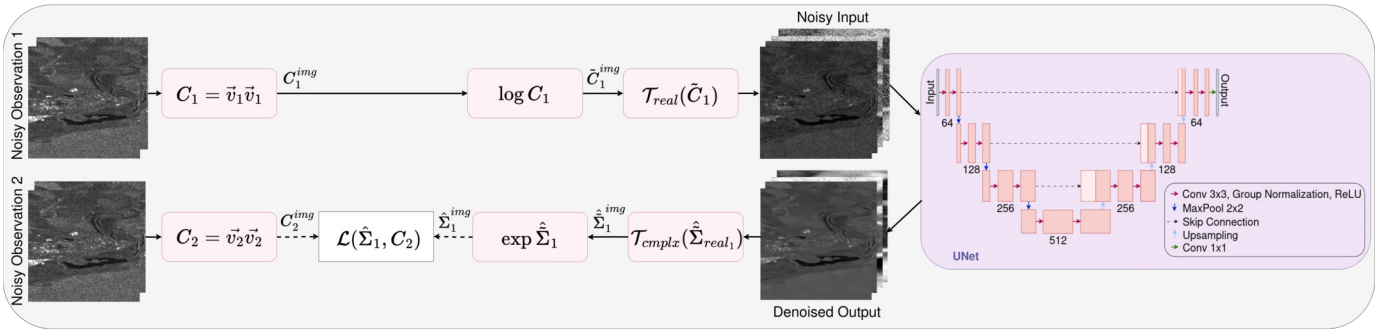

Fig. 1: Overview of InSAR2InSAR: the matrix logarithm of $\boldsymbol{C}_1$ is computed and the matrix is transformed into a real-valued vector, which is fed to the network. The inverse transforms are then performed to obtain $\hat{\boldsymbol{\Sigma}}_1$ which is compared with $\boldsymbol{C}_2$.

A. Data Simulations

The noisy interferometric acquisitions are simulated following the model described in [3]:

$$\begin{bmatrix} z_1 \\ z_2 \end{bmatrix} = M_C \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \ , \tag{8}$$

where x_1 and x_2 are two complex standard Gaussian random variables $[x_1, x_2]^T \sim \mathcal{CN}(0, \mathbf{I})$ and M_C denotes the lower triangle matrix in the Cholevsky decomposition of the covariance matrix, as expressed in (9):

$$M_C = \begin{bmatrix} A_1 & 0 \\ A_2\rho_{12}e^{-j\phi_{12}} & A_2\sqrt{1-\rho_{12}^2} \end{bmatrix} \ . \tag{9}$$

M_C depends on the coherence ρ_{12}, phase ϕ_{12} and the noise-free amplitudes A_1 and A_2.

The objective is to generate a synthetic dataset that is realistic and preserves to the fullest extent feasible the properties of real images. Nevertheless, the simulations generated with the simulation model described above will not take into account the imaging system and its Point Spread Function (PSF). The noise-free InSAR parameters are generated using real Sentinel 1 Interferometric Wide Swath SLC interferometric pairs:

1) **Amplitudes**: the real noisy amplitudes are denoised using MERLIN [13], a denoising method that uses the real and imaginary parts of single-look complex images to restore the amplitudes.
2) **Phase simulation**: the phase is simulated using the Shuttle Radar Topography Mission (SRTM) DEM that has been interpolated on the Sentinel 1 product grid. The wrapped phase ϕ can be simulated using the interpolated DEM and the acquisition parameters [14] as show in (10):

$$\phi = \angle exp(j\frac{4\pi B_{orth}}{\lambda R_{srange} \sin\theta} h) \ , \tag{10}$$

where h is the height provided by the DEM, B_{orth} is the baseline, λ is the sensor wavelength, R_{srange} is the range distance from the center of the scene, and θ is the incidence angle.
Following the approach in [9], we used varying baseline values for each DEM to simulate diverse phase patterns, enhancing the dataset's generalization capability.

3) **Coherence**: the coherence ρ can be estimated using the cross-correlation r of the intensities of the interferometric pairs as described in [15]:

$$\rho = \begin{cases} \sqrt{2r-1} & \text{r} > 0.5 \ , \\ 0 & \text{otherwise} \ . \end{cases} \tag{11}$$

In order to estimate the noise-free coherence, the denoised intensities are used to compute r as in (12):

$$r(A_1, A_2) = \frac{E[A_1^2 A_2^2]}{\sqrt{E[A_1^4]E[A_2^4]}}. \tag{12}$$

The noise free coherence is then estimated according to (11). Different linear transformations are then applied to the estimated coherence to simulate different levels of coherence, hence different levels of noise. The coherence values range from 0.3 to 0.9.

Finally, after computing the noise-free parameters, the noisy interferometric images can be simulated according to (8). Fig. 2 shows an example of 12 simulated phases for three levels of coherences and four different baselines for one patch. The top line represents the noise-free phase with increasing baseline from left to rightand the first column represents the three levels of noise-free coherence. Two noisy simulations were created for the same noise-free parameters, which will be used to compute the two noisy covariance matrices $(\boldsymbol{C}_1, \boldsymbol{C}_2)$.

B. Experimental Setup

In order to train the network, a synthetic dataset was simulated following the protocol described in Section III-A. The simulated data were generated over a burst acquired in Luthenay-Uxeloup, in central France. The burst was then divided into 470 patches of size 256x256 pixels. The training was carried out on 80% of these patches, while the remaining 20% were used for the evaluation. Each patch was used to simulate 3 different levels of coherence and 6 different phase patterns with different baselines as described in Section III-A. This yields to 18 different simulations per patch, which results in a total of 6768 patches for training and 1692 for testing.

The network was trained on batches of 4 patches of size 64x64 which were randomly cropped from the 256x256 training patches, using the Adam optimizer. The training carried

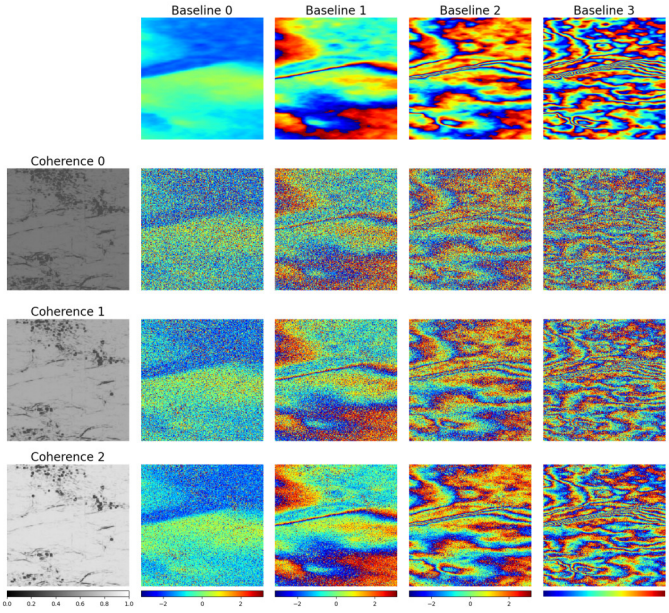

Fig. 2: Example of 12 phase simulations with multiple baselines and coherence levels.

on for 50 epochs, with an initial learning rate of 0.0001 that decreased by a factor of 0.04 after each epoch.

C. Computed Metrics

The metrics used for the evaluation of the performance of the model on simulated data are the following:

1) The Mean Squared Error MSE given by (13).

$$MSE = \frac{1}{N}\sum_i (x_i - \hat{x}_i) \ , \tag{13}$$

where N is the number of pixels, x_i is the ground truth and $\hat{x}_i$ is the estimation of pixel i. A perfect estimation results in a MSE equal to 0. This metric is computed for both phase and coherence.

2) The Structural Similarity Index Measure (SSIM) given by (14) that measures the structural similarities between the estimation and the real quantity.

$$SSIM(x,y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \ , \tag{14}$$

where μ_x and μ_y are the means of x and y, σ_x and σ_y are the standard deviation of x and y, σ_{xy} is the co-variance between x and y, and c_1 and c_2 are constants.
A perfect similarity results in a SSIM equal to 1. This metric is computed for both phase and coherence.

3) The cosine dissimilarity CM, computed as in (15):

$$CM = \frac{1}{2N}\sum_i (1 - cos(\phi_i - \hat{\phi}_i)) \ . \tag{15}$$

The values of CM range from 0 and 1, with 0 being the case of perfect estimation. This is computed for the phase only.

D. Results On Simulations

Fig. 3 shows results on two test examples with low coherence and two different levels of phase fringes. The results are also compared with two fully supervised deep learning methods: InSAR-MONet [9], which estimates the phase only, and ϕ-Net [8], which estimates both coherence and phase.

Considering the phase estimation, while the three presented methods have difficulty recreating high rate fringes (see example 2), the estimation of the proposed method InSAR2InSAR is the closest to the noise-free phase. ϕ-Net smooths the phase and has difficulty recreating the small variations. InSAR-MONet creates some artifacts in the denoised output.

Considering the coherence estimation, both methods InSAR2InSAR and ϕ-Net show strong noise suppression, at the cost of reduced sharpness in InSAR2InSAR and severe smoothing for ϕ-Net. However, InSAR2InSAR seems to better preserve the small variations of the coherence. Additionally, both methods recreate strong phase fringes in the coherence estimation (see top right corner of Example 2).

A quantitative analysis was also conducted on the validation set, which consists of 1692 patches of size 256x256 pixels. The metrics listed in Section III-C were computed and the results are displayed in Table I. The best results are highlighted in bold. InSAR2InSAR achieves the best performance for all metrics, which validates the observed results in Fig. 3.

TABLE I: Quantitative results on the validation dataset. The best results are highlighted in bold.

	Coherence Metrics		**Phase Metrics**		
	MSE	***SSIM***	***MSE***	***CM***	***SSIM***
ϕ-Net	0.0091	0.529	1.196	0.0249	0.66
InSAR-MONet	-	-	0.992	0.023	0.678
InSAR2InSAR	**0.0045**	**0.587**	**0.771**	**0.0099**	**0.772**

E. Discussion On Coherence Estimation

While InSAR2InSAR achieves satisfactory results for phase estimation, it still has difficulty estimating the coherence. A possible explanation for this behavior is the low sensitivity of the considered loss to coherence variations. The study of the variations of the negative log-likelihood with respect to the coherence validates this assumption. Additionally, InSAR2InSAR tends to estimate a low coherence for some phase fringe areas. A similar low coherence estimation in phase fringe areas was also observed in Φ-Net. This could be because such fringes are interpreted as noise, resulting in lower coherence values. We will dedicate further attention and effort to address these challenges effectively.

IV. Conclusion

In this article, we propose InSAR2InSAR, a new self-supervised denoising method that extends *Noise2Noise* for the joint estimation of InSAR parameters. The proposed method shows promising results on simulated data. In future works, we plan to train InSAR2InSAR on a real dataset, which consists of the burst overlap areas extracted from Sentinel 1 Interferometric Wide Swath SLC data.

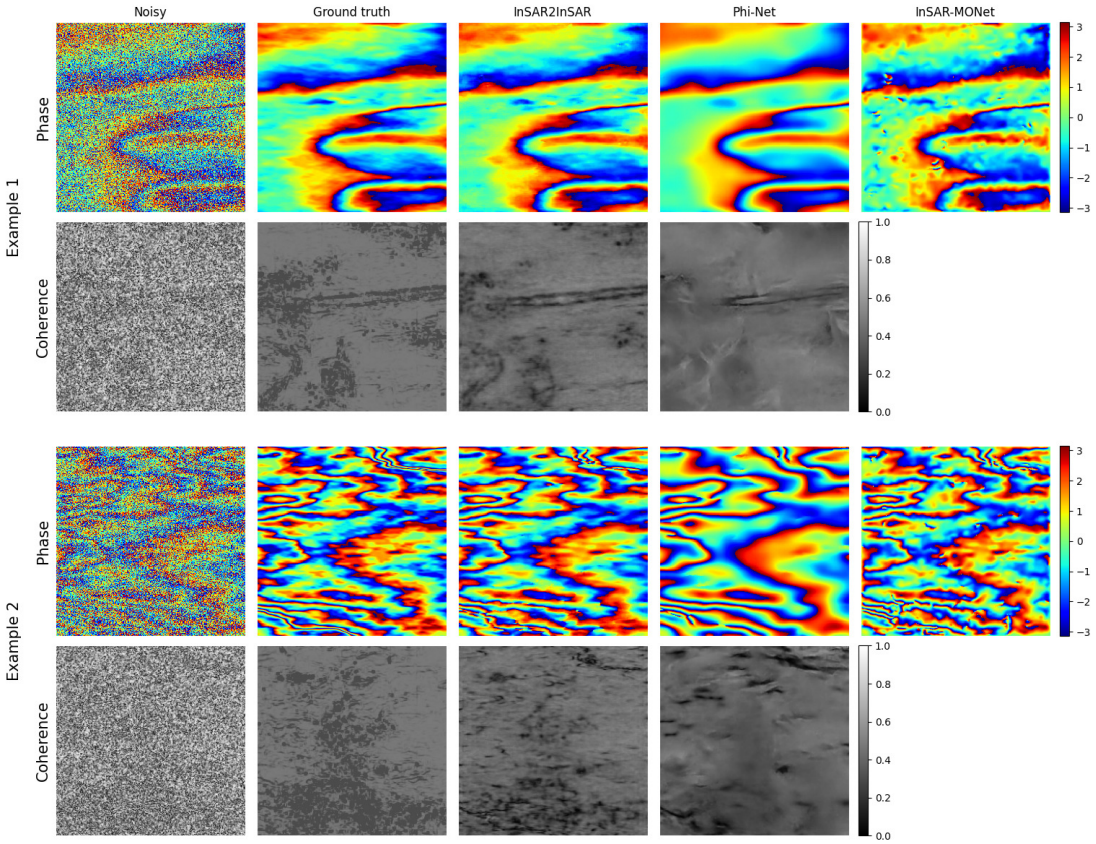

Fig. 3: Two examples of phase and coherence estimations on simulated data. From left to right: noisy simulations, noise-free simulations, estimations of the proposed InSAR2InSAR, ϕ-Net [8] and InSAR-MONet [9].

ACKNOWLEDGMENT

The authors are thankful to Cristiano Ulondu-Mendes (LTCI, Télécom Paris) for the helpful scientific discussions. The authors are also thankful to Jeremy Anger and Roland Akiki (Centre Borelli, ENS Paris-Saclay) for preliminary discussions on this topic.

REFERENCES

[1] P. D. P. W. J. S. Lee, L. Jurkevich and A. Oosterlinck, "Speckle filtering of synthetic aperture radar images: A review," *Remote Sensing Reviews*, vol. 8, no. 4, pp. 313–340, 1994.

[2] J.-S. Lee, K. Papathanassiou, T. Ainsworth, M. Grunes, and A. Reigber, "A new technique for noise filtering of SAR interferometric phase images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 36, no. 5, pp. 1456–1465, 1998.

[3] C. Deledalle, L. Denis, and F. Tupin, "NL-InSAR: Non-Local Interferogram Estimation," *IEEE Transaction on Geoscience and Remote Sensing*, vol. 49, no. 4, pp. 1441–1452, 2011.

[4] F. Sica, D. Cozzolino, X. X. Zhu, L. Verdoliva, and G. Poggi, "InSAR-BM3D: A Nonlocal Filter for SAR Interferometric Phase Restoration," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 6, pp. 3456–3467, Jun. 2018.

[5] 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.

[6] C.-A. Deledalle, L. Denis, S. Tabti, and F. Tupin, "MuLoG, or How to Apply Gaussian Denoisers to Multi-Channel SAR Speckle Reduction?" *IEEE Transactions on Image Processing*, vol. 26, no. 9, pp. 4389–4403, 2017.

[7] S. V. Venkatakrishnan, C. A. Bouman, and B. Wohlberg, "Plug-and-Play priors for model based reconstruction," in *2013 IEEE Global Conference on Signal and Information Processing*, Dec. 2013, pp. 945–948.

[8] F. Sica, G. Gobbi, P. Rizzoli, and L. Bruzzone, "ϕ-Net: Deep Residual Learning for InSAR Parameters Estimation," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 5, pp. 3917–3941, 2021.

[9] S. Vitale, G. Ferraioli, V. Pascazio, and G. Schirinzi, "InSAR-MONet: Interferometric SAR Phase Denoising Using a Multiobjective Neural Network," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–14, 2022.

[10] 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*, ser. Proceedings of Machine Learning Research, J. Dy and A. Krause, Eds., vol. 80. PMLR, 10–15 Jul 2018, pp. 2965–2974.

[11] N. R. Goodman, "Statistical Analysis Based on a Certain Multivariate Complex Gaussian Distribution (An Introduction)," *The Annals of Mathematical Statistics*, vol. 34, no. 1, pp. 152 – 177, 1963.

[12] C.-A. Deledalle, L. Denis, and F. Tupin, "Speckle Reduction in Matrix-Log Domain for Synthetic Aperture Radar Imaging," *Journal of Mathematical Imaging and Vision*, vol. 64, no. 3, pp. 298–320, Mar. 2022.

[13] E. Dalsasso, L. Denis, and F. Tupin, "As If by Magic: Self-Supervised Training of Deep Despeckling Networks With MERLIN," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–13, 2022.

[14] G. Ferraioli, C.-A. Deledalle, L. Denis, and F. Tupin, "PARISAR: Patch-Based Estimation and Regularized Inversion for Multibaseline SAR Interferometry," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 3, pp. 1626–1636, 2018.

[15] A. Guarnieri and C. Prati, "SAR interferometry: a "Quick and dirty" coherence estimator for data browsing," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 35, no. 3, pp. 660–669, 1997.