Φ-Net: Deep Residual Learning for InSAR Parameters Estimation

Francescopaolo Sica, *Member, IEEE*, Giorgia Gobbi, Paola Rizzoli, and Lorenzo Bruzzone, *Fellow, IEEE*

***Abstract*— Nowadays, deep learning (DL) finds application in a large number of scientific fields, among which the estimation and the enhancement of signals disrupted by the noise of different natures. In this article, we address the problem of the estimation of the interferometric parameters from synthetic aperture radar (SAR) data. In particular, we combine convolutional neural networks together with the concept of residual learning to define a novel architecture, named Φ-Net, for the joint estimation of the interferometric phase and coherence. Φ-Net is trained using synthetic data obtained by an innovative strategy based on the theoretical modeling of the physics behind the SAR acquisition principle. This strategy allows the network to generalize the estimation problem with respect to: 1) different noise levels; 2) the nature of the imaged target on the ground; and 3) the acquisition geometry. We then analyze the Φ-Net performance on an independent data set of synthesized interferometric data, as well as on real InSAR data from the TanDEM-X and Sentinel-1 missions. The proposed architecture provides better results with respect to state-of-the-art InSAR algorithms on both synthetic and real test data. Finally, we perform an application-oriented study on the retrieval of the topographic information, which shows that Φ-Net is a strong candidate for the generation of high-quality digital elevation models at a resolution close to the one of the original single-look complex data.**

***Index Terms*— Coherence, convolutional neural network (CNN), deep learning (DL), denoising, interferometric phase, residual learning, synthetic aperture radar (SAR) interferometry.**

I. Introduction

SYNTHETIC aperture radar interferometry (InSAR) is by far one of the most exploited radar remote sensing techniques for the study of the geosphere. It is mostly used for the retrieval of Earth's surface topography [i.e., through the generation of digital elevation models (DEMs)] and the observation of its deformations. InSAR is based on the coherent acquisition of at least a pair of SAR complex images, also called single-look complex (SLC). The phase difference between two SLCs acquired from two slightly different positions is usually indicated as the interferometric phase and encodes the information about the terrain topography. Likewise, if a temporal separation between the two SLCs is present, a further phase term can be retrieved, due to changes of the ground topography occurred during the observation time.

The interferometric phase measurement is normally affected by noise, which arises depending on the nature of the observed scattering mechanism, the atmospheric conditions, and the acquisition geometry and sensor parameters. Furthermore, being the interferometric phase wrapped in a 2π interval, a further processing step, named phase unwrapping, is necessary to retrieve the absolute phase value. Far from being a trivial procedure, the unwrapping is the process to add or subtract to each pixel multiples of 2π phase values in order to reconstruct a phase field that is consistent all over the image. This procedure is drastically impaired by phase noise. For this reason, it is normally performed on a set of reliable pixels that are selected according to the degree of correlation between the interferometric pair. This last quantity is measured as the module of the complex correlation between the two SLCs and is normally indicated as coherence. A precise estimation of both the interferometric phase and the coherence is, therefore, a key aspect for the correct application of the InSAR technique.

Since the first application of SAR interferometry back in 1993, which was aimed to produce an image representing the topographic deformation consequent to the 1992 earthquake in Landers, CA, USA [1], several approaches have been proposed to estimate both the interferometric phase and the coherence. By considering the signal statistics of coherent radar measurements [2] as well as the interferometric signal statistics [3], Seymour and Cumming [4] assume wide-sense local stationarity (smooth variation) of the interferometric signal within a local neighborhood and derive the maximum-likelihood estimator of the interferometric phase and coherence. This is the simplest approach and is still nowadays widely used since it involves the application of a simple plain moving average procedure. The dimension of the averaging window clearly represents a tradeoff between variance and bias of the estimate. For this reason, further developments have been proposed in order to better deal with signal nonstationarities, by selecting the window shape among a set of predefined ones [5] or by arbitrarily adapting the shape to the direction of the interferometric fringes as in [6].

Alternatively, starting from the observation that the interferometric signal can be well approximated using a set of Fourier basis functions, Goldstein and Werner [7] proposed an adapted filter in the frequency domain, which enhances

Manuscript received December 20, 2019; revised March 25, 2020 and June 25, 2020; accepted August 13, 2020. Date of publication September 15, 2020; date of current version April 22, 2021. *(Corresponding author: Francescopaolo Sica.)*

Francescopaolo Sica, Giorgia Gobbi, and Paola Rizzoli are with Microwaves and Radar Institute, German Aerospace Center, 82234 Weßling, Germany (e-mail: francescopaolo.sica@dlr.de; giorgia.gobbi@dlr.de; paola.rizzoli@dlr.de).

Lorenzo Bruzzone is with the Department of Information Engineering and Computer Science, University of Trento, 38123 Trento, Italy (e-mail: lorenzo.bruzzone@unitn.it).

Color versions of one or more of the figures in this article are available online at https://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2020.3020427

the signal power spectrum. Also, in this case, further modifications have improved the performance of such a procedure by regulating the intensity of the filtering according to the local noise power [8]. Following the idea of alternative domain representation, Lopez-Martinez and Fabregas [9] modeled the interferometric phase noise in the complex domain and proposed a filtering strategy based on the wavelet transform.

Sparse representation has also been applied to phase estimation in [10]. Here, a dictionary of image slices is generated from the noisy image itself in order to provide a new representation of the interferometric signal as sparse as possible (where only a few samples differ from zero). In order to retain the most informative slices of the image, while discarding the noisy ones, the dictionary is formed by matrix factorization under an L1-norm constraint. The high computational complexity, which is caused by the exhaustive search of independent image slices for the creation of the dictionary, is justified by the high performance of this algorithm, which still today deserves to be mentioned among the state-of-the-art (SOA) methods.

Among the approaches that apply to the image spatial domain, a large variety of patch-based methods have been recently proposed for the interferometric phase and coherence estimation. Following the example of the pioneering nonlocal-means approach [11], patch-based methods are able to preserve spatial details while providing strong noise suppression. Here, the basic idea is to search for similar pixels prior to the actual estimation. In this way, the predictors are selected according to a suitable distance metric rather than to their spatial proximity (from here the name of "nonlocal"). Precisely, a "patch" is a small image segment taken as a neighborhood of a pixel, which is used to provide enough samples for the computation of the statistical similarity between two pixels. Deledalle *et al.* [12] proposed an iterative approach to the definition of a weighted maximum-likelihood estimator. In the first step, a distance metric is chosen according to the interferometric signal statistics [3], while the Kullback–Leibler divergence is used to refine the weights at the subsequent steps. A generalization of this method to polarimetric data is then provided in [13], where an estimate of the covariance matrix is presented and the distance metric is derived according to the Wishart distribution. Following the processing scheme of the block-matching and 3-D filtering [14], Sica *et al.* [15] proposed to separately process the real and imaginary parts of the complex interferogram in order to estimate the interferometric phase. This is a two-step approach that exploits a cosine distance metric to form groups of similar patches of fixed size and then performs the estimation in the wavelet domain. Once a group of similar patches is formed, the algorithm computes the wavelet transform on the whole group. In the first step, shrinkage of the wavelet coefficients is applied in order to provide a first rough estimate. In the second step, since a prior estimation is available, it is possible to utilize a Wiener filter in the wavelet domain.

In [16], a general criterion, which is valid for every patch-based approach, is presented. It is aimed to increase the number of predictors and, therefore, the final estimation performance. By considering that the interferometric information is given by the spatial variation of the phase rather than by its absolute value, this criterion extends the selection of similar patches to those which show analogous topography variations, i.e., the ones that differ only for a constant phase offset over the patch. This approach provides results that overcome all the existing patch-based methods, by achieving significant improvements especially over areas characterized by a low number of similar patches. Thus, patch-based approaches represent, nowadays, the SOA for the estimation of the interferometric parameters and have shown a great potential for the generation of accurate high-resolution DEMs [17].

Only recently, following the fast development of deep learning (DL) techniques, novel approaches based on convolutional neural networks (CNNs) have been introduced in this field, showing very promising results [18], [19]. It is worth recalling that the application of CNNs to denoising problems in signal and image processing dates back to 2009 when the first shallow architecture was proposed for this purpose [20], setting the groundwork for CNN denoising. On the basis of the VGG model (proposed by the Visual Geometry Group, University of Oxford, in [21]), Zhang *et al.* [18] presented the use of shortcut connections that add the input to the output of the network in conformity with the principle of residual networks [22]. This is the first network that exploits residual connections for the image denoising task and introduces a novel methodology to address this issue. Following this approach, some denoising algorithms have already been proposed in the SAR remote sensing field, e.g., for despeckling SAR images [23], [24]. While such approaches slightly differ in terms of architectural solutions, they all process the signal in the homomorphic domain, in order to obtain an additive noise model and then apply the residual connection according to [18].

The first attempt at using CNN with residual connections has been recently made for the estimation of the interferometric phase [25]. The authors propose to process the real phase, which is characterized by interferometric fringes, i.e., wrapped phase patterns. The reason for this choice relies on a simple signal modeling and handling, even if it shows a very nonstationary behavior. The network architecture follows the one proposed in [18], with four convolutional layers and without pooling layers. It is trained on synthetic data of open-pit mines, which is a very challenging case characterized by abrupt phase changes. This network shows promising preliminary results.

In this article, we propose Φ-Net, a novel CNN architecture for the joint estimation of the interferometric phase and coherence. Φ-Net addresses very challenging goals: 1) capability to perform blind denoising, without any prior assumption on the noise power; 2) capability to cope with inputs characterized by the presence of noise levels that may vary both among different input patches and within a single patch itself; and 3) capability to preserve high-frequency signal components in both the interferometric phase and in the coherence. In this way, we can guarantee good preservation and localization of distributed spatial patterns, edges, and point-like scatterers, typical of real InSAR data.

We design Φ-Net on the basis of the U-Net architecture [26] and exploit the concept of residual learning by mapping the input toward the output [18]. Accordingly, we replace the single U-Net layers with residual blocks (RBs) [22]. Note that similar architectural solutions, based on residual

U-Net, have been exploited in other applications, such as biomedical image denoising [27] and road extraction from remotely sensed optical data [28]. However, they have not been investigated for the specific and challenging task of InSAR parameter estimation. The choice of designing Φ-Net starting from a U-Net architecture resides in the fact that U-Net-based networks are able to preserve both local context information and fine structures and textures of images. On the one hand, the encoder, through the cascade of convolution and max-pooling layers, squeezes the input features in fewer spatial samples that carry the information about the context and, thus, on the local noise level. On the other hand, the decoder, while reconstructing this information by upsampling and convolution operations, also includes additive features that are copied from the encoder stage. This operation is done through skip connections, which allows the network to preserve most of the structural information of the input signal.

In the design of Φ-Net, we also address a very crucial issue in deep neural networks: the creation of a large, various, and reliable training data set.[1] Here, we propose a strategy by considering a large variety of cases that can be found in real InSAR data. In particular, we take into account several variables related to the acquisition of InSAR data, including the SAR system parameters, the acquisition geometry, and the land cover.

This article is structured in seven sections. Section II describes the basics of the adopted methodology, starting from the analysis of related works on DL approaches for denoising. Section III defines the used InSAR signal model and all the involved input and output variables and their handling. It also details the proposed CNN architecture, justifying the design choices and its implementation. Then, Section IV extensively describes the choices made for the generation of an effective synthetic data set for training the proposed CNN, discussing the crucial aspects that have been considered. Section V details all different steps of the network training phase. We then compare the performance of the proposed Φ-Net with SOA algorithms. In particular, Section VI presents and discusses the results using an independent data set of synthetic images, while Section VII investigates the performance on real TanDEM-X and Sentinel-1 InSAR data, showing that the proposed algorithm outperforms SOA methods. Finally, Section VIII discusses our findings and highlights open points and the next challenges related to the topic.

II. Background Concepts

In this section, we recall the basic principles of CNNs, which led to the adoption of the proposed architectural solution. In particular, we discuss both the main properties of deep residual learning and its application to image processing tasks, including image restoration.

A CNN is a machine learning architecture made from the cascade of several layers of learnable and nonlearnable operations, including convolutions. Pooling layers are often present in this kind of networks, in order to downsample the input feature by preserving only the largest value in a given window. The local nature of the convolution operation and pooling process (which often follows it) is a very intuitive way to extract the informative content from an image, by transferring the information about the position of the local feature into a higher semantic level. The joint use of fully connected (FC) layers increases the capability of the network to learn very complex nonlinear functions. For this reason, an FC layer is often exploited as the last layer of a CNN when the latter is used for image recognition purposes. An example is given by the CNN architectures AlexNet [29], VGG [21], and GoogLeNet [30], which are still today SOA architectures for computer vision tasks. These networks have extraordinary capabilities in extracting high-level semantic features from an image and have been widely used, achieving an unprecedented level of accuracy. The key aspect that such networks share is the use of deeper architectures in order to provide a higher level of abstraction and, consequently, achieve better performance. Nevertheless, unlike shallow architectures, deeper networks frequently experience training performance losses [22]. This problem has often been ascribed to vanishing gradient phenomena, which takes place during the process of network optimization. Essentially, the update of the network's internal parameters is impaired and the learning stops. Even if this issue has already been addressed by normalization processes of the input data [31], [32] and batch [33], the phenomena are still present and need to be considered. Moreover, this aspect cannot be ascribed to overfitting either. Indeed, by stacking identity layers to an existing network, while the complexity of the network does not increase, the training performance may decrease anyway [22]. In order to overcome this issue, He *et al.* [22] suggested introducing identity connections between the different layers of the network. In particular, they proposed a novel building unit, the RB, formed by the cascade of two convolutional layers and a skip connection just before the second rectified linear unit (ReLU) activation function, as shown in Fig. 1. The identity connection is used to add the input features of the block to its output. Due to this principle, the RB will likely better reproduce an identity mapping with respect to a plain convolutional layer. Indeed, the RB optimizes the residual function $g(x) = F(x)+x$, and an identity mapping is simply obtained for $F(x) = 0$. This solution is similar to the gated skip connection proposed in [34], which, however, adds new hyperparameters to the network at the cost of an increased training complexity. Further configurations of RBs have been investigated in [35] and [36], showing that alternative implementations can lead to better results in terms of training accuracy and robustness, with respect to the network depth. Anyway, the most suitable RB depends on the specific field of application.

Shortcut connections have shown great potential to ease the training of the network and improve the optimization performance. Indeed, they can be found in other network architectures as well, designed with different functionalities and for other applications. For example, the skip connections introduced in the U-Net architecture [26] (which was originally developed for biomedical image segmentation) allow for better propagation of the input signal toward the output. Based on the architecture of a fully convolutional network

[1]Patent pending: European Patent, filing number 2020/017/E3202DE00.

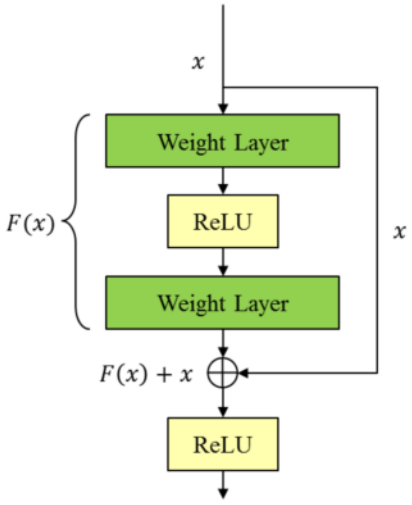


Fig. 1. RB operations.

(FCN) [37], the U-Net is formed by the cascade of an encoder and a decoder stage. The former is the actual architecture of an FCN and is used to compress the spatial information of the input into a higher number of features characterized by a lower spatial dimension. Such features carry information about the image context, i.e., the semantic of the object in the image. Besides, the latter stage spatially expands the considered feature maps in order to provide a precise localization. The output of the network is a semantic segmented map of the same size of the input. The skip connection is used in this context to copy the output feature maps of one convolutional layer (of the encoder) and concatenate it to the feature maps in input to the corresponding layer of the same spatial dimension (of the decoder), skipping, in this way, several convolutional layers. In [26], it has been shown that this mechanism is able to improve the capability of the network to preserve the spatial information by increasing the localization accuracy of the target objects.

Conceptually different and introduced for different purposes, the basic residual connection (identity mapping) and the U-Net skip connection are substantially the same copy operation, even though with a slightly different meaning. Indeed, while the residual connection generates at its output a residual function that is used for the optimization of the network, classical skip connections are normally used to concatenate feature maps related to different networks' layers. Moreover, the concept of residual DL has been further developed in [38] with the residual of residuals (RoR) network. This approach achieves better training performance by mutually connecting RBs in order to form bigger ones.

The residual learning concept has been applied to the denoising of natural images for the very first time in [18], outperforming SOA algorithms. Based on the VGG architecture [21], this network simply adds a unitary mapping of the input toward the output. By subtracting the estimated output to the input, the network is optimized to reproduce the input noise rather than the cleaned input signal. Similar to this approach, the residual network architecture proposed in [27] modifies the U-Net architecture by adding a shortcut residual connection between the input and the output in order to retrieve the input noise according to [18].

Following the aforementioned works, we set up the development of our network architecture starting from the residual U-Net and further extend the residual concept to each of the network layers so that each layer is composed of a single RB. The detailed description of the proposed CNN is presented in Section III.

III. Φ-Net

In this section, we detail the proposed methodology for the estimation of the interferometric phase and coherence. First, we present all involved quantities: the used signal model and the handling of input and output data. Second, we present the architecture of Φ-Net together with its main implementation details.

A. Interferometric Signal Model

For statistically modeling the interferometric SAR signal, we assume the SLC being a complex circular Gaussian variable, according to the statistical description of coherent radar measurements of [2] and, as exploited in [3], for the statistical description of the interferometric signal.

The complex interferogram Γ is defined as the complex conjugate product of the two SLCs (z_1, z_2)

$$\Gamma = z_1 z_2^*. \tag{1}$$

We can write the complex interferogram as a function of its noise-free parameters by considering the following notation [15], [16]:

$$\begin{pmatrix} z_1 \\ z_2 \end{pmatrix} = T \begin{pmatrix} u_1 \\ u_2 \end{pmatrix} \tag{2}$$

where $(u_1, u_2)^{\mathrm{T}}$ is a vector of two standard circular Gaussian random variables. T is the Cholesky decomposition of the data covariance matrix and can be expressed as a function of the InSAR noise-free parameters as

$$T = \begin{pmatrix} A & 0 \\ A\rho e^{-j\phi} & A\sqrt{1-\rho^2} \end{pmatrix} \tag{3}$$

where A is the amplitude (which is assumed to be equal for both SLCs), ϕ is the interferometric phase, and ρ is the coherence.

We further define the normalized complex interferogram as

$$\gamma = \frac{\Gamma}{A^2}. \tag{4}$$

The observed normalized complex interferogram can, therefore, be written as the sum of the noise-free normalized interferogram γ_0 and the noise term n as

$$\gamma = \gamma_0 + n \tag{5}$$

where γ_0 depends on the noise-free interferometric phase and coherence only as

$$\gamma_0 = \rho e^{-j\phi} \tag{6}$$

and n is a zero-mean signal-dependent contribution, which can be expressed as

$$n = \rho e^{j\phi}(|u|^2 - 1) + \sqrt{1-\rho^2}u_1 u_2^* \tag{7}$$

and is characterized by the following covariance matrix:

$$C_n = \frac{1}{2}\begin{pmatrix} 1+\rho^2\cos(2\phi) & \rho^2\sin(2\phi) \\ \rho^2\sin(2\phi) & 1-\rho^2\cos(2\phi) \end{pmatrix} \tag{8}$$

being the terms on the main diagonal the variances of the real and imaginary parts of the noise and the off-diagonal term their mutual correlation.

B. Input Data and Decorrelating Transform

In this section, we detail all the aspects concerning the handling of the input data before the application of the Φ-Net. From a pair of SLCs, we first compute the normalized complex interferogram as in (4). Therefore, we estimate the amplitude A by averaging the intensities of the two SLCs and then applying the maximum likelihood estimation of the amplitude, computed within a 2-D window with size $w \times w$ (in this case $w = 3$)

$$\hat{A} = \sqrt{\sum_{w\times w} \frac{|z_1|^2 + |z_2|^2}{2}}. \tag{9}$$

Given that the interferogram is a complex number, we apply our estimation process on its real and imaginary parts γ_{R} and γ_{I} separately, where

$$\gamma_{\mathrm{R}} = \mathcal{R}\left(\frac{\Gamma}{\hat{A}^2}\right) \quad \gamma_{\mathrm{I}} = \mathcal{I}\left(\frac{\Gamma}{\hat{A}^2}\right).$$

From the noise model presented in Section III-A, we know that the real and imaginary parts of the complex interferogram are mutually correlated. Joint processing of these two quantities is required in order to exploit this property in the estimation process. This requirement could be handled by using complex-valued neural networks that are a relatively new and less investigated field of research. As already done in [15] and [16], we resort to the use of a decorrelating transform prior to the actual processing. This solution, even if suboptimal, has already been applied in a large variety of image processing problems with good results. In our case, it allows us to separately process the real and imaginary parts of the complex interferogram and, therefore, to build our architecture on established network models already successfully applied to different research fields.

We then partition the input data in 64×64 pixels overlapping patches with a stride of 8 pixels along both rows and columns, as further detailed in Section V. For each input patch, we apply the Karhunen–Loève transform D

$$x = \begin{pmatrix} a \\ b \end{pmatrix} = D\begin{pmatrix} \gamma_{\mathrm{R}} \\ \gamma_{\mathrm{I}} \end{pmatrix} \tag{10}$$

where a and b are the decorrelated quantities corresponding to the real and imaginary parts of the noisy normalized complex interferogram, respectively. These quantities are the actual input to the Φ-Net: $x = (a, b)$.

The decorrelating matrix D has the following expression:

$$D = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \tag{11}$$

and we resort to the maximum likelihood approach [4] for the estimation of the noise-free parameter ϕ over the whole patch.

C. Output Data and Aggregation

After the application of the Φ-Net to noisy data, we obtain the output feature map $F(x) = (\hat{a}, \hat{b})$. In order to retrieve the estimation of the real and imaginary parts of the normalized complex interferogram over the patch, we need to invert the transform in (10) as

$$\begin{pmatrix} \hat{\gamma_{\mathrm{R}}} \\ \hat{\gamma_{\mathrm{I}}} \end{pmatrix} = D^{-1}\begin{pmatrix} \hat{a} \\ \hat{b} \end{pmatrix}. \tag{12}$$

We then aggregate all the estimated patches to form a whole image with the same dimension of the input one. Particularly, we bring each patch back to its original position and average overlapping areas in order to avoid discontinuity in the reconstructed output image. Finally, the estimated interferometric phase and coherence are given by

$$\hat{\phi} = \arctan\left(\frac{\hat{\gamma_{\mathrm{I}}}}{\hat{\gamma_{\mathrm{R}}}}\right) \quad \hat{\rho} = \sqrt{\hat{\gamma_{\mathrm{R}}}^2 + \hat{\gamma_{\mathrm{I}}}^2}.$$

D. Proposed Φ-Net Architecture

In this section, we present the architecture of Φ-Net, which is based on the U-Net model and optimized for the estimation of InSAR parameters. The proposed architecture is depicted in Fig. 2. Particularly, we remove one layer from the original U-Net and use a three-layer ($N_{\mathrm{L}} = 3$) implementation. The reason for this choice relies on the fact that we process small input patches (64×64 pixels) and that a shallower architecture is easier to train. Indeed, it has a smaller number of hyperparameters to be optimized and, thus, requires a lower number of input samples (training patches). The encoder is followed by a bridge layer and, finally, by a three-layer decoding path with skip connections. At each level, instead of applying a classical cascade of convolutions, we use the RB architecture represented in Fig. 3. Here, the input is connected to the output through a shortcut connection (1×1 convolution) in order to match the output dimensions with the input ones. The use of residual connections for each layer and the type of RB has been determined on the basis of the output performance by exploratory ablation experiments.

As input, we use a feature map x of size $[M \times M \times C]$, where M indicates the size of the 2-D input array [note that, for the sake of simplicity, in Fig. 2, we indicate only the first dimension of the input array (i.e., M)], while C identifies the number of input channels. As already mentioned in Section III-B, we exploit patches of dimension $M = 64$ and the real and imaginary parts of the normalized complex interferogram ($C = 2$) as input to our network. Fig. 2 depicts all the involved features (input, output, and the internally generated ones) as arrays with dimensions that visually reflect the actual feature dimension. The operations applied to each feature are indicated with continuous lines of different colors,

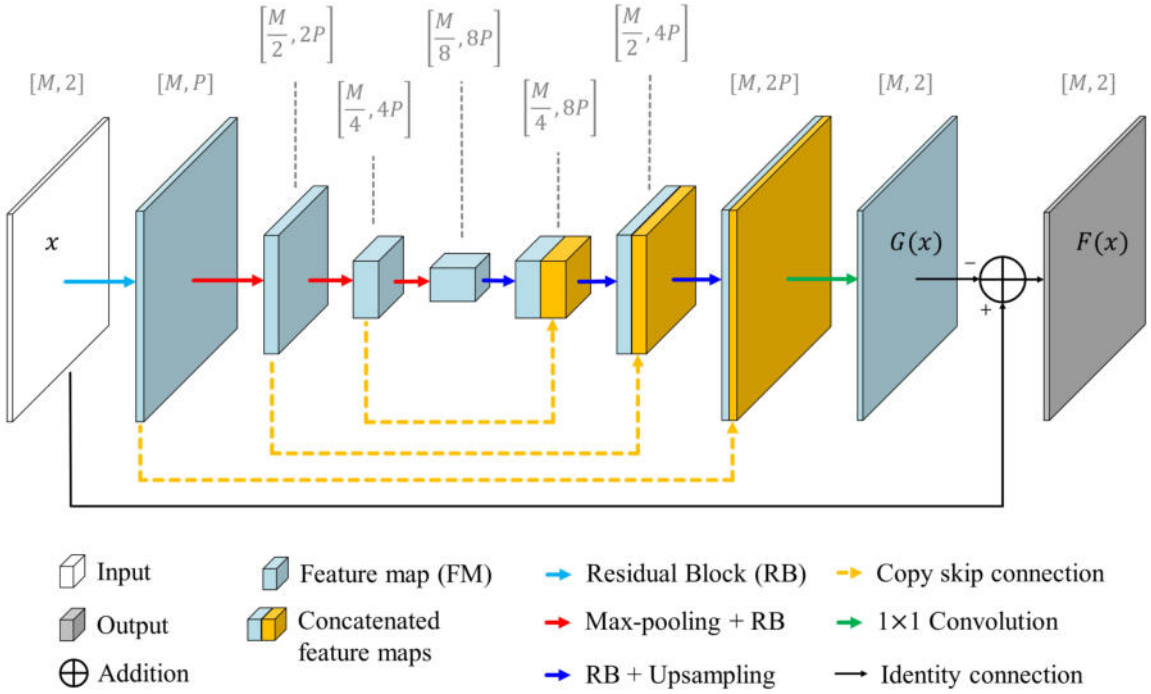


Fig. 2. Proposed residual network architecture. The output dimensions at each single layer are shown in gray brackets. For the sake of simplicity, the used notation implies that the first dimension (i.e., M) identifies a 2-D ($M \times M$) array. The internal structure of the RBs is detailed in Fig. 3.

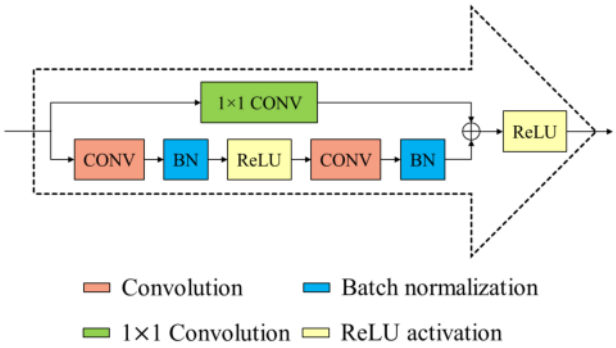


Fig. 3. Φ-Net RB operations.

which stands for the type of operation as reported in the legend. Skip connection operations, which copy the input from the encoder to the corresponding layer of the decoder, are instead indicated with orange dashed lines. The employed RB operations are the ensemble of convolutional blocks (CONV), batch normalization (BN), and ReLU activation functions, in the order depicted in Fig. 3. All convolutions of the RB have a kernel size of $N \times N$ (with $N = 3$), excluded the 1×1 convolution in the residual connection branch (upper one), while the number of output layers is set for all to a multiple of $P = 64$. The encoder is, therefore, formed by the cascade of an RB and 2×2 nonlearnable max-pooling operations. In particular, we indicate with a light-blue line the RB operations that from the two input features, x generates 64 output features, while we use red lines to indicate the cascade of max-pooling and RB operations. In this way, at each layer of the encoder, the number of features is doubled, achieving a maximum of $2^{N_L} P = 512$ features at the end of the encoding path (bridge), while the size of each feature map is $M/2^{N_L}$, thus $[4 \times 4]$. Practically, the input patches ($x = (a, b)$) are transformed into one narrow array with a smaller spatial dimension and larger depth. This new quantity (formed at the bridge layer) codifies the information about the local noise level and the type of detail contained in the input patches (e.g., smooth variations or steps).

The decoding path is a mirrored version of the encoding one, where, at each level, we use a bilinear upsampling followed by the RB operation (dark blue line). In this way, we expand the number of features to the corresponding one at the related layer of the encoder. Moreover, at each layer of the decoding path, we concatenate each feature map with the corresponding one in the encoding path (skip connection). This procedure decodes the information (present at the bridge layer) into spatially wider arrays, reproducing the original patch dimensions. With the help of skip connections, the original fringe structure of the interferometric phase and the coherence patterns can be preserved, leading to higher estimation accuracy and better resolution preservation.

After the last decoding layer, we use a 1×1 convolution to obtain $C = 2$ output channels $G(x)$, representing the noise superimposed to the complex input. Finally, the estimated real and imaginary parts of the normalized complex interferogram $F(x)$ are given by the subtraction of $G(x)$ from the original input x.

E. Loss Function

For the optimization of the network parameters, we exploit an L2-norm (L_2) loss function, together with a regularization

term (R), which penalizes output values outside the interval $[-1, 1]$. Note that, different from the usual L1-norm regularization, the term R only applies to the output if its absolute value exceeds the unit value. The reason for this choice relies on the fact that the noise-free normalized interferogram has a unitary maximum modulus, different from the corresponding noisy version. We define the overall loss function L as

$$\begin{aligned} &L(a_0, b_0, \hat{a}, \hat{b}) \\ &\quad = L_2(a_0 - \hat{a}, b_0 - \hat{b}) + \lambda R(\hat{a}, \hat{b}) \\ &\quad = ||[a_0 - \hat{a}, b_0 - \hat{b}]||_2 + \lambda \cdot ||\max(0, |[\hat{a}, \hat{b}]| - 1)||_1 \end{aligned}$$

where a and b are defined in (10), while the subscript 0 and the hat symbol indicate the corresponding noise-free and estimated parameters, respectively. The term λ weights the regularization function with respect to the L2-norm and is empirically set after preliminary experiments. Note that since we train the network on decorrelated input features, we also transform the noise-free reference according to (10).

IV. Strategy for the Training Set Generation

A crucial aspect of the use of deep neural networks is the generation of a large and reliable training data set for the learning of the network. We address this task by exploiting simulated data only. For this purpose, we generate a data set of synthetic interferometric images by utilizing the signal model presented in Section III-A. Thus, the simulated noise depends on the choice of the set of initial noise-free parameters (amplitude, phase, and coherence), which is a key aspect for the generation of a training data set that can simulate typical behaviors occurring in the real InSAR data. In the definition of the training data set, we follow the guidelines proposed in [39], which are expected to be considered in order to generate a training set that represents the generality of the problem. Accordingly, we model the approximation and the detail components [39]. The former refers to the representation of the main background physical information, while the latter aims at completing the description of the physical problem by introducing realistic complex variations. In our case, we achieve the approximation component by considering the interdependence between the noise-free interferometric parameters. Then, we introduce the detail component by also taking into account realistic spatial variation patterns for each of them. A high-level scheme of the proposed approach is presented in Fig. 4.

Regarding the approximation component, different relationships among parameters are used to model real InSAR data. For example, a recurring scenario comprises an interferometric phase and amplitude that show weakly correlated patterns or no correlation at all, whereas such dependence is often present between coherence and amplitude.

For the detail component, each parameter can be assumed to present either smooth variation trends or abrupt changes, depending on the type of illuminated target. For example, the interferometric phase presents fringe patterns of different spatial density depending on the local terrain slope, while the amplitude and the coherence may show slowly varying

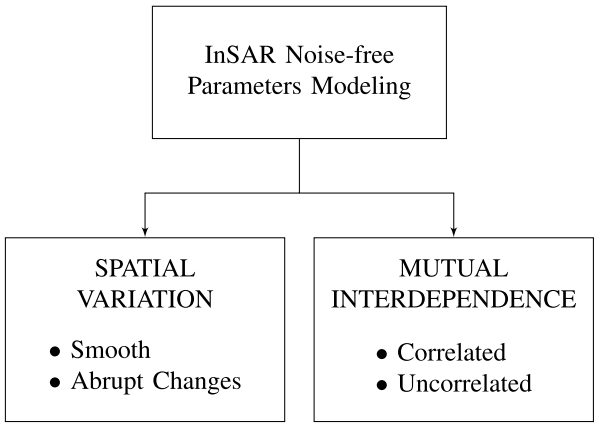


Fig. 4. High-level scheme of the proposed strategy for the generation of the training data set.

TABLE I
Values of Acquisition Parameters Used for the Simulation of Synthetic Phase Patterns

Acquisition parameters				
ID	Orbit height [km]	Swath dimension [km]	Look angle [degrees]	Height of ambiguity [m]
1227812	514	30×50	46.3	76.2
1228537	514	30×50	46.8	68.6

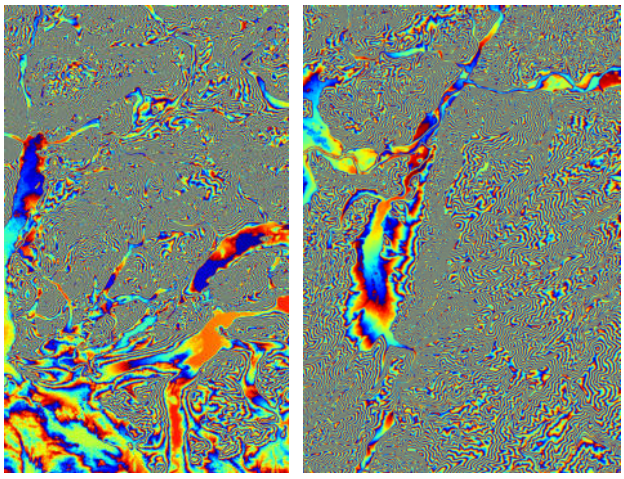
textures, as well as edges and tiny details, in a similar way as natural patterns that can be found in optical images.

In the following, we first extensively describe all used patterns for the generation of the noise-free amplitude, phase, and coherence images and, thereafter, motivate the choice of their mutual combination. Overall, we generate 600 images of dimension 256×256 pixels.

A. Interferometric Phase Patterns

For generating realistic phase patterns, we employ an actual InSAR acquisition geometry and further rely on real topographic information. In this specific work, we selected an area over the Austrian territory. We point out that a certain phase fringe pattern can be obtained by different combinations of acquisition parameters and surface topography. Therefore, in order to generate a large variety of data, we fix the geometry and variate the surface topography by selecting a wide area. We employ the geometry parameters of two TanDEM-X StripMap single-polarization acquisitions, as summarized in Table I, where, for each data-take, we report the orbit height, the swath dimension, the look angle, and the height of ambiguity (please refer to Section VII for more details on the TanDEM-X mission). We then consider an external DEM for the generation of the synthetic topographic phase. For this purpose, we choose the edited DEM from the Shuttle Radar Topography Mission (SRTM) at 30-m resolution [40]. This product is gap-filled and allows for the generation of synthetic interferometric phase images, which are not affected by the presence of phase inconsistencies caused by residual void pixels. We then back-geocode the DEM, retaining the topographic phase component only, while discarding the flat-Earth contribution caused by the side-looking geometry of

(a)

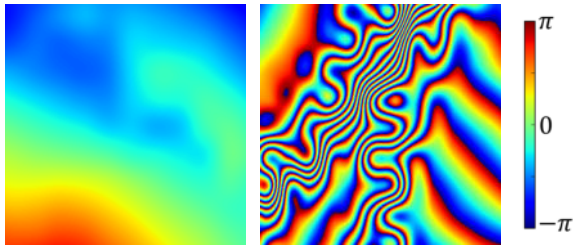


(b) (c)

Fig. 5. (a) Footprints of the two nominal TanDEM-X acquisitions used for generating the synthetic interferometric phase images, superimposed to GoogleEarth: (blue) acquisition date, August 12, 2014, AcqItemID: 1227812, and (red) acquisition date, August 29th, 2014, AcqItemID: 1228537. (b) and (c) Synthetic interferometric phase images associated with the two TanDEM-X footprints (blue) and (red), respectively.

SAR sensors. The synthesized interferometric phase images are shown in Fig. 5(b) and (c). The presence of both relatively flat and high-relief areas over the considered territory results in the generation of synthetic phase patterns characterized by the presence of low- and high-frequency fringes, respectively. Out of all available data, we form two subsets of 250 randomly selected patches of 256×256 pixels, characterized by the presence of fringe patterns at low and high frequencies. Two examples of synthetic interferometric phase patterns from the two subsets are depicted in Fig. 6.

In addition, we consider another variety of abrupt changes that may appear in interferometric images: phase steps. These steps may occur in the presence of very strong scatterers (e.g., artificial surfaces) that normally show high coherence values. We, therefore, simulate phase steps superimposed to flat fringe patterns, by considering 100 additional low-frequency phase patches. The strategy for the generation of this patch subset is extensively explained in Section IV-C.



(a) (b)

Fig. 6. Example of synthetic interferometric phase patterns generated from SRTM DEM data. (a) Low-frequency fringes (corresponding to relatively flat terrain). (b) High-frequency fringes (identifying high-relief terrain).

B. Amplitude and Coherence Patterns

As for the synthetic phase, we aim at generating smooth patterns and abrupt changes for both amplitude and coherence. On the one hand, the first behavior is simulated by considering a linear trend with low spatial variation. On the other hand, abrupt changes, as previously mentioned, may comprise a much larger variety of patterns, ranging from regular textures to isolated edges, straight, and curved lines, as well as small details of different shapes. Moreover, such patterns may have different orientations. Therefore, it becomes clear that the generation of an exhaustive synthetic data set from scratch is far away from being a trivial task.

In order to cope with this challenge, we rely now on real natural patterns extracted from remotely sensed optical images. In particular, we consider the data set named NWPU-RESISC45, which comprises 45 classes [41] and can be freely downloaded at [42]. Among all available images within the data set, we select a subset of 400 images equally chosen from each of the data set classes. Since we are only interested in the spatial pattern itself, we then compute the brightness from the colored optical images and use it for the generation of synthetic amplitude and coherence images. In particular, for both the slowly varying linear ramp and the natural pattern cases, we scale the optical brightness values to the intervals $[25, 255]$ and $[0, 1]$ for the amplitude and the coherence, respectively. Note that amplitude values are directly linked to the measured power of the backscattered radar signal. Since all SAR sensors are characterized by a defined noise floor (which is caused by the intrinsic system noise), an amplitude measurement greater than zero is always recorded. Therefore, we set the simulated amplitude values to a minimum value of 25.

We further point out that we use optical data in order to just extract shapes and contours that can be normally observed in nature, while we ignore the actual meaning of the optical measurements. A similar approach has been already used in the literature for the generation of speckled SAR images [23], [24].

C. Noise-Free Parameters Interdependence

Noise levels are driven by coherence and amplitude values according to the interferometric signal model presented in Section III-A. Therefore, in this section, we address the issue

of noise-free parameters interdependence in order to generate the noisy data set of complex interferograms.

In particular, we normally assume that the noise-free interferometric phase shows independent fluctuations with respect to noise-free amplitude and coherence trends. This is a reasonable assumption since the phase mainly depends on the scene topography rather than on the type of ground surface and land cover, as it applies instead for amplitude and coherence. We consider all noise-free parameters to be mutually dependent only in one case that will be illustrated later on in this section.

Amplitude and coherence might show some degree of interdependence according to the nature of the imaged scene. In order to better address this aspect, we consider the coherence factorization presented in [43] and extended in [44], where

$$\rho_{\mathrm{Tot}} = \rho_{\mathrm{Quant}}\,\rho_{\mathrm{Amb}}\,\rho_{\mathrm{Rg}}\,\rho_{\mathrm{Az}}\,\rho_{\mathrm{Temp}}\,\rho_{\mathrm{SNR}}\,\rho_{\mathrm{Vol}} \tag{13}$$

being ρ_{Tot} the interferometric coherence, and where the terms on the right-hand side account for the possible causes of decorrelation related to quantization (ρ_{Quant}), ambiguities (ρ_{Amb}), baseline decorrelation (ρ_{Rg}), relative shift of the Doppler spectra (ρ_{Az}), temporal decorrelation (ρ_{Temp}), thermal noise (ρ_{SNR}), and volume decorrelation (ρ_{Vol}). The first four terms are related to the specific sensor parameters and acquisition geometry, while the last three (ρ_{Temp}, ρ_{SNR}, and ρ_{Vol}) show a dependence on the specific characteristics of the illuminated scene on the ground and, therefore, on the kind of on-going backscattering mechanism. For this reason, in order to generate a meaningful training set, we simulate the possible relationships between coherence and backscatter by assuming the predominance of one among these last three decorrelation contributions at a time.

In particular, ρ_{Temp} identifies the decorrelation caused by changes in the illuminated scene occurred between the two SAR acquisitions forming the interferometric pair. In the case of bistatic InSAR data, $\rho_{\mathrm{Temp}} = 1$ since both master and slave images are simultaneously acquired. In repeat-pass InSAR, the amount of temporal decorrelation depends on several factors, such as the revisit time, the operative wavelength, and the type of land cover on the ground. The presence of temporal decorrelation might lead to independent amplitude and coherence patterns.

The term ρ_{SNR} identifies the amount of decorrelation due to thermal noise and can be written as

$$\rho_{\mathrm{SNR}} = \frac{1}{1 + \mathrm{SNR}^{-1}} \tag{14}$$

where SNR represents the signal-to-noise ratio of both master and slave acquisitions that are here assumed to be equal. This contribution is proportional to the level of backscatter on the ground and may, therefore, introduce a direct correlation between the image amplitude and the coherence patterns.

Finally, ρ_{Vol} identifies the decorrelation occurring because of volumetric scattering phenomena. This characterizes areas where the radar waves can penetrate into a volume, such as forests or snow- and ice-covered regions. In this case, the amount of decorrelation depends on the radar wavelength, the length of the orthogonal baseline, and the characteristics of the illuminated target on the ground (e.g., forest type and density or dielectric properties of snow) [45]. A direct relationship between backscatter and coherence decay is, therefore, not expected. As an example, one can consider a tropical rainforest, which is normally characterized by quite high and stable backscatter levels (e.g., -6.5 dB at the C-band [46] in the γ^0 projection [47]). As demonstrated by the bistatic TanDEM-X mission (which allows for the precise separation between ρ_{Vol} and ρ_{Temp}), a certain amount of volume decorrelation occurs depending on the acquisition geometry [45].

From these observations, we can assert that realistic scenarios can be successfully simulated only if amplitude and coherence patterns are generated as both spatially dependent patterns and completely unrelated ones. Therefore, in Section IV-D, we describe the strategy developed for the creation of the synthetic data set, which properly combines the phase, amplitude, and coherence patterns presented in Sections IV-A and IV-B.

As a final component, we consider the possible influence that correlated amplitude and coherence patterns may have on interferometric phase values. Indeed, in real InSAR data, phase steps can often be observed in correspondence of coherence and amplitude steps. This behavior is typical of artificial surfaces, e.g., buildings, which presents high levels of coherence and amplitude, together with phase steps caused by the abrupt elevation change from the ground to the top of the building. In order to simulate this behavior, we first segment the coherence image for values above 0.6. We exploit a watershed technique and set two segmentation intervals: $0.6 < \rho < 0.8$ and $0.8 < \rho < 1$. We then associate a random constant phase jump to each segment, allowing for jumps up to $\pm\pi$. These phase jumps are extracted from a Gaussian distribution with zero mean and standard deviation $\sigma = \pi\sqrt{2}/6$. Indeed, one should note that a certain phase jump Δ_Φ might be the result of the sum of two adjacent ones Δ_1 and Δ_2, as schematized in Fig. 8(b). The selected σ assures that the 99.7% (3σ) of the jumps Δ_Φ lie within the $]-\pi, \pi]$ interval.

We then added the derived phase jumps to 100 low-frequency fringe patches, achieving phase patterns like those presented in Fig. 8(a).

D. Patterns Combination

By summarizing all the abovementioned analyses, we group together the amplitude, coherence, and phase patterns into six different categories, each one comprising 100 images. The types of amplitude and coherence patterns used for each case are presented in Fig. 7 (one column per case).

Concerning the phase patterns, we utilized the synthetic phase images presented in Section IV-A. In particular, we have the following.

1) *Case 1:*
   a) *Amplitude:* Left-to-right ramp.
   b) *Coherence:* Left-to-right ramp.
   c) *Phase:* 50 images with low-frequency fringes synthetic phase patterns and 50 images with high-frequency fringes synthetic phase patterns.

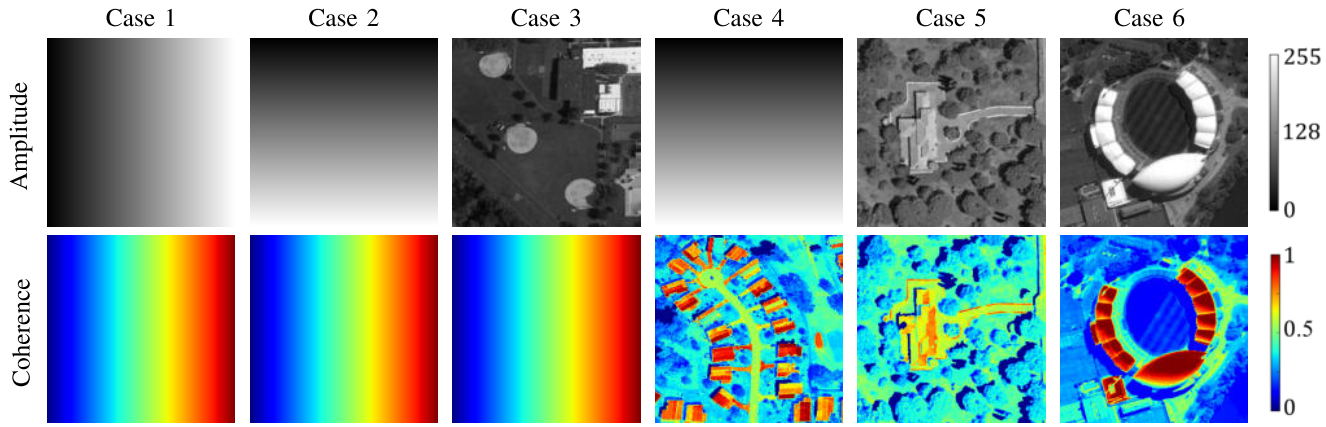


Fig. 7. Combination of amplitude and coherence patterns for each of the six cases that are considered for the generation of the synthetic data set. (Case 1) correlated ramps for amplitude and coherence, (Case 2) uncorrelated ramps for both quantities, (Case 3) natural pattern for the amplitude and ramp for the coherence, (Case 4) ramp for the amplitude and natural pattern for the coherence, (Case 5) natural pattern for both quantities, and (Case 6) natural pattern for both quantities.

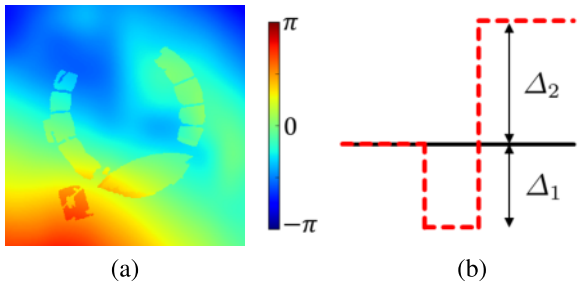


Fig. 8. (a) Step phase image associated with the amplitude and coherence in Fig. 7 (case 6). (b) Simplified plot describing the total phase jump resulting from the sum of Δ_1 and Δ_2.

2) *Case 2:*
   a) *Amplitude:* Top-to-bottom ramp.
   b) *Coherence:* Left-to-right ramp.
   c) *Phase:* As in Case 1.
3) *Case 3:*
   a) *Amplitude:* Natural patterns.
   b) *Coherence:* Left-to-right ramp.
   c) *Phase:* As in Case 1.
4) *Case 4:*
   a) *Amplitude:* Top-to-bottom ramp.
   b) *Coherence:* Natural patterns.
   c) *Phase:* As in Case 1.
5) *Case 5:*
   a) *Amplitude:* Natural patterns.
   b) *Coherence:* Natural patterns.
   c) *Phase:* As in Case 1.
6) *Case 6:*
   a) *Amplitude:* Natural patterns.
   b) *Coherence:* Natural patterns.
   c) *Phase:* Low-frequency patterns with steps, as shown in Fig. 8(a).

According to our analysis, the proposed combination of different patterns simulates most of the possible behaviors, which can be found in the real InSAR data. In particular, one can notice that amplitude and coherence can either be directly dependent as in Case 1 (where two left-to-right ramps are used) and in Cases 5 and 6 (where natural patterns are used) or independent as in Cases 2–4. From Cases 1–5, both low- and high-frequency patterns are equally shared: 50 patches are chosen from each subset. Case 6 represents a phase discontinuity that is complementary to fringe patterns. It is simulated by adding phase jumps to 100 flat phase patterns. For this reason, Fig. 9 shows one image only for this case.

Finally, we use all the described combinations of amplitude, coherence, and interferometric phase to derive the observed normalized complex interferogram, as previously described in (4). As already mentioned, noise levels depend on the amplitude and coherence values, and therefore, they vary according to the selected patterns in Fig. 7. An example of resulting noisy interferometric phase images is depicted in Fig. 9. Here, we combined the amplitude and coherence patterns presented in Fig. 7 (Cases 1–6) with the interferometric phase images at low- and high-frequency fringe patterns presented in Fig. 6. Please note that this is just an example meant to show the effects of different amplitude and coherence on the same phase image, but, in practice, different phase patterns are used for each combination.

V. Φ-Net Training Phase

Φ-Net has been trained starting from the data set created, as explained in Section IV, which has been split into two subsets, one for the training (90%) and one for the validation (10%), leading to two groups of 540 and 60 images, respectively. The data set is equally partitioned among all six cases, resulting in a validation data set composed of ten images for each case.

From the 256×256 pixels training images, we extract 64×64 pixels patches with a stride of 8 pixels. No data augmentation is used since the generated training data set already comprises a large variety of features at different scales and rotation angles. In order to avoid feeding the network with exactly the same input data because of overlapping patches,

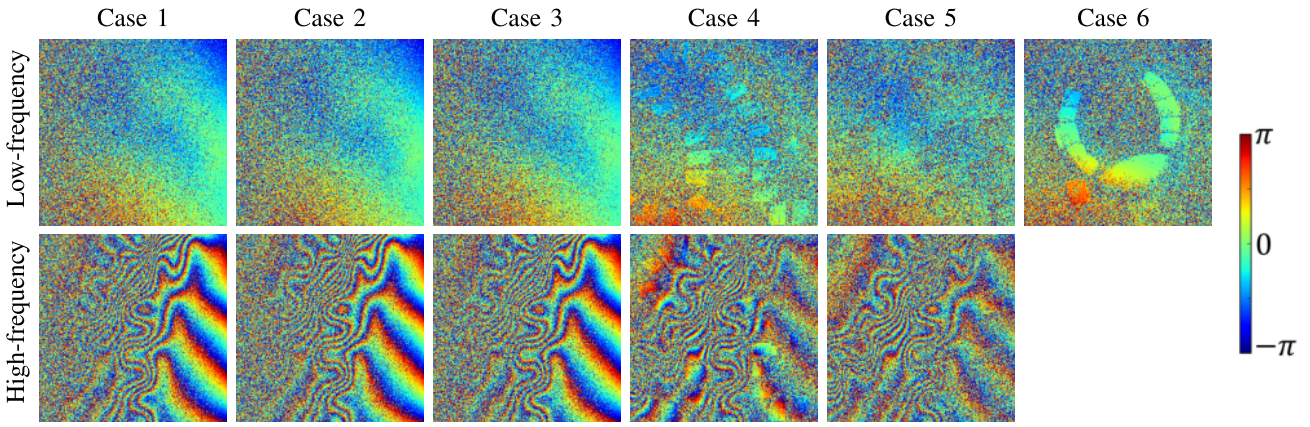


Fig. 9. Noisy phase images resulting from the combination, for each case, of the corresponding amplitude and coherence patterns shown in Fig. 7 with the two exemplary synthetic phase images (characterized by low- and high-frequency fringes, respectively) depicted in Fig. 6.

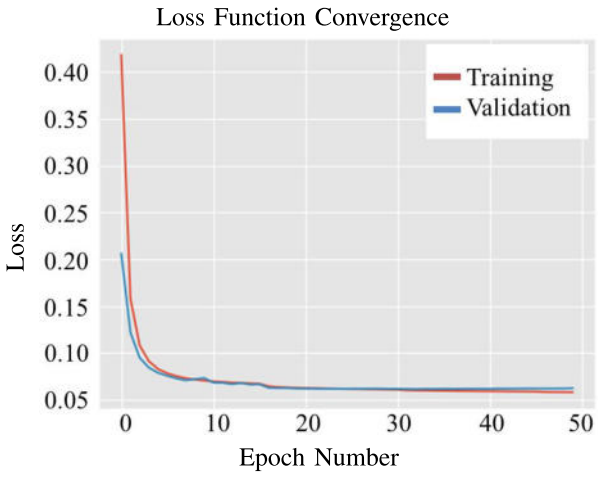


Fig. 10. Loss function convergence graph for training (red) and validation (blue) using overall 50 epochs.

we rotate each patch of 90° with respect to the neighboring ones and switch the phase sign every four patches. In particular, considering the whole data set, we generate, in total, $N_{\mathrm{T}} = 337\,408$ patches for the training and $N_{\mathrm{V}} = 37\,376$ patches for the validation. We then group them into minibatches of dimension $B_{\mathrm{s}} = 128$, leading to $N_{\mathrm{it}} = N_{\mathrm{T}}/B_{\mathrm{s}} = 2636$ iterations per epoch, which also corresponds to the number of times that the network weights are updated for each epoch.

After preliminary experiments, we fix the number of epochs at $N_{\mathrm{e}} = 50$. Indeed, around 50 epochs both the training and the validation losses converge. In particular, we notice from Fig. 10 that, while the training loss slightly decreases, the validation loss remains constant. In order to avoid overfitting, we stop the training at 50 epochs. The corresponding overall number of training iterations is, therefore, given by $Tot_{\mathrm{it}} = N_{\mathrm{e}} \times N_{\mathrm{it}} = 1\,31\,800$.

We then apply the decorrelating transform to the input data, as explained in Section III-B. This practically consists of a phase rotation that is constant all over the patch. The same operation is also applied to the noise-free data that are used for the computation of the loss function. The weight given to the regularization function R is set according to preliminary experiments to $\lambda = 10^{-2}$.

The network has been optimized by using the Adam algorithm [48] with an initial learning rate of $l_{\mathrm{r}} = 10^{-4}$, which has been decreased every 15 epochs by a factor [10, 20, 30]. Moreover, it has been implemented in Python 3.7 using Keras 2.2.4 framework and has been trained on 6x TeslaV100 GPUs (with 32-GB RAM each) for a total time of about 3 h.

VI. Experimental Results: Synthetic Data

We compare the proposed Φ-Net with the most recent SOA algorithms: SpInPhase [10], NL-InSAR [12], and OC-InSAR-BM3D [16]. All the algorithms parameters are set according to their original implementation in order to guarantee their robustness with respect to a large set of data. We further consider the ML estimator in [4] with a 5×5 boxcar window size in order to have a low-resolution baseline, which is used for the visual inspection of the results. In particular, for each method, we consider the following settings:

1) boxcar window with dimWindow = 5;
2) SpInPHASE with dimPatch = 10 and dictionaries of 512 atoms;
3) NL-InSAR with dimPatch = 7 and dimWindow = 21;
4) OC-InSAR-BM3D with dimPatch = 8 and dimWindow = 21.

In order to prove the validity of the proposed Φ-Net architecture, in Tables II, III, and V, we also present the numerical results for the original U-Net implementation.*

We test the network performance on the independent data set of synthetic images described in Section VI-A. It has to be noted that instead of InSAR-BM3D, we consider its offset-compensated version only since it has already been demonstrated that the latter performs either equally or better [16]. Likewise, we consider only NL-InSAR in place of its polarimetric extension (NLSAR [13]) since the former performs better on single-pair interferometric data [15], [16], [25], which is the case of this work.

*Note that, in this case, we adopt the original U-Net architecture, but we retain the proposed training strategy, including the training data set, the processing of the input data, and the loss function.

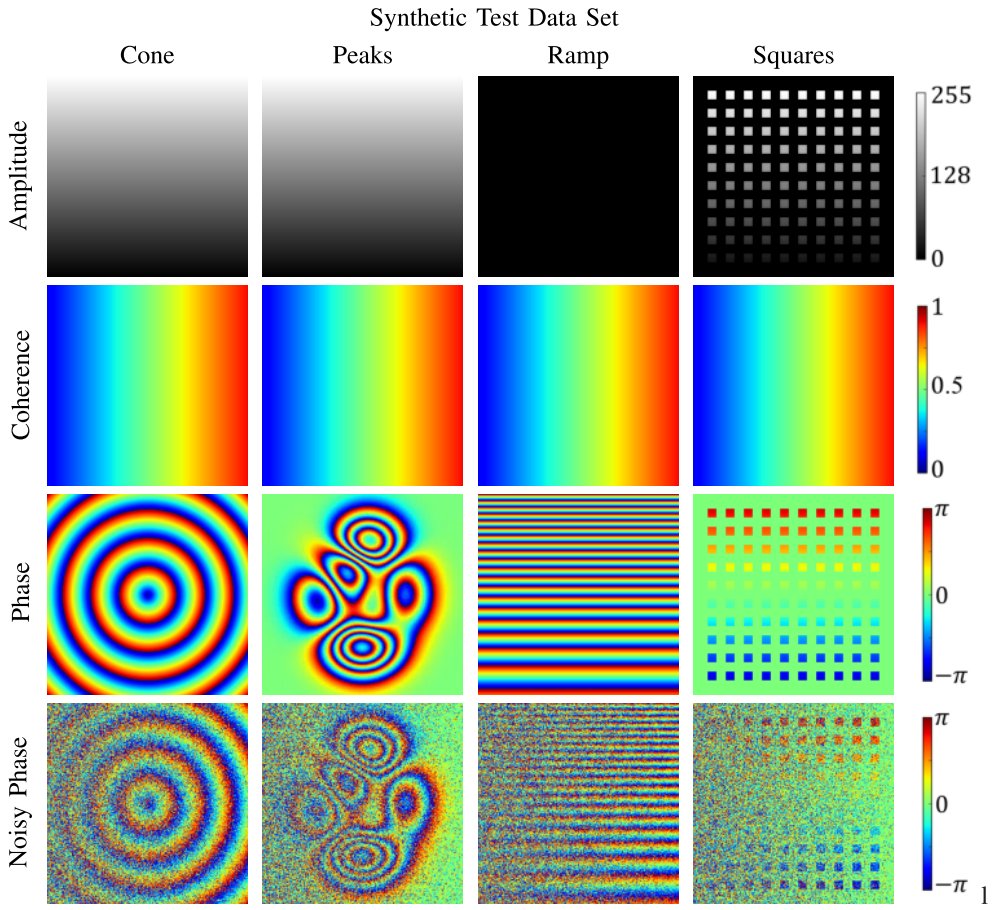


Fig. 11. Synthetic data set of images used for testing the Φ-Net performance. The amplitude values of the cone, peaks, and squares cases vary from 25 to 255, while, for the ramp case, it is set to a constant value of 25. The coherence values vary between 0.1 and 0.9 in all cases.

A. Definition of the Synthetic Test Data Set

In order to test Φ-Net on an independent data set, we select the one used in the recent works in [15] and [16].

The data set is shown in Fig. 11. Four different cases are considered.

1) *Cone:*
   a) *Amplitude:* Bottom-to-top increasing ramp (values between 25 and 255).
   b) *Coherence:* Left-to-right ramp (values between 0.1 and 0.9).
   c) *Phase:* Wrapped geometric cone.
2) *Peaks:*
   a) *Amplitude:* Bottom-to-top increasing ramp (values between 25 and 255).
   b) *Coherence:* Left-to-right ramp (values between 0.1 and 0.9).
   c) *Phase:* Series of wrapped peaks.
3) *Ramp:*
   a) *Amplitude:* Constant value equal to 25.
   b) *Coherence:* Left-to-right ramp (values between 0.1 and 0.9).
   c) *Phase:* Bottom-to-top oriented wrapped phase ramp with increasing fringe density. (Note that the amplitude is here kept constant in order not to interfere with the frequency variations of the phase fringes.)
4) *Squares:*
   a) *Amplitude:* Constant background (value set to 25) and small geometric squares with intensities decreasing top-to-bottom (values between 25 and 255).
   b) *Coherence:* Left-to-right ramp (values between 0.1 and 0.9).
   c) *Phase:* Abrupt phase jumps in correspondence of each square's borders.

For each combination of amplitude, coherence, and phase, we generated ten noisy images by using ten different random noise realizations (one of those is depicted in the last row of Fig. 11) so that, in total, we utilize 40 different images.

It is important to note that, even though similar patterns can be found in nature, the synthetic patterns considered in the test data set are not specifically used for training the network.

B. Results

The interferometric phase and coherence images, estimated in all four cases of the synthetic data set, and for all SOA methods, considering a single noise realization, are depicted in Figs. 12 and 13, respectively. Only three methods, namely,

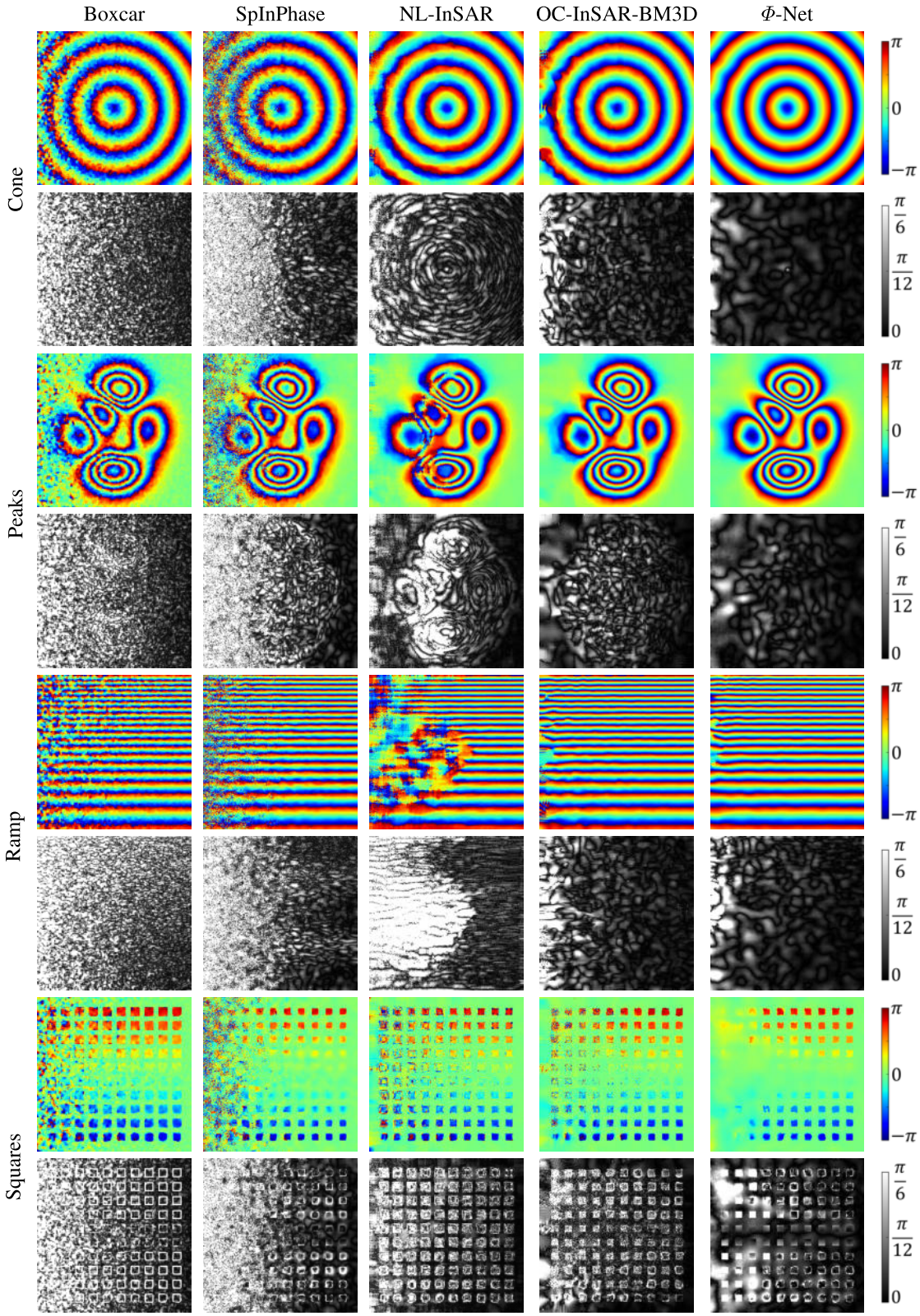


Fig. 12. Test on synthetic data: interferometric phase comparison between SOA methods and our proposed Φ-Net. The estimated phase images are color-coded, while the corresponding error matrices are in gray scale.

the boxcar, NL-InSAR, and Φ-Net, provide an estimation of the coherence, while the latter is not available for SpInPHASE and OC-InSAR-BM3D. Moreover, for each case, we also show the error map, corresponding to the absolute value of the difference between the estimated quantity and its corresponding clean reference. The error map is depicted between zero and

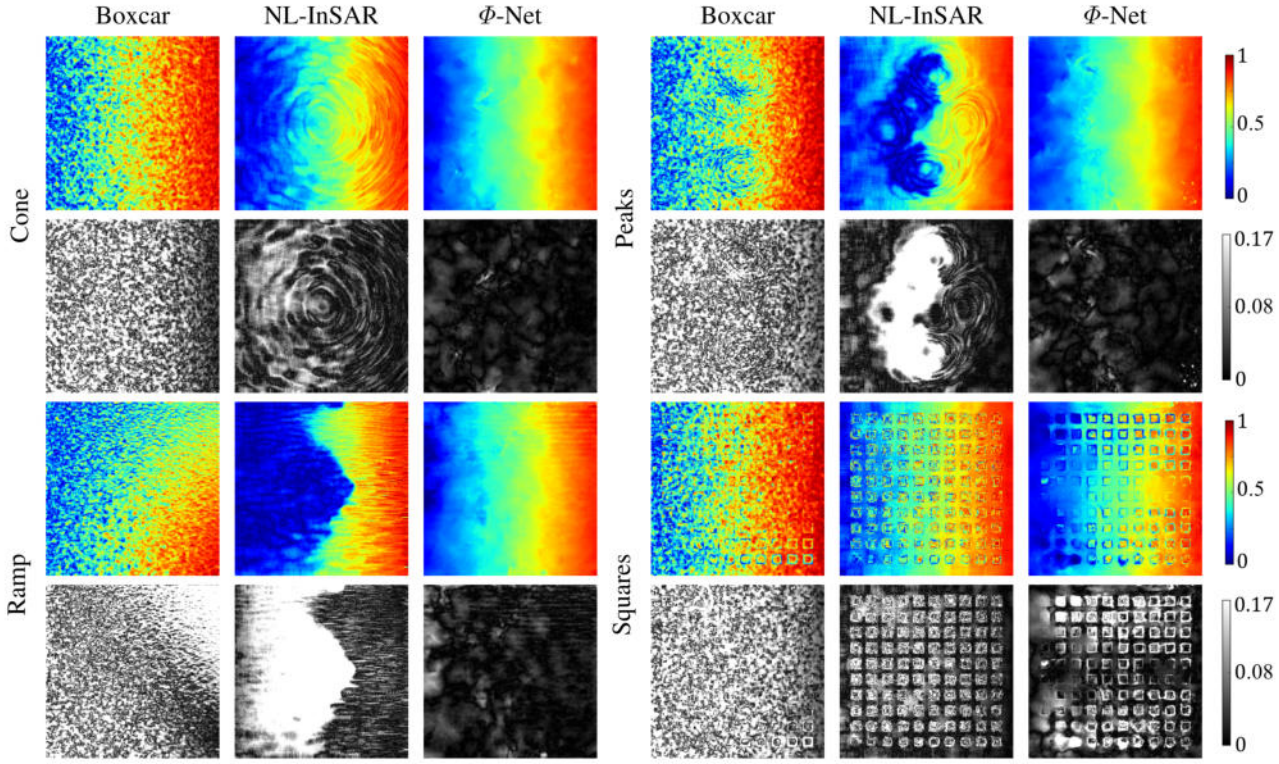


Fig. 13. Test on synthetic data: coherence comparison between SOA methods and our proposed Φ-Net. The estimated coherence images are color-coded, while the corresponding error matrices are in gray scale.

$1/6$ of its possible maximum value: $\pi/6$ and 0.17 for phase and coherence errors, respectively.

From the first visual inspection, it is clearly visible that, in all considered cases, the proposed Φ-Net architecture performs better than the other methods, being able to correctly reconstruct details also in the presence of strong noise levels, which has a severe impact on all other SOA methods.

Furthermore, Φ-Net is more effective than other techniques in the separation of the noise contribution from the underlying informative signal. Indeed, SOA methods show error maps that follow the fringe structure for both the phase and the coherence. This behavior indicates that part of the signal is considered as noise and consequently erased during the estimation process. At the same time, Φ-Net shows better performance for any fringe density, comprising very dense fringe patterns as well. This is a very critical aspect of SOA methods, in particular for the nonlocal patch-based method NL-InSAR. Indeed, it is affected by a performance drop in correspondence with phase patterns at certain fringe densities, which are determined by the employed patch and search window size. This behavior can be well observed from the Ramp data in Figs. 12 and 13, in which NL-InSAR is not able to reconstruct especially the patterns at medium fringe density. Differently, OC-InSAR-BM3D, due to its offset-compensation mechanism, is specifically designed to perform well on any fringe pattern and is, therefore, less sensitive to fringe density. Nevertheless, the proposed Φ-Net achieves a lower phase error with respect to OC-InSAR-BM3D for any noise level as well. Moreover, Φ-Net performs fairly well and still better than any other SOA method in the presence of phase steps as well (squares case of Figs. 12 and 13). From the phase image, we notice that the phase steps are very well estimated for high coherence values (right part of the image), presenting only small errors at the squares' borders, similar to SOA methods. For very low coherence values, instead, Φ-Net is not really able to discriminate between phase jumps due to the noise or to the underlying signal. Anyway, the error is not larger than the one of SOA algorithms.

These observations are supported by numerical results. For this purpose, we compute the following performance metrics.

1) The root mean square error (RMSE) between the estimated quantity and the reference one, calculated for both the interferometric phase and the coherence.
2) The total number of residues, i.e., the number of phase inconsistencies that are detected as nonzero values when computing a closed integral loop over four spatially adjacent pixels.
3) The cosine dissimilarity measure, defined as

$$\mathrm{CM} = \frac{1}{2N}\sum_{i=0}^{N-1}(1-\cos(\phi_i - \hat{\phi}_i)) \tag{15}$$

where N is the number of pixels of the image. This metric assumes values in the interval [0, 1], with zero indicating a perfect match between the estimated and the noiseless phase images.

For each simulated data, we compute the considered performance metrics over ten independent realizations, and

TABLE II

PERFORMANCE RESULTS ON THE SYNTHETIC DATA SET INTRODUCED IN SECTION VI-A. THE TABLE DISPLAYS THE MEAN (RMSE) AND STANDARD DEVIATION (STDDEV) OF THE RMSE FOR THE ESTIMATED INTERFEROMETRIC PHASE AND COHERENCE. THE LAST COLUMN SHOWS THE AVERAGE RMSE OF ALL THE FOUR SIMULATED DATA

RMSE on Interferometric Phase [rad]									
	Cone		Peaks		Ramp		Squares		Average
	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	
boxcar	0.5285	0.0087	0.5461	0.0131	0.6618	0.0076	0.7754	0.0097	0.6280
SpInPhase	0.7386	0.0079	0.7185	0.0135	0.6735	0.0113	0.6957	0.0106	0.7066
NL-InSAR	0.3158	0.0133	0.5165	0.0158	1.1465	0.0109	0.5503	0.0077	0.6323
OC-InSAR-BM3D	0.2725	0.0198	0.1728	0.0088	0.3249	0.0316	0.4668	0.0098	0.3093
U-Net	0.1275	0.0204	0.1196	0.0053	0.2742	0.0279	0.4035	0.0048	0.2312
***Φ*-Net**	**0.1007**	0.0122	**0.1085**	0.0066	**0.2124**	0.0111	**0.3915**	0.0096	**0.2033**

RMSE on Coherence [ ]									
	Cone		Peaks		Ramp		Squares		Average
	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	
boxcar	0.1413	0.0012	0.1407	0.0017	0.1625	0.0007	0.168	0.0016	0.1531
NL-InSAR	0.0646	0.0011	0.1456	0.0031	0.1932	0.0013	0.1214	0.0007	0.1312
U-Net	0.0251	0.0016	0.0249	0.0017	0.0305	0.0027	0.1155	0.0021	0.0490
***Φ*-Net**	**0.0123**	0.0014	**0.0148**	0.0014	**0.0192**	0.0016	**0.0945**	0.0015	**0.0352**

TABLE III

PERFORMANCE RESULTS ON THE SYNTHETIC DATA SET INTRODUCED IN SECTION VI-A. THE TABLE DISPLAYS THE MEAN AND STANDARD DEVIATION (STDDEV) OF THE NUMBER OF RESIDUES AND OF THE COSINE DISSIMILARITY MEASURE FOR THE ESTIMATED INTERFEROMETRIC PHASE. THE LAST COLUMN SHOWS THE AVERAGE OF ALL THE FOUR SIMULATED DATA

Number of Residues on Interferometric Phase [ ]									
	Cone		Peaks		Ramp		Squares		Average
	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	
boxcar	414.1	41.3	495.1	34.4	856.5	22.6	756.9	28.2	630.7
SpInPhase	2695.1	106.1	2572.1	154.4	2346.8	124.3	2233.3	142.3	2461.8
NL-InSAR	35.2	13.4	155.1	27.6	869.4	57.5	1118.5	42.3	544.6
OC-InSAR-BM3D	33	18.3	4.5	13.2	155.9	66.6	807.9	71.6	250.3
U-Net	**0**	(-)	**0**	(-)	16.1	6	10.2	3.2	6.6
***Φ*-Net**	**0**	(-)	**0**	(-)	**7.5**	2.5	**8.4**	4.7	**3.9**

CM on Interferometric Phase [ ]									
	Cone		Peaks		Ramp		Squares		Average
	Mean	StdDev	Mean	StdDev	Mean	StdDev	Mean	StdDev	
boxcar	0.0539	0.0014	0.0583	0.002	0.0827	0.0017	0.1064	0.002	0.0753
SpInPhase	0.0972	0.0016	0.0917	0.0029	0.0807	0.0024	0.0872	0.0022	0.0892
NL-InSAR	0.0226	0.0015	0.0491	0.0027	0.2005	0.0032	0.0558	0.0014	0.082
OC-InSAR-BM3D	0.0159	0.0019	0.0073	0.0006	0.0203	0.0031	0.0409	0.0016	0.0211
U-Net	0.004	0.0012	0.0035	0.0003	0.0149	0.0023	0.0332	0.0005	0.0139
***Φ*-Net**	**0.0025**	0.0006	**0.0029**	0.0003	**0.0094**	0.0008	**0.0316**	0.0013	**0.0116**

we provide both their mean value and standard deviation. Tables II and III provide a summary of the obtained results. In order to have an overall indicator for the performance of the methods, the last column shows the average metric values that are computed by averaging all single values reported in the other columns.

As one can observe, the proposed Φ-Net architecture performs better than all other methods in all four considered test cases, with all performance metric values always better than those of the best SOA method.

Focusing now on the estimated interferometric phase, one can note that the best SOA method is always OC-InSAR-BM3D, with an average RMSE of 0.3093 rad. On the other hand, the proposed Φ-Net achieves an overall mean RMSE of 0.2033 rad, 0.1 rad better than OC-InSAR-BM3D. Regarding the estimated coherence, NL-InSAR is the best SOA method for the cone and squares cases, while the simple boxcar outperforms it in the other two cases. Nevertheless, in all cases, the proposed Φ-Net shows a much better performance, with an overall mean RMSE of just 0.0352, which is more than four times better than NL-InSAR. Regarding the total number of residues, Φ-Net is able to provide a residues-free reconstruction in two cases, cone and peaks, showing significant improvement with respect to all SOA methods. The same applies to the cosine dissimilarity measure, with an average value almost twice better than OC-InSAR-BM3D. Furthermore, Φ-Net also performs better than the original U-Net architecture, by providing lower values for all the considered metrics. The improvement, with respect to the U-Net, is relatively small compared with the performance gain that is obtained with respect to SOA methods. This is expected as Φ-Net shares the same basic architecture of the U-Net. Therefore, in this case, the network architecture has a lower impact on the performance with respect to the data set used for training both networks. We can, therefore, assert that, in order to further improve the estimation performance of DL approaches, both the training data set and the network architecture should be jointly optimized.

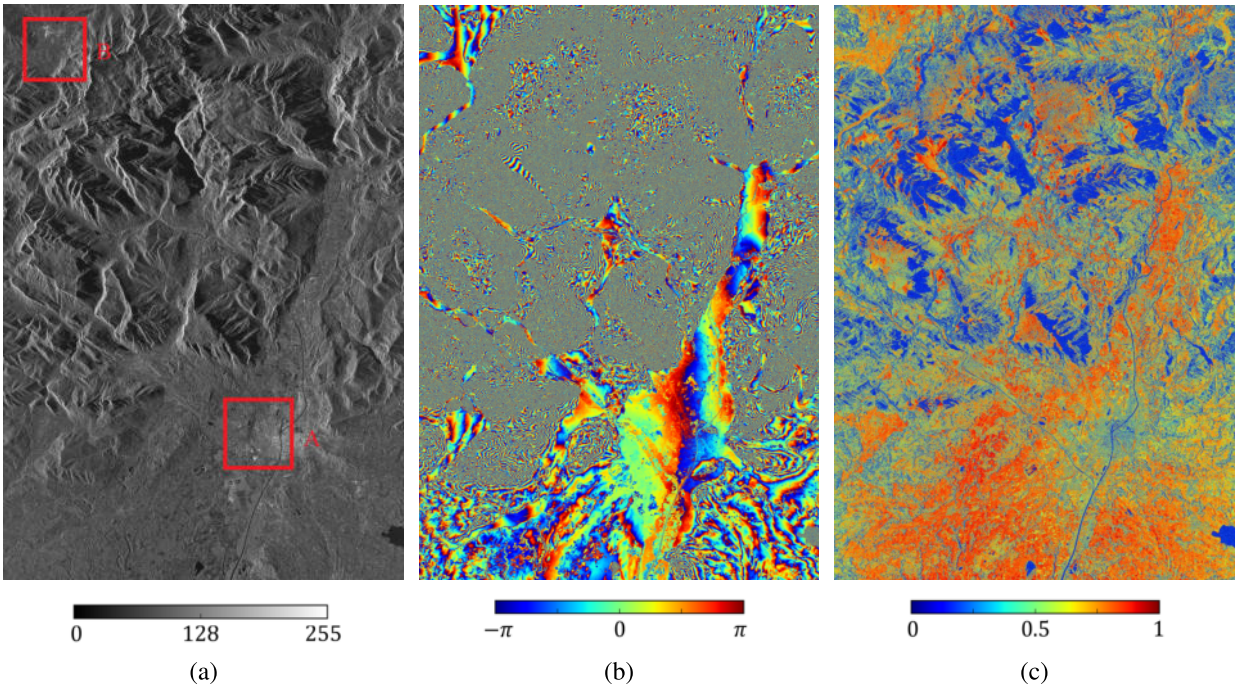


Fig. 14. TanDEM-X full-resolution data (StripMap mode, acquisition date May 23, 2015, AcqItemID: 1023484), used for testing the algorithm performance for the generation of high-resolution DEMs. (a) SAR amplitude, (b) estimated interferometric phase, and (c) estimated coherence obtained applying the Φ-Net. The red squares A and B in (a) identify the location of the two selected areas (macropatches), which are analyzed in detail.

TABLE IV

COMPUTATIONAL PERFORMANCE: HARDWARE (HW)/SOFTWARE (SW) SPECIFICATIONS AND COMPUTING TIMES FOR PROCESSING A SINGLE PATCH OF 256 × 256 PIXELS WITH ALL ANALYZED ALGORITHMS

Computational performance and HW/SW specification			
Algorithm	HW	SW	Time [s]
boxcar	4x Intel(R) Xeon(R) CPU E5-4650Q @ 2.70GHz, 512 GB RAM	Python	0.06
SpInPhase	1x Intel(R) Core(TM) CPU i7-7700HQ @ 2.80GHz, 16 GB RAM	C	155.10
NL-InSAR	1x Intel(R) Core(TM) CPU i7-7700HQ @ 2.80GHz, 16 GB RAM	C	84.22
OC-InSAR-BM3D	1x Intel(R) Core(TM) CPU i7-7700HQ @ 2.80GHz, 16 GB RAM	C	11.10
Φ-Net	1x Tesla V100 GPUs 32GB RAM 2x Intel(R) Xeon(R) CPU E5-2698 v4 @ 2.20GHz , 512 GB RAM	Python	0.32

Finally, Table IV provides the computing times that we measured for processing a noisy image of 256 × 256 pixels, together with the software language and the hardware details. Interesting to note is the difference between the best SOA method (OC-InSAR-BM3D) and our proposed Φ-Net. The much faster processing time of the Φ-Net makes it a very good candidate for future operational processing of large volumes of data.

VII. EXPERIMENTAL RESULTS: REAL DATA

In this section, we provide the results obtained when applying Φ-Net to real interferometric SAR acquisitions from both the TanDEM-X and the Sentinel-1 missions.

TanDEM-X Data: TanDEM-X is an on-going German spaceborne SAR mission, comprising the two X-band twin satellites TerraSAR-X and TanDEM-X, launched in 2007 and 2010, respectively [44]. They are currently flying in a close orbit formation, which allows for the acquisition of high-resolution bistatic InSAR data, characterized by the absence of temporal decorrelation. An example of the detected amplitude from the master image in a TanDEM-X bistatic acquisition is depicted in Fig. 14(a). The primary goal of the mission was the generation of a global high-resolution DEM with unprecedented accuracy, which has been successfully completed in 2016 [49].

In the light of this experience, we test the proposed Φ-Net on real TanDEM-X bistatic InSAR data for two main purposes: 1) to compare its performance with SOA methods on real InSAR data and 2) to explore the potential of TanDEM-X data for the generation of high-quality DEMs, with a resolution close to the one of the original SLC product.

As input InSAR data, we utilized one TanDEM-X bistatic acquisition acquired in StripMap mode over the area of Salzburg, Austria, on May 23, 2011 (AcqItemID: 1023484), with a height of ambiguity of 48.31 m, as depicted in Fig. 14. The interferometric phase and coherence images, generated by applying the proposed Φ-Net, are shown in Fig. 14(b) and (c), respectively. Also, in this case, we have the availability of both flat and high-relief regions, together with urban areas, which allows us to test the algorithms' performance over different kinds of terrain.

Sentinel-1 Data: The Sentinel-1 mission operates at the C-band and is formed by a constellation of two satellites (Sentinel-1a and -1b), which fly on the same nominal orbit with an angular shift of 180°. They provide repeat-pass interferometric data at six-day revisit time. The Sentinel-1 mission has been designed for the main purpose of differential interferometry applications. Therefore, with respect to TanDEM-X, it typically provides InSAR data with a smaller spatial baseline

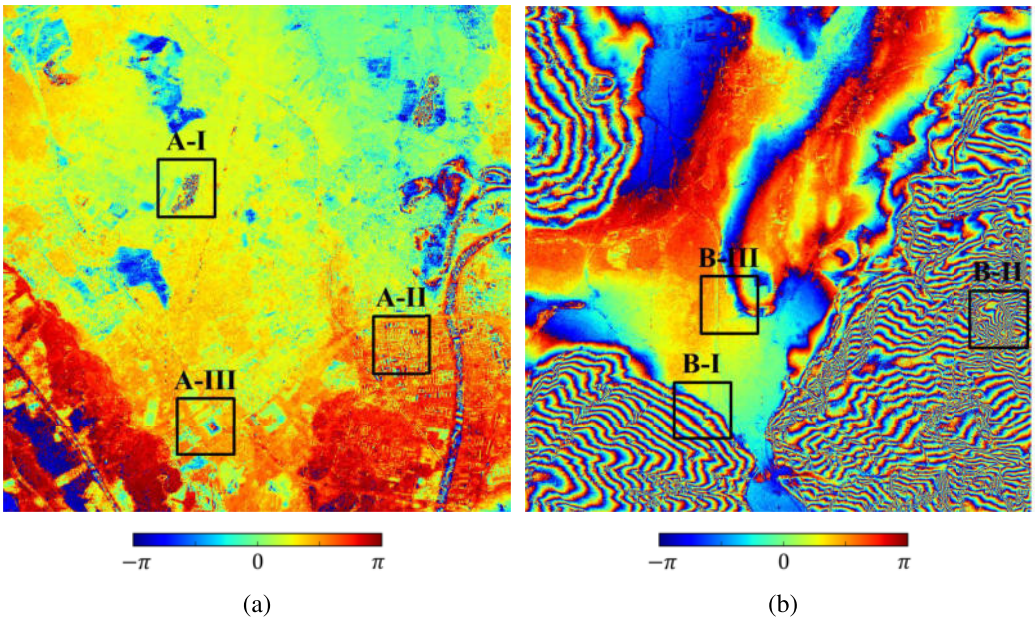


Fig. 15. Detail of the analyzed macropatches A and B introduced in Fig. 14. (a) Macropatch A, showing the urban area surrounding the Salzburg Airport and (b) macropatch B, depicting a mountainous region. Within each macropatch, three smaller patches of 300×300 pixels are identified.

and a larger temporal one. For these reasons, compared with TanDEM-X, Sentinel-1 interferograms are characterized by less dense fringe patterns (due to both the smaller spatial baselines and the use of the C-band instead of the X-band), and the coherence is normally lower due to the presence of temporal decorrelation (six-day revisit time).

A. Analysis on InSAR Parameters Estimation

For performing the first task, we select a collection of patches from the full interferogram of Fig. 14. We first identify two specific areas (A) and (B), as shown in Fig. 14(a), which are characterized by (A) urban area (surrounding the town of Salzburg) and (B) vegetated and mountainous environment. For each of these areas, depicted in Fig. 15, we then select three smaller patches of 300×300 pixels, which are significant for assessing the quality of the estimation over different types of targets.

The estimated interferometric phases and coherences are shown in Figs. 16 and 17, respectively. Again, the coherence is available for a subset of methods only, namely boxcar, NL-InSAR, and Φ-Net.

Regarding the interferometric phase, the first row of Fig. 16 shows the optical picture of the area, taken from GoogleEarth. The corresponding noisy interferograms at full resolution (3 m) are depicted in the second row, while each column corresponds to a different patch, in particular the following.

1) *A-I:* Urban area surrounding the Salzburg Airport, characterized by the presence of man-made flat surfaces (runway and airplanes parking slots).
2) *A-II:* Typical complex-structured urban area with buildings and streets ordered in a geometric-like plan.
3) *A-III:* Urban area including the Salzburg Red Bull Soccer Arena.
4) *B-I:* Mountainous terrain with almost constant moderate slopes.
5) *B-II:* Extremely high-relief and complex mountainous terrain.
6) *B-III:* Agricultural area, characterized by the presence of a river and cultivated fields.

From the first visual inspection, it can be inferred that the proposed Φ-Net (last row) has very powerful denoising capabilities on real InSAR data as well. We can observe that, contrary to other methods, Φ-Net provides a strong noise suppression. Indeed, it is possible to note in A-I how the extension of the incoherent noisy area (characterized by flat man-made structures) is significantly reduced, while, in B-I and B-II, the very dense interferometric fringe patterns (associated with mountainous areas) appear cleaner and smoother. Moreover, Φ-Net is able to preserve the phase patterns details as clearly visible in A-II, A-III, and B-III. On the contrary, estimation methods, such as NL-InSAR and OC-InSAR-BM3D, which also preserve the signal resolution, show a much noisier result, as confirmed by a closer analysis of the Red Bull Arena in A-III and the river path in B-III. This last observation is coherent with the results obtained on synthetic data. Indeed, Φ-Net is able to better estimate phase steps, showing good detail preservation and, at the same time, a strong noise suppression. SpInPhase shows very good results in the presence of moderate slope fringes even though it seems that a consistent residual noise persists, especially for high noise power. Moreover, it is interesting to note that both Φ-Net and OC-InSAR-BM3D are able to preserve details over homogeneous areas that, otherwise, disappear for the other methods. This behavior is especially observed for NL-InSAR, which severely smooths homogeneous phase regions and creates phase steps in place of slowly varying phase ramps. This behavior is typical for nonlocal-means-based approaches and is normally indicated as the staircasing effect [50]. Even though the result is appealing for a visual inspection of the interferogram, it is often not accurate, as already shown in [15]. A similar result

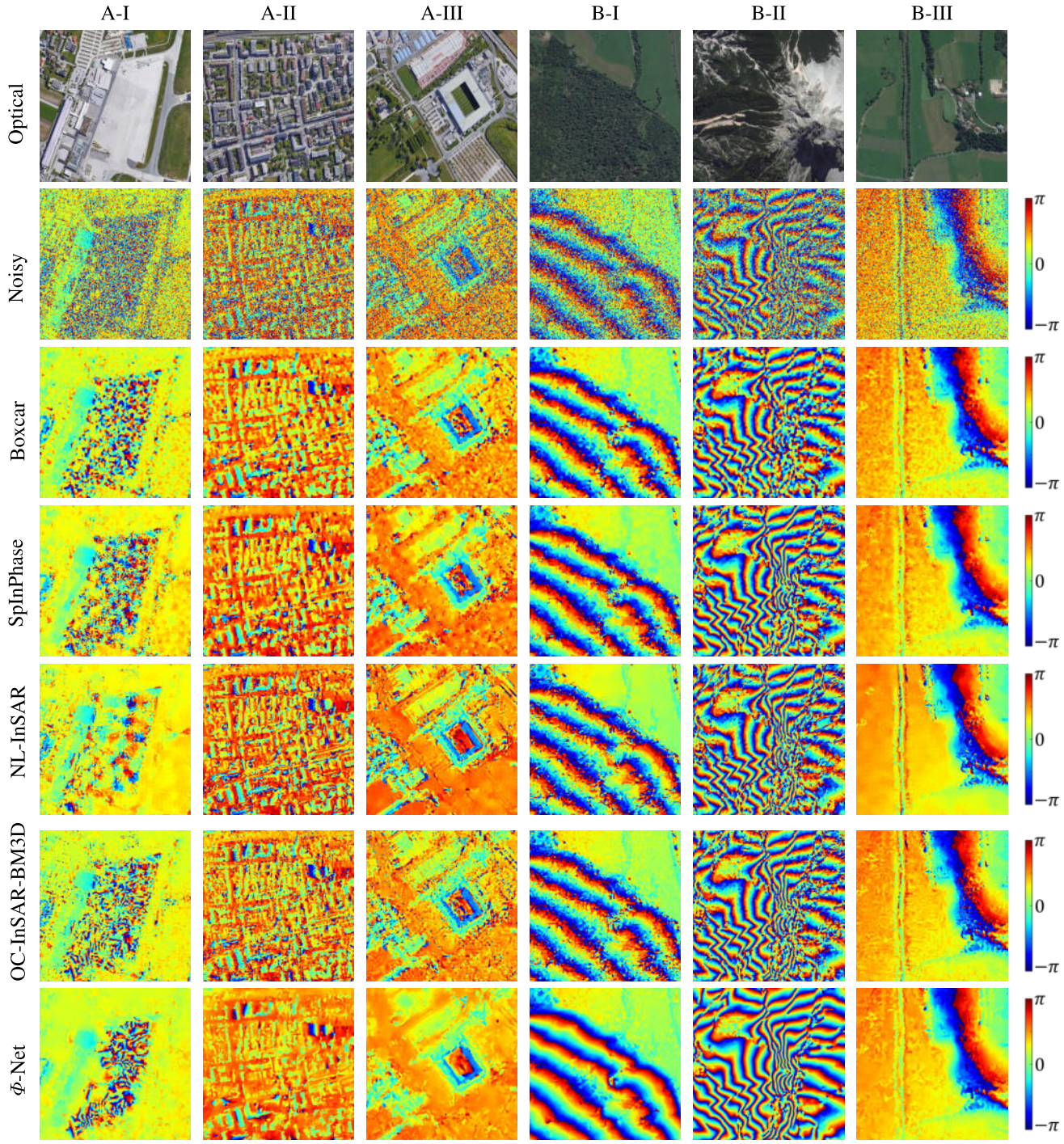


Fig. 16. Interferometric phase images of the considered TanDEM-X patches from Fig. 15, estimated using SOA methods and our proposed Φ-Net. The first and second rows show the optical image from GoogleEarth and the noisy interferogram at full resolution (3 m), respectively.

is observed for the coherence as well. Indeed, as a first impression, Φ-Net estimates appear noisier than NL-InSAR ones since NL-InSAR tends to smooth slow-varying textures and preserve abrupt changes. However, such variability could actually be the effect of high-resolution details preservation. Indeed, single-point-like coherent targets are better preserved by Φ-Net. This is visible in A-I, where single targets over the flat low-coherence area can be better separated than for the other methods. Moreover, Φ-Net is able to better estimate the coherence of homogeneous surfaces with respect to NL-InSAR. This can be noticed in B-II, where the coherence estimate suffers less from the errors induced by dense fringe patterns, as consistently observed in the analysis of synthetic data.

In order to sustain the drawn considerations on real data, we also computed the total number of residues per image, as already done for the synthetic test data set. The results are presented in Table V, columns A- and B-. For all considered

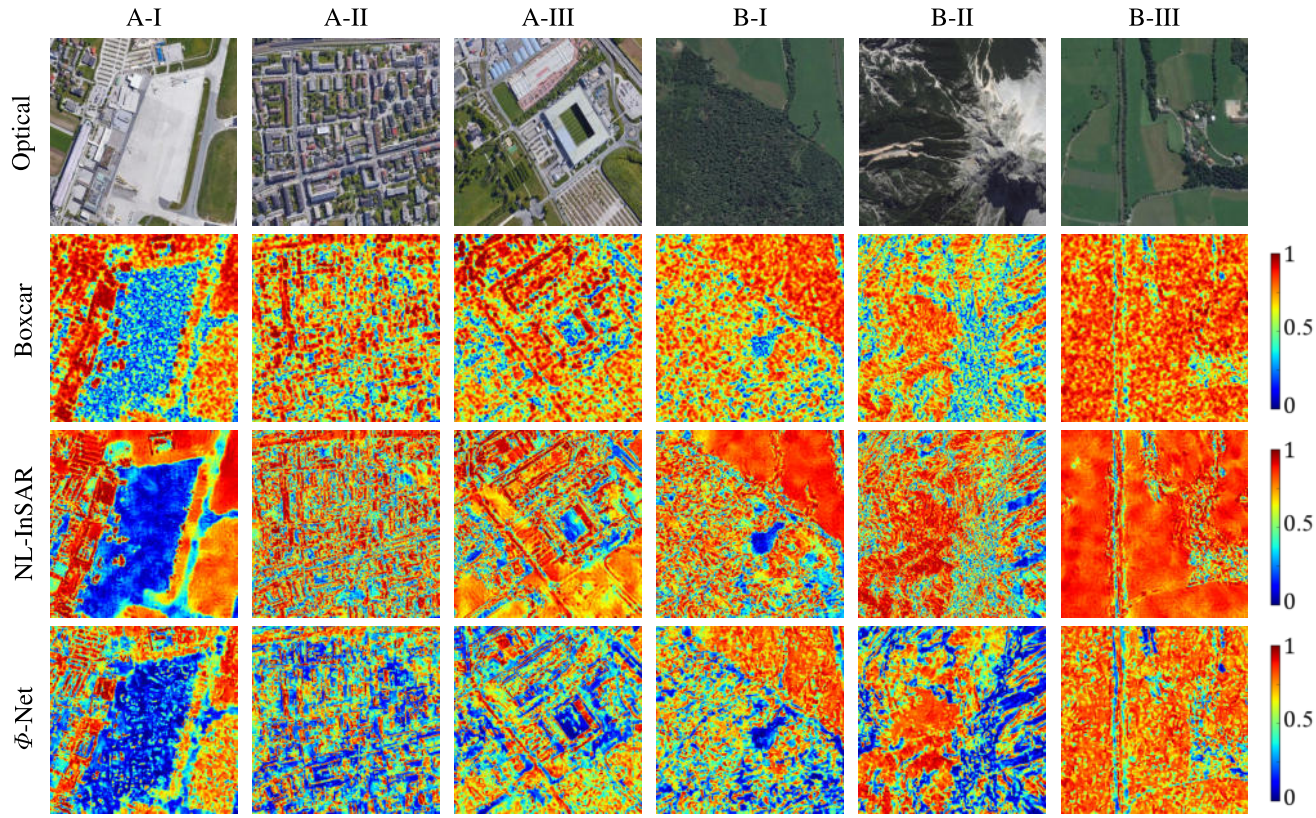


Fig. 17. Coherence images of the considered TanDEM-X patches from Fig. 15, estimated using SOA methods and our proposed Φ-Net. The first row shows the corresponding optical images taken from GoogleEarth.

test sites, such a number considerably decreases when applying the Φ-Net, which is a clear indicator of the effectiveness of the proposed approach.

Similar results can be observed on the Sentinel-1 data. For this analysis, we exploit an interferometric wide-swath (IW) data pair acquired on March 13, 2018, (master date) and March 19, 2018, over the city of Frankfurt, Germany. For the selected subswath, the SAR amplitude, the estimated phase, and the estimated coherence, obtained by applying the Φ-Net, are depicted in Fig. 18. From these data, we select three image patches of dimension 300×600 pixels over three meaningful areas.

1) *C-I:* Urban area surrounding the Frankfurt airport.
2) *C-II:* Agricultural fields around the city of Frankfurt.
3) *C-III:* Urban scenario showing the presence of a river with bridges.

These patches are indicated in red in Fig. 18.

We present the results obtained by the Φ-Net in comparison with those achieved by each of the considered SOA methods for the estimation of the interferometric phase and coherence in Figs. 19 and 20, respectively. According to the results obtained on TanDEM-X data, we note here that the Φ-Net is able to suppress the noise by preserving the resolution of small details in both the phase and the coherence images. This behavior is visible in the phase maps when looking at the buildings surrounding the airport in C-I and C-II or the bridges in C-III. At the same time, Φ-Net presents excellent performance also in the presence of severe noise levels, as it can be clearly observed from the patch C-III, which exhibits

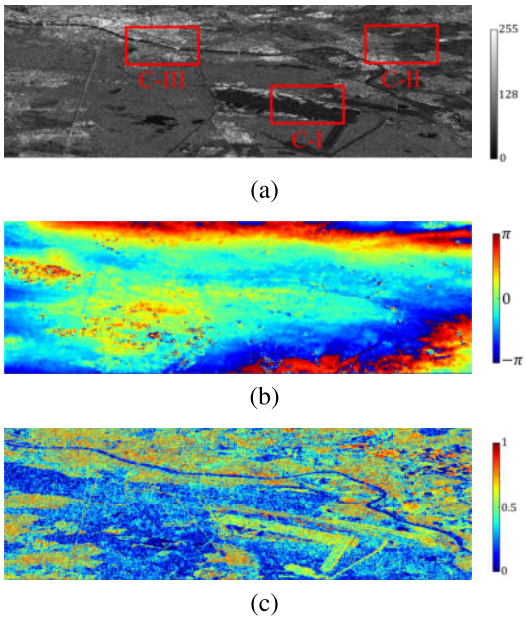


Fig. 18. Analyzed Sentinel-1 subswath. (a) Amplitude, (b) interferometric phase, and (c) interferometric coherence obtained by applying the Φ-Net. The three considered patches of size 300×600 pixels are shown in (a) as red boxes.

areas with strong noise due to the presence of water surfaces. On the contrary, SOA methods show smoothed details and a poor performance for strong noise levels, which results in residual phase noise and artifacts. The latter can be observed

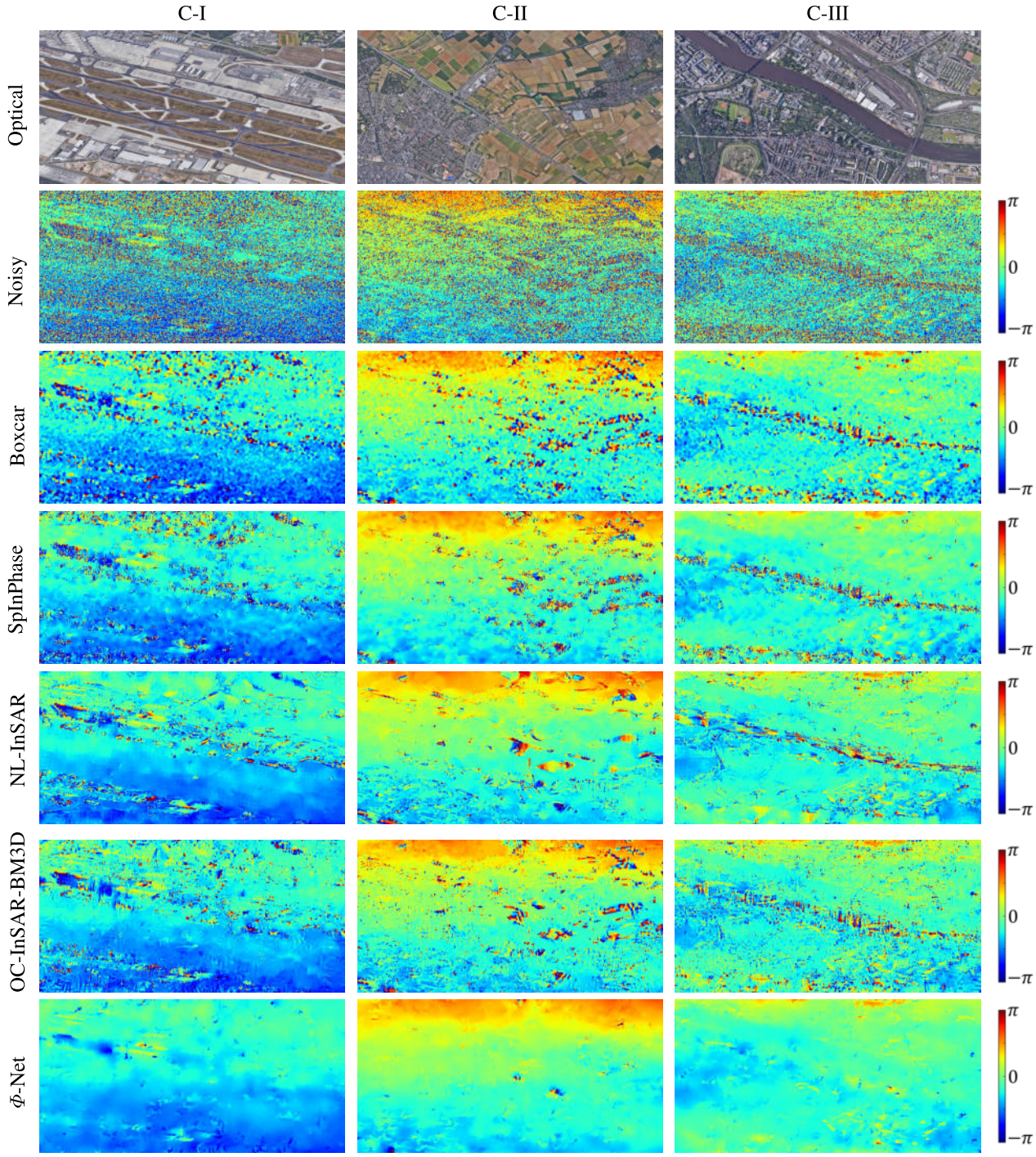


Fig. 19. Interferometric phase images of the considered Sentinel-1 patches from Fig. 18, estimated using SOA methods and our proposed Φ-Net. The first and second rows show the optical image from GoogleEarth and the noisy interferogram at azimuth/ground-range full resolution (14×3.7 meters), respectively.

in C-III, especially for the NLInSAR method, which shows horizontal stripes oriented along the direction of the river. Moreover, the results on the coherence estimation also confirm the previous observations. Indeed, Φ-Net provides better detail preservation for both man-made structures in C-I and C-III and agricultural fields in C-II, as depicted in Fig. 20. As already done for the TanDEM-X data, we also computed the total number of residues (see Table V, C-columns). All test sites are either residues-free or close to zero, achieving a significant improvement not only compared with all SOA methods but also with respect to the similar U-Net architecture.

B. Analysis on High-Resolution DEM Generation

We further assess the performance of Φ-Net by applying it to the case of high-resolution DEM generation. For this purpose, we compare Φ-Net with both the boxcar filter (which is used as standard processing for the generation of TanDEM-X DEM products) and OC-InSAR-BM3D (which is the best

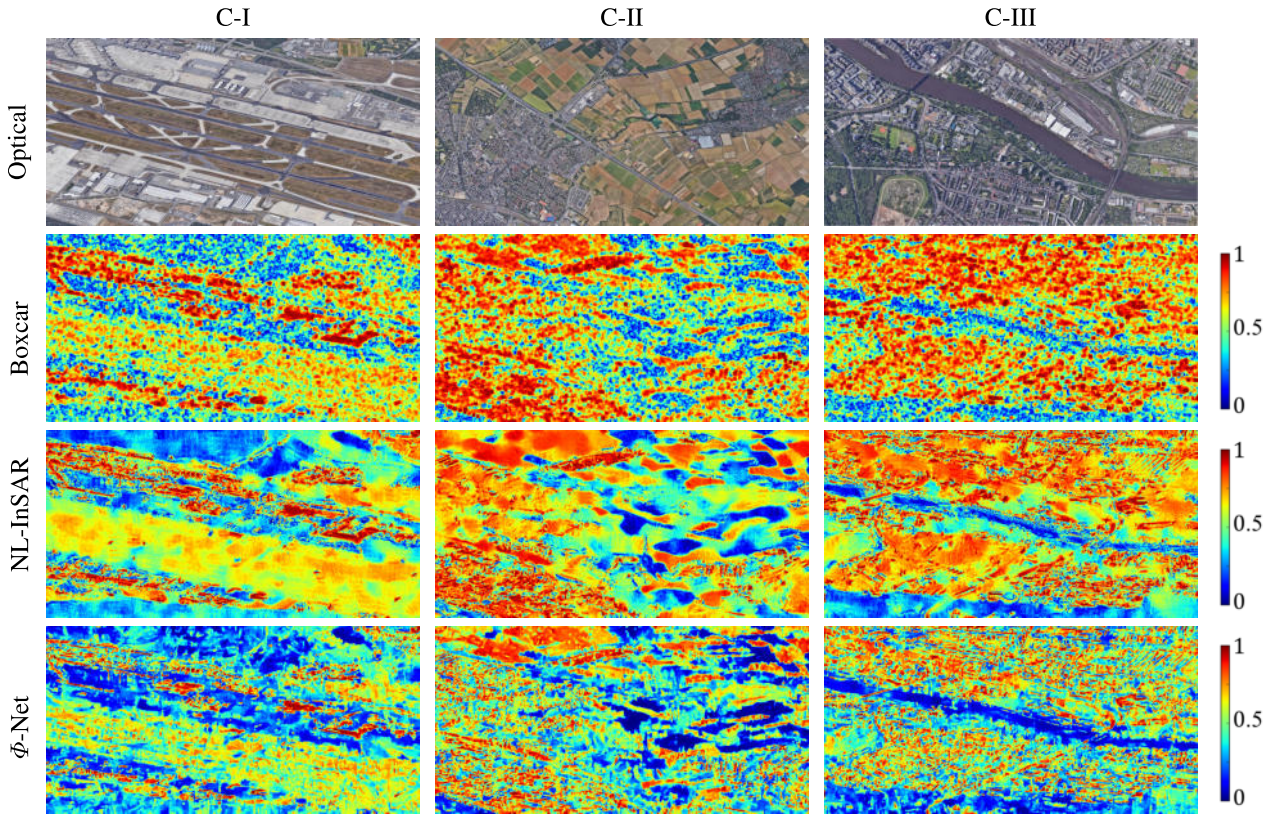


Fig. 20. Coherence images of the considered Sentinel-1 patches from Fig. 18, estimated using SOA methods and our proposed Φ-Net. The first row shows the corresponding optical images taken from GoogleEarth.

TABLE V
NUMBER OF RESIDUES COMPUTED ON THE TanDEM-X (A- AND B-COLUMNS) AND SENTINEL-1 (C-COLUMNS) REAL DATA SETS

Number of Residues on Interferometric Phase [ ]									
	A-I	A-II	A-III	B-I	B-II	B-III	C-I	C-II	C-III
boxcar	1148	812	589	437	1160	80	1611	1501	1237
SpInPhase	1454	924	624	564	1109	63	3336	2637	2339
NL-InSAR	748	2452	1255	971	1949	287	1571	870	1658
OC-InSAR-BM3D	1487	2347	1552	660	1593	150	1866	1645	2735
U-Net	938	267	196	136	822	17	22	10	32
***Φ*-Net**	**480**	**139**	**55**	**22**	**318**	**13**	**5**	**0**	**0**

TABLE VI
RMSE AND 90 PERCENTILE (90%) OF THE ABSOLUTE HEIGHT ERROR BETWEEN THE REFERENCE LIDAR DTM AND ALL THREE CONSIDERED InSAR DEMs, EVALUATED USING PIXELS CLASSIFIED AS NONVEGETATED AREAS ONLY

Absolute Height Error [m]		
Algorithm	RMSE	90%
Boxcar	7.18	10.95
OC-InSAR-BM3D	6.13	7.65
Φ-Net	5.35	7.35

among the considered SOA methods for the DEM generation, as already shown in previous studies [17]).

We integrated OC-InSAR-BM3D, together with our Φ-Net, within the interferometric processor TAXI [51], available at the Microwaves and Radar Institute of the German Aerospace Center (DLR). In this specific application, the TAXI processor is used to perform: 1) phase unwrapping; 2) phase to height conversion; and 3) geocoding. Phase unwrapping is performed by means of the SNAPHU algorithm [52], which is among SOA methods. Then, we perform the analysis by comparing the processed InSAR DEMs with an external reference over the same area. In particular, as reference data, we use an airborne laser scanning (ALS) digital terrain model (DTM) with a resolution on the ground of 5 m [53]. Such a product is provided over the Austrian territory only so that it is not available over the whole considered TanDEM-X acquisition, which is located at the border between Austria and Germany. For this reason, we perform a performance analysis over one selected patch only, aiming at reproducing a worst case scenario. In particular, we selected a patch of 300 × 300 pixels, characterized by the presence of mountainous terrain and only moderately affected by extreme geometric distortions (shadow and layover), since, in such cases, the InSAR height retrieval would not be meaningful. The considered DEM patch is depicted in Fig. 21. The InSAR DEMs are processed at a resolution of 6 m (independent pixel spacing), and the reference Lidar DTM is accordingly resampled.

It should be pointed out that the available Lidar DTM does not include vegetation height, while InSAR DEMs represent the height of the mean phase center resulting from the superimposition of all possible returns within a resolution cell. For this reason, in order to perform a fair comparison between the Lidar DTM and the InSAR DEMs, we masked out vegetated areas using the FROM-GLC land cover map presented in [54] and [55], which is displayed in the second row of Fig. 21.

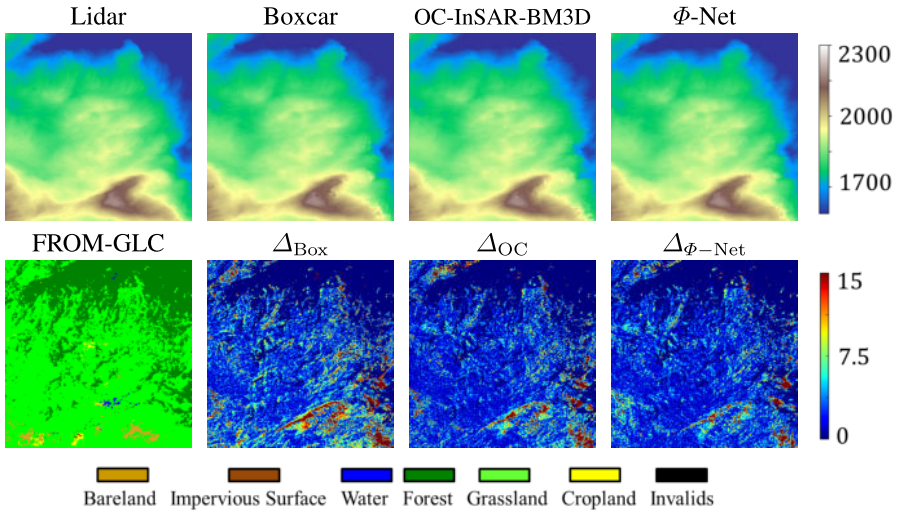


Fig. 21. DEM performance analysis over a selected patch using TanDEM-X bistatic data. From left to right, the top row shows the reference Lidar DTM, the Boxcar DEM, the OC-InSAR-BM3D DEM, and the Φ-Net DEM. The second row depicts the corresponding FROM-GLC land cover map and the absolute height error maps between Lidar/Boxcar, Lidar/OC-InSAR-BM3D, and Lidar/Φ-Net DEMs (Δ_{Box}, Δ_{OC}, and $\Delta_{\Phi-\mathrm{Net}}$, respectively). All color bars on the right-hand side are in meters, while the FROM-GLC legend is depicted at the bottom.

The absolute height error maps with respect to the LiDAR DTM and the generated InSAR DEMs are labeled in Fig. 21 as Δ_{Box}, Δ_{OC}, and $\Delta_{\Phi-\mathrm{Net}}$. As performance parameters, we evaluate the RMSE and the 90 percentile of the error.

From the error maps in Fig. 21, we can observe that Φ-Net better reconstructs the slopes in correspondence with the mountain at the bottom right of the DEM image, where most of the errors occur. The numerical results of Table VI confirm this observation.

From this analysis, visible improvements with respect to both the boxcar filter and OC-InSAR-BM3D can be observed. Nevertheless, different from the performance assessment made on the estimated phase and coherence images, this DEM study shows more comparable performance for all the considered methods. The reason for this behavior relies on two main aspects. On the one hand, the process of generating a DEM introduces a certain inaccuracy, which has the same impact on all processed DEMs. The larger such an uncertainty is, the more similar the resulting DEM errors will appear. The same applies to the inaccuracy present in the InSAR acquisition itself due to geometrical distortions caused by the side-looking acquisition geometry typical of SAR sensors. On the other hand, in order to generate a 6-m product, we reduce the resolution of the original data by a factor 2 in both azimuth and range. This is done during the geocoding step, which formerly applies low-pass filtering to the input phases, thus reducing the differences between the generated DEMs. Nevertheless, these results are very promising and confirm the strong potential of the proposed Φ-Net for the generation of high-resolution DEMs. As a future investigation, we also aim at comparing the generated DEM products by using different unwrapping approaches, as the PUMA algorithm [56], and by using recent methodologies that embed interferometric phase estimation and unwrapping in a single approach, as for the PARISAR algorithm [57].

VIII. Conclusion and Discussion

In this article, we presented Φ-Net, a novel DL network for the estimation of both the interferometric phase and the coherence from SAR data. We provided two main contributions: 1) the design of a CNN architecture suitable for the processing of the interferometric signal and 2) the development of an effective strategy for the generation of a large, variegate, and reliable training data set.

The first contribution is based on the exploitation of residual learning connections embedded into the architecture of the U-Net. This kind of network, which has never been investigated before for the challenging task of interferometric parameters estimation, has shown very powerful characteristics. The cascade of encoder and decoder stages, used in combination with skip connections and residual shortcuts, enables an effective representation of the interferometric signal and the superimposed noise. It results that Φ-Net is able to preserve fringe structures of any density as well as abrupt changes of phase and coherence while strongly reducing the noise. This is observed from the visual inspection of the phase and coherence images of synthetic data. Indeed, Φ-Net better preserves phase fringe structures that do not appear in the phase and coherence error maps, as it is the case for the other SOA methods. Furthermore, we observed consistent results when applying Φ-Net to real InSAR data. In particular, Φ-Net shows a very good capability of preserving high-resolution details and spatial textures that, on the contrary, appear mostly blurred or distorted with the other considered methods.

A key role on the estimation performance is played by the used training data set. The generation of the training data set has been approached at different levels of approximation. According to [39], we focused on modeling both the approximation component (which refers to the representation of the main background physical information) and the detail

component (which introduces more precise modeling based on realistic scenarios). On the one hand, the former is tackled by exploiting the knowledge of physics behind the SAR theory and by accordingly modeling the relationship of interdependence among the noise-free interferometric parameters. On the other hand, the latter is considered when simulating realistic spatial patterns for each of the parameters. For example, the used interferometric phase patterns have been generated according to real DEMs and by means of *ad hoc* designed phase steps, while amplitude and coherence are generated on the basis of natural patterns. The inclusion of a meaningful detail component significantly improved network performance. Indeed, Φ-Net is able to perform very well on a large variety of synthetic test patterns and on real InSAR data, showing better results than SOA methods on the considered test cases. The obtained results over real data from both the TanDEM-X and the Sentinel-1 missions prove that Φ-Net is a robust architecture. From the obtained results, we can assert that the created synthetic data set has more influence on the final performance with respect to the introduction of residual connections in the U-Net, given the proposed training strategy, which includes the processing of the input data and the selected loss function. Even though the introduction of residual links has shown a further increase in the overall performance, this improvement is less sharp than the one obtained with respect to SOA algorithms. In conclusion, to further improve the obtained results, both the network architecture and the training data set generation should be jointly optimized.

As a possible application scenario, we eventually considered the generation of DEM by using single-pass InSAR. In particular, we generated DEMs from a single TanDEM-X acquisition, by exploiting the boxcar filter, OC-InSAR-BM3D, and Φ-Net. In order to assess the performance, we investigated the mismatch between the InSAR DEMs with respect to a Lidar DTM, which has been used as a reference. As confirmed by the computed absolute height error metrics, Φ-Net shows a high potential for the generation of high-resolution DEMs. By also considering its much shorter computational time with respect to OC-InSAR-BM3D (which is the best SOA method), Φ-Net shows to be a good candidate for establishing an operational high-resolution processing chain.

Further developments of this work will also be focused on the improvement of the current training data set, by increasing its specificity for dedicated scenarios [39] through the use of real InSAR data as well.

ACKNOWLEDGMENT

The authors would like to thank Philipp Posovszky for his valuable IT support. They are also very grateful to the anonymous reviewers for their comments and suggestions that helped improving the final quality of this article.

REFERENCES

[1] D. Massonnet *et al.*, "The displacement field of the landers earthquake mapped by radar interferometry," *Nature*, vol. 364, no. 6433, pp. 138–142, Jul. 1993.

[2] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution (an introduction)," *Ann. Math. Statist.*, vol. 34, no. 1, pp. 152–177, Mar. 1963.

[3] R. Bamler and P. Hartl, "Synthetic aperture radar interferometry," *Inverse Problems*, vol. 14, no. 4, pp. R1–R54, 1998.

[4] M. S. Seymour and I. G. Cumming, "Maximum likelihood estimation for SAR interferometry," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Aug. 1994, pp. 2272–2275.

[5] J.-S. Lee, K. P. Papathanassiou, T. L. Ainsworth, M. R. Grunes, and A. Reigber, "A new technique for noise filtering of SAR interferometric phase images," *IEEE Trans. Geosci. Remote Sens.*, vol. 36, no. 5, pp. 1456–1465, Sep. 1998.

[6] S. Fu, X. Long, X. Yang, and Q. Yu, "Directionally adaptive filter for synthetic aperture radar interferometric phase images," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 552–559, Jan. 2013.

[7] R. M. Goldstein and C. L. Werner, "Radar interferogram filtering for geophysical applications," *Geophys. Res. Lett.*, vol. 25, no. 21, pp. 4035–4038, Nov. 1998.

[8] I. Baran, M. P. Stewart, B. M. Kampes, Z. Perski, and P. Lilly, "A modification to the goldstein radar interferogram filter," *IEEE Trans. Geosci. Remote Sens.*, vol. 41, no. 9, pp. 2114–2118, Sep. 2003.

[9] C. Lopez-Martinez and X. Fabregas, "Modeling and reduction of SAR interferometric phase noise in the wavelet domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 12, pp. 2553–2566, Dec. 2002.

[10] H. Hongxing, J. M. Bioucas-Dias, and V. Katkovnik, "Interferometric phase image estimation via sparse coding in the complex domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 1072–1963, Oct. 2015.

[11] A. Buades, B. Coll, and J. M. Morel, "A review of image denoising algorithms, with a new one," *Multiscale Model. Simul.*, vol. 4, no. 2, pp. 490–530, Jan. 2005.

[12] C.-A. Deledalle, L. Denis, and F. Tupin, "NL-InSAR: Nonlocal interferogram estimation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 4, pp. 1441–1452, Apr. 2011.

[13] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, "NL-SAR: A unified nonlocal framework for resolution-preserving (Pol)(In)SAR denoising," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 4, pp. 2021–2038, Apr. 2015.

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

[15] F. Sica, D. Cozzolino, X. X. Zhu, L. Verdoliva, and G. Poggi, "InSAR-BM3D: A nonlocal filter for SAR interferometric phase restoration," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 6, pp. 3456–3467, Jun. 2018.

[16] F. Sica, D. Cozzolino, L. Verdoliva, and G. Poggi, "The offset-compensated nonlocal filtering of interferometric phase," *Remote Sens.*, vol. 10, no. 9, p. 1359, Aug. 2018.

[17] F. Sica and N. Gollin, "Analysis of offset-compensated nonlocal filtering for InSAR DEM generation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 5201–5204.

[18] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[19] K. Zhang, W. Zuo, and L. Zhang, "FFDNet: Toward a fast and flexible solution for CNN-based image denoising," *IEEE Trans. Image Process.*, vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[20] V. Jain and S. Seung, "Natural image denoising with convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2009, pp. 679–776.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[23] 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.

[24] P. Wang, H. Zhang, and V. M. Patel, "SAR image despeckling using a convolutional neural network," *IEEE Signal Process. Lett.*, vol. 24, no. 12, pp. 1763–1767, Dec. 2017.

[25] N. K. Kottayil, A. Zimmer, S. Mukherjee, X. Sun, P. Ghuman, and I. Cheng, "Accurate pixel-based noise estimation for InSAR interferograms," in *Proc. IEEE Sensors*, Oct. 2018, pp. 1–4.

[26] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, Oct. 2015, pp. 234–241.

[27] M. P. Heinrich, M. Stille, and T. M. Buzug, "Residual U-Net convolutional neural network architecture for low-dose CT denoising," *Current Directions Biomed. Eng.*, vol. 4, no. 1, pp. 297–300, Sep. 2018.

[28] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual U-Net," *IEEE Geosci. Remote Sens. Lett.*, vol. 15, no. 5, pp. 749–753, May 2018.
[29] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[30] C. Szegedy *et al.*, "Going deeper with convolutions," 2014, *arXiv:1409.4842*. [Online]. Available: http://arxiv.org/abs/1409.4842
[31] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[32] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
[33] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 1–11.
[34] R. K. Srivastava, K. Greff, and J. Schmidhuber, "Highway networks," 2015, *arXiv:1505.00387*. [Online]. Available: http://arxiv.org/abs/1505.00387
[35] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.*, 2016, pp. 630–645.
[36] D. Han, J. Kim, and J. Kim, "Deep pyramidal residual networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 5927–5935.
[37] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.
[38] K. Zhang, M. Sun, T. X. Han, X. Yuan, L. Guo, and T. Liu, "Residual networks of residual networks: Multilevel residual networks," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 6, pp. 1303–1314, Jun. 2018.
[39] L. Bruzzone, "Multisource labeled data: An opportunity for training deep learning networks," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 4799–4802.
[40] T. G. Farr *et al.*, "The shuttle radar topography mission," *Rev. Geophys.*, vol. 45, no. 2, pp. 1–33, 2007.
[41] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proc. IEEE*, vol. 105, no. 10, pp. 1865–1883, Oct. 2017.
[42] G. Cheng. (Sep. 2019). *NWPU-Resisc45 Data Set*. [Online]. Available: http://www.escience.cn/people/JunweiHan/NWPU-RESISC45.html
[43] H. A. Zebker and J. Villasenor, "Decorrelation in interferometric radar echoes," *IEEE Trans. Geosci. Remote Sens.*, vol. 30, no. 5, pp. 950–959, Sep. 1992.
[44] G. Krieger *et al.*, "TanDEM-X: A satellite formation for high-resolution SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 11, pp. 3317–3341, Nov. 2007.
[45] M. Martone, P. Rizzoli, and G. Krieger, "Volume decorrelation effects in TanDEM-X interferometric SAR data," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 12, pp. 1812–1816, Dec. 2016.
[46] R. Hawkins, E. Attema, R. Crapolicchio, P. Lecomte, J. Closa, P. J. Meadows, and S. K. Srivastava, "Stability of Amazon backscatter at C-band: Spaceborne results from ERS-1/2 and RADARSAT-1," in *Proc. SAR Workshop, CEOS Committee Earth Observ. Satell., Working Group Calibration Validation*, Toulouse, France, vol. 450, R. A. Harris and L. Ouwehand, Eds. Paris, France: European Space Agency, Mar. 2000, p. 99.
[47] R. K. Raney, A. Freeman, R. W. Hawkins, and R. Bamler, "A plea for radar brightness," in *Proc. Int. Geosci. Remote Sens. Symp.*, vol. 2, 1994, p. 1090.
[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*. [Online]. Available: http://arxiv.org/abs/1412.6980
[49] P. Rizzoli *et al.*, "Generation and performance assessment of the global TanDEM-X digital elevation model," *ISPRS J. Photogramm. Remote Sens.*, vol. 132, pp. 119–139, Oct. 2017.
[50] A. Buades, B. Coll, and J.-M. Morel, "The staircasing effect in neighborhood filters and its solution," *IEEE Trans. Image Process.*, vol. 15, no. 6, pp. 1499–1505, Jun. 2006.
[51] P. Prats *et al.*, "Taxi: A versatile processing chain for experimental TanDEM-X product evaluation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2010, pp. 1–4.
[52] C. W. Chen and H. A. Zebker, "Phase unwrapping for large SAR interferograms: Statistical segmentation and generalized network models," *IEEE Trans. Geosci. Remote Sens.*, vol. 40, no. 8, pp. 1709–1719, Aug. 2002.
[53] *Digital Terrain Models of Austria*. Accessed: Oct. 18, 2019. [Online]. Available: http://data.opendataportal.at/dataset/dtm-austria
[54] P. Gong *et al.*, "Finer resolution observation and monitoring of global land cover: First mapping results with Landsat TM and ETM+data," *Int. J. Remote Sens.*, vol. 34, no. 7, pp. 2607–2654, 2013.
[55] P. Gong *et al.*, "Stable classification with limited sample: Transferring a 30-m resolution sample set collected in 2015 to mapping 10-m resolution global land cover in 2017," *Sci. Bull.*, vol. 64, no. 6, pp. 370–373, Mar. 2019.
[56] J. M. Bioucas-Dias and G. Valadao, "Phase unwrapping via graph cuts," *IEEE Trans. Image Process.*, vol. 16, no. 3, pp. 698–709, Mar. 2007.
[57] G. Ferraioli, C.-A. Deledalle, L. Denis, and F. Tupin, "Parisar: Patch-based estimation and regularized inversion for multibaseline SAR interferometry," *IEEE Trans. Geosci. Remote Sens.*, vol. 56, no. 3, pp. 1626–1636, Mar. 2018.

Francescopaolo Sica (Member, IEEE) received the Laurea (M.S.) degree *(summa cum laude)* in telecommunication engineering and the Dr.Ing. (Ph.D.) degree in information engineering from the University of Naples Federico II, Naples, Italy, in 2012 and 2016, respectively.

From November 2014 to February 2016, he visited the Remote Sensing Technology Institute (EOC), German Aerospace Center (DLR), Weßling, Germany, working as a Guest Ph.D. Student on statistical methods for the estimation of interferometric parameters. Since 2012, he has been with the Italian National Research Council (IREA-CNR), Naples, where he started working on multitemporal/multibaseline synthetic aperture radar (SAR) interferometry applications for deformation monitoring. He is a Research Engineer with the Microwaves and Radar Institute, German Aerospace Center (DLR). Since January 2019, he has been pursuing the Living Planet Fellowship Post-Doctoral Program of the European Space Agency with the project, High-Resolution Forest Coverage with InSAR & Deforestation Surveillance (HI-FIVE). Within this project, he is investigating how to achieve a global and systematic observation of the world's forests and measure deforestation through Earth observation and machine learning algorithms. His research interests include the processing of synthetic aperture radar (SAR) images for single-baseline and multibaseline interferometry with specific application to surface deformation monitoring, digital elevation model (DEM) generation, and land cover classification; recent interests concern the development of deep learning algorithms for statistical inference and land cover classification from SAR, optical, and multispectral data.

Dr. Sica was a recipient of the IEEE Student Prize 2012 for the best master thesis. He regularly serves as a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Giorgia Gobbi received the M.Sc. degree in information and communications engineering from the University of Trento, Trento, Italy, in 2019, with a thesis on InSAR parameters retrieval with deep residual learning approach, pursued at the Microwaves and Radar Institute, German Aerospace Center (DLR), Weßling, Germany. She is pursuing the Ph.D. degree with the System Performance Group, Microwaves and Radar Institute, DLR.

Her research interests include statistical signal estimation of remotely sensed images in application to interferometric synthetic aperture radar (SAR) data and the development of machine learning algorithms applied to interferometric SAR parameter retrieval.

Paola Rizzoli received the bachelor's and master's degrees in telecommunication engineering from the Politecnico di Milano (Polimi), Milan, Italy, in 2003 and 2006, respectively, and the Dr.Ing. (Ph.D.) degree *(summa cum laude)* in electrical engineering and information technology from the Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, in 2018.

From 2006 to 2008, she was a Scientific Researcher and a Project Engineer with the Politecnico di Milano and Aresys s.r.l., a Polimi spin-off company. At the end of 2008, she joined the Microwaves and Radar Institute, German Aerospace Center (DLR), Weßling, Germany, as a Project Engineer, where she has been involved in the development and optimization of the TerraSAR-X and TanDEM-X spaceborne synthetic aperture radar (SAR) missions, concentrating, in particular, on the generation of the TanDEM-X global digital elevation model. Since 2016, she has been leading the System Performance Research Group, Satellite SAR Systems Department, German Aerospace Center, being responsible for the final performance assessment of the global TanDEM-X DEM and the generation of the global TanDEM-X forest/nonforest map. Her main research interests include SAR systems design, data reduction techniques, estimation theory, variable modeling, signal processing, and artificial intelligence algorithms. She is concentrating on the development of innovative machine learning- and deep learning-based approaches for the study of the biosphere and the cryosphere from remotely sensed earth observation data, ranging from land cover classification and deforestation monitoring to snow facies classification and glaciers dynamics.

Dr. Rizzoli was a recipient of the DLR Science Award in 2018 and the Best Paper Award at the German Microwave Conference in 2019. She regularly serves as a Reviewer for the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING, the IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, and the IEEE JOURNAL OF SELECTED TOPICS IN APPLIED EARTH OBSERVATIONS AND REMOTE SENSING.

Lorenzo Bruzzone (Fellow, IEEE) received the Laurea (M.S.) degree *(summa cum laude)* in electronic engineering and the Ph.D. degree in telecommunications from the University of Genoa, Genoa, Italy, in 1993 and 1998, respectively.

He is a Full Professor of telecommunications with the University of Trento, Trento, Italy, where he teaches remote sensing, radar, and digital communications. He is also the Founder and the Director of the Remote Sensing Laboratory, Department of Information Engineering and Computer Science, University of Trento. He is also a principal investigator of many research projects. Among others, he is the Principal Investigator of the Radar for Icy Moon Exploration (RIME) instrument in the framework of the JUpiter ICy moons Explorer (JUICE) mission of the European Space Agency (ESA) and the Science Lead for the High Resolution Land Cover project in the framework of the Climate Change Initiative of ESA. He is the author (or coauthor) of 276 scientific publications in refereed international journals (209 in IEEE journals), more than 330 articles in conference proceedings, and 22 book chapters. He is an editor/coeditor of 18 books/conference proceedings and one scientific book. His articles are highly cited, as proven from the total number of citations (more than 34 000) and the value of the H-index (87) (source: Google Scholar). He was invited as a keynote speaker in more than 40 international conferences and workshops. His research interests are in the areas of remote sensing, radar and synthetic aperture radar (SAR), signal processing, machine learning, and pattern recognition. He promotes and supervises research on these topics within the frameworks of many national and international projects.

Dr. Bruzzone has been a member of the Administrative Committee of the IEEE Geoscience and Remote Sensing Society (GRSS) since 2009, where he has been the Vice-President for Professional Activities since 2019. He ranked first place in the Student Prize Paper Competition of the 1998 IEEE International Geoscience and Remote Sensing Symposium (IGARSS), Seattle, July 1998. Since then, he was a recipient of many international and national honors and awards, including the IEEE GRSS 2015 Outstanding Service Award, the 2017 and 2018 IEEE IGARSS Symposium Prize Paper Awards, and the 2019 WHISPER Outstanding Paper Award. Since 2003, he has been the Chair of the SPIE Conference on Image and Signal Processing for Remote Sensing. He was a guest coeditor of many special issues of international journals. He is also the Co-Founder of the IEEE International Workshop on the Analysis of MultiTemporal Remote-Sensing Images (MultiTemp) Series and is a member of the Permanent Steering Committee of this series of workshops. He is the Founder of the *IEEE Geoscience and Remote Sensing Magazine* for which was the Editor-in-Chief from 2013 to 2017. He is also an Associate Editor of the IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING. He was a Distinguished Speaker of the IEEE Geoscience and Remote Sensing Society from 2012 to 2016.