ISPRS Journal of Photogrammetry and Remote Sensing 220 (2025) 192–206

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Synthesis of complex-valued InSAR data with a multi-task convolutional neural network

Philipp Sibler [a,b], Francescopaolo Sica [a], Michael Schmitt [a,*]

[a] Department of Aerospace Engineering, University of the Bundeswehr Munich, Werner-Heisenberg-Weg 39, 85577, Neubiberg, Germany
[b] Hensoldt Sensors GmbH, Graf-von-Soden-Str., 88090, Immenstaad, Germany

ARTICLE INFO

Keywords:
Synthetic aperture radar (SAR)
Deep learning
Multitask learning
Image synthesis
SAR interferometry (InSAR)

ABSTRACT

Simulated remote sensing images bear great potential for many applications in the field of Earth observation. They can be used as controlled testbed for the development of signal and image processing algorithms or can provide a means to get an impression of the potential of new sensor concepts. With the rise of deep learning, the synthesis of artificial remote sensing images by means of deep neural networks has become a hot research topic. While the generation of optical data is relatively straightforward, as it can rely on the use of established models from the computer vision community, the generation of synthetic aperture radar (SAR) data until now is still largely restricted to intensity images since the processing of complex-valued numbers by conventional neural networks poses significant challenges. With this work, we propose to circumvent these challenges by decomposing SAR interferograms into real-valued components. These components are then simultaneously synthesized by different branches of a multi-branch encoder–decoder network architecture. In the end, these real-valued components can be combined again into the final, complex-valued interferogram. Moreover, the effect of speckle and interferometric phase noise is replicated and applied to the synthesized interferometric data. Experimental results on both medium-resolution C-band repeat-pass SAR data and high-resolution X-band single-pass SAR data, demonstrate the general feasibility of the approach.

1. Introduction

Synthetic Aperture Radar (SAR) Interferometry is an essential technique in the field of remote sensing due to its ability to accurately provide high-resolution topographic information and measure changes in the Earth's surface. SAR interferometry (InSAR) utilizes a pair of SAR images acquired with a spatial baseline.

Changes in the spatial position or the time of acquisition of the SAR sensor, forming the spatial or temporal baseline, lead to changes in the complex reflectivity of a scene that can be evaluated by InSAR processing. The importance of the InSAR technique lies in its ability to accurately measure topography and surface deformation, which can have applications in various fields such as natural hazard assessment, infrastructure monitoring, change detection, and environmental studies. SAR interferometry techniques start from combining a pair of coregistered complex SAR images but, for higher-level InSAR processing, can extend to the processing of a stack of multiple related and coregistered images.

Deep Learning (DL) algorithms have been used to address various challenges in SAR interferometry, including improving the accuracy of digital elevation models (DEMs), reducing the complexity of phase unwrapping, and mitigating phase noise in SAR interferograms. To our knowledge, the majority of DL algorithms applied in the InSAR realm are so far based on supervised learning. This type of approach requires large and diverse sets of labeled data to learn from and generalize to unseen data. However, acquiring large amounts of labeled InSAR data can be challenging and even not always possible, as not every task can be associated with ground truth, such as for InSAR phase and coherence estimation as well as phase unwrapping. Therefore, the usually adopted strategy is generating synthetic pairs of true and noisy signals. Synthetic InSAR data can be generated, for example, by simulating SAR acquisitions over a virtual terrain, taking into account various types of scatterers and acquisition geometries. Synthetic data

* Corresponding author.
E-mail addresses: philipp.sibler@unibw.de, philipp.sibler@hensoldt.net (P. Sibler), francescopaolo.sica@unibw.de (F. Sica), michael.schmitt@unibw.de (M. Schmitt).

https://doi.org/10.1016/j.isprsjprs.2024.12.007
Received 13 July 2023; Received in revised form 28 October 2024; Accepted 5 December 2024
Available online 18 December 2024
0924-2716/© 2024 The Authors. Published by Elsevier B.V. on behalf of International Society for Photogrammetry and Remote Sensing, Inc. (ISPRS). This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

generation can be used to improve the development and evaluation of Deep Learning-based InSAR algorithms in a controlled environment.

1.1. Synthesis of remote sensing data

The generation of synthetic data for remote sensing applications can itself be supported by Deep Learning architectures and has already been performed for numerous EO applications in the past years.

The most direct way to generate artificial images using deep learning has been offered by Generative Adversarial Networks (GANs) (Goodfellow et al., 2014) and later by their conditional variants (cGANs) (Isola et al., 2016; Wang et al., 2018). GANs consist of two neural networks: a generator network and a discriminator one. The generator network generates synthetic images, while the discriminator network evaluates the quality of the synthetic images and provides feedback to the generator. Conditional GANs can improve the quality of synthetic images by providing additional information to the generator network.

This technique has been successfully applied to generate artificially labeled datasets of optical remote sensing images in Howe et al. (2019) by applying labeled semantic maps to realistic optical EO images. In this work, the authors demonstrated that augmenting real datasets with their synthetic data increased the accuracy of object detection tasks. Similarly, in Abady et al. (2022) the authors rely on GANs for land cover and season transfer on hyperspectral Sentinel-2 datasets. The land cover is used as input to modify image content from vegetation to barren and vice versa, whereas the season information is used to perform a transfer of image content from winter to summer or vice versa.

Moreover, convolutional neural networks (CNN) have been largely used for image-to-image translation tasks. Among the EO applications, we can mention the tasks of scene matching between optical and SAR image modalities (Mou et al., 2017; Merkle et al., 2018; Hughes et al., 2018), SAR image synthesis (Guo et al., 2017), SAR automatic target recognition (ATR) tasks (Shi et al., 2019; Cao et al., 2020; Brosch and Neumann, 2021; Song et al., 2022), aiding interpretation of SAR data using artificially created SAR patches (Fuentes Reyes et al., 2019), and multimodal optical and SAR image synthesis (Baier et al., 2021).

1.2. Synthesis of complex-valued SAR images

The generation of synthetic complex-valued SAR images or complex-valued SAR interferograms is still a largely unexplored application area of DL-based methods for Earth observation. The main challenges arise from the fact that InSAR data are inherently complex-valued, consisting of amplitude and phase components. Their generation requires accurate modeling of the underlying physical processes such as radar scattering, acquisition geometry, and wave propagation. This includes an appropriate network architecture, loss function, and training data. Complex-valued convolutional neural networks could be a viable solution for processing and analyzing complex data. However, they must be carefully designed to account for the unique properties of InSAR data, including the preservation of the phase component (Asiyabi et al., 2023). Furthermore, an appropriate loss function is needed to measure the similarity between the synthetic and real InSAR data. This loss function must take into account the amplitude and phase components of the data.

Nonlinear activation functions and loss terms are typically non-holomorphic when applied to complex-valued inputs, meaning they lack the complex derivatives needed for training. Wirtinger derivatives can help with this problem (Hirose, 2012).

Despite the rapid growth in the use of complex-valued CNNs, there is still no approach that can satisfy all of the above criteria, i.e., phase preservation by the network cannot always be guaranteed, and loss functions usually consider only the phasor amplitude, ignoring the phase. Finding an appropriate way to handle complex-valued data in real-valued CNNs is necessary to avoid direct CV-CNN implementations. In fact, input data with simply concatenated complex channels in CNN processing is prone to suffer from spectral aliasing and modulation artefacts caused by nonlinearities in the activation function layers when traversing the CNN (Sibler et al., 2021).

Therefore, the usual alternative is to resort to real-valued CNNs and split the input signal into real and imaginary components after decorrelating the two input quantities, as proposed in Sica et al. (2021).

Following the same principle, preliminary work shows that the use of real-valued CNN can be further applied to the generation of complex InSAR data (Sibler et al., 2021). The authors propose an upsampling and modulation approach that converts a data set from complex to real values while preserving all spectral information: The complex analytic image is upsampled by a factor of 2 and modulated to the center of the Nyquist zone. By enforcing the symmetry of its discrete Fourier spectrum the full spectral content of the image is preserved, and a purely real-valued image is retrieved. Thus, real-valued network models can be used. However, this approach has limitations due to the reduction of usable image size by the upsampling factor of 2 and in its ability to replicate high-frequency image content.

For super-resolution applications on complex SAR data (Addabbo et al., 2023) are employing a split network architecture with separated networks for the real and imaginary parts. In an attempt to properly preserve phase information also in the high-resolution complex output images they only allow for a single crossover connection and thus a limited sharing of information between both networks. Apart from this connection both networks are trained and operated individually.

1.3. Paper contributions

In this manuscript, we aim to provide a DL-based methodology for the generation of synthetic complex interferometric SAR data. We approach this task by using a CNN and formulating the complex signal estimation by retrieving reflectivity, coherence and phase images separately. The correlation of these three quantities allows us to formulate the problem as a multitask optimization problem and to use a multi-objective loss. We show that the proposed approach provides a general framework for interferogram generation that can be adapted to different types of datasets if properly trained for each specific scenario, including variations in acquisition modalities, resolution, spatial and temporal baselines.

We extend the topic of complex-valued CNN synthesis for SAR applications by providing the following main contributions:

1. Synthesis of artificial complex-valued SAR interferograms in ground coordinates on a multitask CNN and a supervised training approach
2. Prediction of the achievable coherence level for SAR interferometry of a given scene
3. Synthesis of artificial speckle and interferometric phase noise based on scene coherence

As an additional remark on the employed coordinate system we see no general limitation for generating artificial complex-valued SAR data in the original sensor slant range-azimuth coordinate system. However, as all input constraint images are referenced to a ground coordinate system, they can directly be interpreted and further processed in this system using a variety of tools, and as SAR data can readily be projected to ground coordinates, working in ground coordinates appeared as a reasonable basis for the general approach presented in this paper.

Moreover, by directly using ground coordinates we aimed to avoid back-geocoding input constraint information to sensor coordinates, with the risk of introducing errors due to imperfect knowledge of the local topography, especially as the typical resolution of available DEMs is much lower than the information within the input data.

The manuscript is structured as follows: Section 2 introduces InSAR imaging and its statistical model, Section 3 gives a complete description

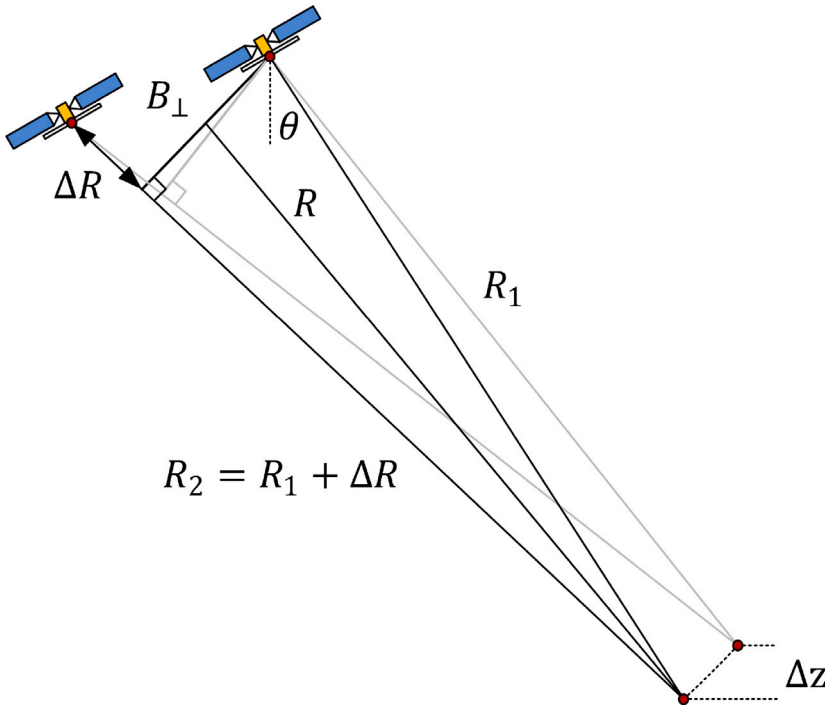

Fig. 1. InSAR acquisition geometry.

of the proposed methodology. Section 4 describes all the materials used, including the datasets used and the preprocessing applied. The experimental results and discussion follow in Sections 5 and 6, respectively. Finally, Section 7 presents the conclusion and future work.

2. InSAR data model

2.1. Complex-valued InSAR data model

Complex-valued SAR imagery is acquired from a moving radar sensor in a side-looking configuration that repeatedly and coherently illuminates the scene with pulses in its side-looking direction (range) and collects echo vectors along its direction of motion (azimuth). Focusing algorithms compress the pulse bandwidth in the range direction and compensate for the Doppler shift of the scattered echoes in the azimuth one. By compressing the resulting Doppler bandwidth of the echoes in the azimuth direction as well, a two-dimensional SLC image u is generated (Bamler and Hartl, 1998; Fornaro and Pascazio, 2014).

If the same scene is imaged on two nearby parallel trajectories separated by a spatial baseline B, a range difference ΔR to a scatterer between the two SAR sensor positions is transformed by interferometric processing into a measured interferometric phase ϕ_{12}. This type of interferometry is also called across-track interferometry.

The formation of an interferogram $\boldsymbol{v}$ from two co-registered complex-valued images $\boldsymbol{u_1}, \boldsymbol{u_2} \in \mathbb{C}$ is achieved by

$$\boldsymbol{v} = \boldsymbol{u_1} \cdot \boldsymbol{u_2}^* = |\boldsymbol{u_1}|\,|\boldsymbol{u_2}| \cdot \exp\left(j\phi_{12}\right) \tag{1}$$

with the interferometric phase

$$\phi_{12} = \varphi_1 - \varphi_2. \tag{2}$$

The phase terms of the two images are given by

$$\varphi_1 = -\frac{4\pi}{\lambda} R_1 + \varphi_{1,s} \tag{3}$$

$$\varphi_2 = -\frac{4\pi}{\lambda} R_2 + \varphi_{2,s}, \tag{4}$$

where R_n is the distance to the scatterer for image n and $\varphi_{n,s}$ is the phase shift caused by the scatterer itself. If the scatterer phase shift $\varphi_{n,s}$ can be assumed to be the same for both images, and with $\Delta R = R_2 - R_1$ the interferometric phase can be written as

$$\phi_{12} = \frac{4\pi}{\lambda} \Delta R. \tag{5}$$

For two scatterers at a constant range R but a topographic height difference Δz the *phase-to-height sensitivity* for an InSAR acquisition can be found with some derivations from the acquisition geometry, with $B_\perp$ as the perpendicular baseline and θ as sensor observation angle.

$$\frac{\partial \phi}{\partial z} = \frac{4\pi}{\lambda} \frac{B_\perp}{R \sin\theta} \tag{6}$$

A local topographic height change Δz thus translates to a change of $\Delta\phi$ in the interferometric phase according to Eq. (6). Often, the *height of ambiguity* for a full fringe (2π) of interferometric phase is provided to characterize the height sensitivity and the non-ambiguous height interval of the interferogram for the acquisition geometry depicted in Fig. 1.

$$z_{2\pi} = \frac{\lambda}{2} \frac{R \sin\theta}{B_\perp} \tag{7}$$

Especially for single-pass interferometry, based on two SAR sensors separated by a spatial baseline B but receiving parallel in time, the magnitudes $|\boldsymbol{u_1}|$, $|\boldsymbol{u_2}|$ of both SLCs can be assumed similar or almost equal. Under that assumption a general scene reflectivity β can be estimated from both SLC amplitudes as (Deledalle et al., 2011; Sica et al., 2018)

$$\beta = (|\boldsymbol{u_1}|^2 + |\boldsymbol{u_2}|^2)/2 \tag{8}$$

2.2. SAR image statistics

The SAR signal is nowadays well-known and statistically tractable. The statistical properties in terms of amplitude and phase, as well as in terms of real and imaginary parts, and the statistical dependencies linking these quantities have been extensively studied in several publications since the pioneering work in Bamler and Hartl (1998).

In general, two opposite cases are considered that can describe the majority of situations found in real SAR images: (1) distributed Gaussian scattering and (2) point scattering. While a point scatterer can be treated as a deterministic complex-valued response of a dominant target in the scene, Gaussian scatterers are formed by a coherent superposition of an arbitrary number of sub-scatterers within a resolution cell. These individual contributions are unknown and cannot be reconstructed individually. For a sufficiently large number of sub-scatterers, and if there is no dominant scatterer in the resolution cell, the central limit theorem holds and the SAR pixel z can be assumed to be a circular Gaussian random variable. The Gaussian scattering model typically holds for homogeneous scenes, such as rural or vegetated areas, or almost for any land cover type for low-to-medium resolution SAR imagery.

The circular Gaussian statistic translates in random variations of the magnitude of the SAR image, generating a pseudo-noise behavior defined as "speckle". Being an intrinsic characteristic of the radar scattering mechanism, speckle can provide useful information about the imaged targets. In fact, for interferometric applications this information content is essential as speckle patterns with strong correlations between the images are a sign that the contributions of the individual scatterers in both their amplitude and phase relations within certain resolution cells are relatively similar, hence the scene is not decorrelated in the backscatter superposition of its elements, which is also reflected by a high coherence magnitude for this area. On the other hand, speckle for other applications more interested in image intensity often impairs the visual understanding and interpretation of SAR images and is therefore considered as noise and filtered out through despeckling procedures (Goodman, 1975).

Given two single-look complex images $\boldsymbol{u_1}$, $\boldsymbol{u_2}$, the interferogram $\boldsymbol{v}$ is computed as the complex conjugate product of the two: $\boldsymbol{v} = \boldsymbol{u_1} \cdot \boldsymbol{u_2}^*$ and given the circular Gaussian statistic of the SLCs, its PDF is given by:

$$PDF(\boldsymbol{v}) = \frac{1}{\pi^2 \det(\boldsymbol{C})} \exp\left(-\boldsymbol{u}^H \boldsymbol{C}^{-1} \boldsymbol{u}\right). \tag{9}$$

with $\boldsymbol{u} = \begin{pmatrix} u_1 \\ u_2 \end{pmatrix}$ and is fully characterized by its complex covariance matrix

$$\boldsymbol{C} = \mathrm{E}\left\{\boldsymbol{u}\boldsymbol{u}^H\right\} = \begin{pmatrix} \bar{I}_1 & \gamma_{12}\bar{I} \\ \gamma_{12}^*\bar{I} & \bar{I}_2 \end{pmatrix} \tag{10}$$

with $\bar{I} = \sqrt{\bar{I}_1 \bar{I}_2} = \sqrt{\mathrm{E}\left\{|\boldsymbol{u_1}|^2\right\} \mathrm{E}\left\{|\boldsymbol{u_2}|^2\right\}}$ and $\bar{I}_i$ as the expected value of the intensities over a set of pixels of the image u_i. With the definition of complex coherence

$$\gamma_{12} = \frac{\mathrm{E}\left\{\boldsymbol{u_1}\boldsymbol{u_2}^*\right\}}{\sqrt{\bar{I}_1 \bar{I}_2}} = |\gamma_{12}| \exp\left(j\phi_{12}\right) \tag{11}$$

the expected value of the interferogram $v = u_1 u_2^*$ can directly be found as

$$\mathrm{E}\{v\} = \sqrt{\bar{I}_1 \bar{I}_2}\, |\gamma_{12}| \exp\left(j\phi_{12}\right) \tag{12}$$

For the special case of multi-looking, where the expected value actually is approximated by averaging over neighboring pixels, the multi-looked interferogram $\tilde{v}$ thus actually can be written as:

$$\begin{aligned} \tilde{v} &\approx \sqrt{\bar{I}_1 \bar{I}_2}\, |\tilde{\gamma}_{12}| \exp\left(j\tilde{\phi}_{12}\right) \\ &\approx |\tilde{u}_1|\, |\tilde{u}_2|\, |\tilde{\gamma}_{12}| \exp\left(j\tilde{\phi}_{12}\right) \end{aligned} \tag{13}$$

with $\tilde{u}$, $\tilde{\phi}$ as the averaged magnitude and phase, and $\tilde{\gamma}$ as the local coherence estimate for an averaged pixel.

2.3. InSAR phase statistics

With the complex coherence γ_{12} defined in (11) it can be seen that ϕ_{12} is the phase of the interferogram pixel or interferometric phase. The coherence magnitude $|\gamma_{12}|$ is related to the noise content of the interferometric phase ϕ_{12}: As it was shown (Bamler and Hartl, 1998) the marginal PDF for the phase drawn from the joint PDF (9) can be written as

$$\begin{aligned} PDF(\phi_{12}) = \frac{1 - |\gamma_{12}|^2}{2\pi} \frac{1}{1 - |\gamma_{12}|^2 \cos^2\left(\phi_{12} - \phi_0\right)} \\ \left(1 + \frac{|\gamma_{12}| \cos(\phi_{12} - \phi_0) \arccos(-|\gamma_{12}| \cos(\phi_{12} - \phi_0))}{\sqrt{1 - |\gamma_{12}|^2 \cos^2\left(\phi_{12} - \phi_0\right)}}\right) \end{aligned} \tag{14}$$

From this rather involved expression two extreme cases w.r.t. $|\gamma_{12}|$ can be derived: for $|\gamma_{12}| = 0$ (minimum coherence) the phase is a uniform distribution, for $|\gamma_{12}| = 1$ (maximum coherence) the PDF tends to a Dirac delta distribution at mean phase ϕ_0.

3. InSAR data synthesis with multitask CNN

In this section, we present our solution for the generation of synthetic complex-valued InSAR data. Specifically, we address this problem with a multitask encoder–decoder convolutional neural network (MED-CNN). We exploit a complex-valued InSAR data model and introduce basic statistics on SAR images and the SAR interferometric noise. Thus, we provide a rationale for the multitask encoder–decoder architecture that we employ to synthesize the components of complex-valued InSAR data and the subsequent model to reproduce speckle and artificial phase noise on the synthesized components.

Our InSAR data synthesis approach thus is structured in two stages (see Fig. 2):

First, the basic multitask encoder–decoder architecture is based on the single-task generator structure, as similarly done in Baier et al. (2021). We extend it with a multitask learning approach and dedicated loss terms for InSAR data synthesis.

Second, to conclude our approach we disentangle the synthesized InSAR information to virtual single-look complex (SLC) images, simulate and re-apply speckle behavior onto their magnitudes and introduce interferometric phase noise based on the synthesized scene coherence magnitude.

Moreover, for this publication we define the term "renoising" as both the introduction of *speckle behavior*, although speckle is not considered noise as stated before, and the introduction of artificial *interferometric phase noise* to synthesized InSAR data.

3.1. Multitask InSAR image synthesis

Considering the discussed limitations of CNN signal synthesis, we propose a multimodal, multitask synthesis architecture for complex-valued, geocoded SAR interferograms inspired by the network architecture proposed in Baier et al. (2021). The basic encoder and decoder architecture and the connecting ResNet body for the generator is re-used, however, it is further extended to allow the prediction of scene reflectivities, achievable coherence magnitudes and interferometric phases of a given SAR scene. Moreover, we introduce a domain-specific training loss for the interferometric phase based on available DEM data. In contrast to Baier et al. (2021), we abandon the use of GANs as the basic network topology, skipping the use of a discriminator network for adversarial training, but employ a multitask encoder–decoder architecture. As our proposed solution skips the adversarial aspect of the network architecture instabilities observed during training of adversarial networks, such as mode collapse, are avoided.

As input features, we exclusively use DEM rasters and the semantic land cover maps, which are fed into a multitask CNN encoder–decoder architecture.

In this first stage of InSAR signal synthesis we rely on the assumption that the magnitude responses $|\boldsymbol{u_1}|$, $|\boldsymbol{u_2}|$ of both SLCs are almost similar. Thus we synthesize not both magnitudes directly but combine them to a scene reflectivity as introduced in Eq. (8). Disentanglement

to two synthesized SLCs $|\hat{u_1}|$, $|\hat{u_2}|$, related to each other by their interferometric phase $\hat{\phi_{12}}$, is then performed in the second renoising stage of InSAR signal synthesis.

To overcome the limitations of direct complex-valued synthesis in CNNs, in this work we split the task into independent decoder networks for the real-valued parts of scene reflectivity $\hat{\beta}$, coherence magnitude $|\hat{\gamma_{12}}|$, and phase images $\hat{\phi_{12}}$ of artificial SAR interferograms, fed by a common encoder network and a ResNet body that encodes input information to be sourced by all decoder outputs. While the decoder stage is composed by different heads allowing the generation of output variables with different physical meanings and dynamic ranges, the training with a common encoder and a multi-objective loss function ensure the interdependence between these variables.

Treating the different contributing real-valued channels in this way not only results in an architecture that is easy to use, but also allows us to understand its different physical output quantities. Since each of the decoder outputs has its own interpretable physical meaning, additions that affect the properties of these physical parameters can be applied directly to these individual elements of the network architecture.

3.2. InSAR data renoising

Based on the theoretical models in Sections 2.2 and 2.3, the approximation for speckle content and interferometric phase noise as characterized by its PDF in (14) is applied on a model discussed in Sica et al. (2021). The model operates on the synthesized scene reflectivity $\hat{\beta}$ and synthesizes the two constituent SLCs by

- respeckling the amplitude on multiplicative circular Gaussian processes, and
- renoising the synthesized interferometric phase $\hat{\phi_{12}}$ with respect to the estimated coherence magnitude $|\hat{\gamma_{12}}|$.

To simulate speckle and noise contributions, signals $\tilde{a}, \tilde{b}, \tilde{c}, \tilde{d}$ are drawn independently from Gaussian random processes $\mathcal{N}(\mu = 0, \sigma^2 = 1) \in \mathbb{R}^{N_v \times N_h}$. Approximating the resolution characteristics of the real SAR sensor for renoising, the Gaussian processes are filtered with a 2D convolution kernel $h \in \mathbb{R}^{30\times30}$ as

$$a = (\tilde{a} * h)/\sigma_{\tilde{a}*h} \tag{15}$$

$$\vdots$$

$$d = (\tilde{d} * h)/\sigma_{\tilde{d}*h} \tag{16}$$

Division by their own $\sigma_{\tilde{a}*h}, \ldots, \sigma_{\tilde{d}*h}$ ensures unit standard deviations for all filtered noise signals. This resolution approximation intentionally is performed on projected signals in geo-coordinates and not on signals in original sensor coordinates, inevitably having some limitations in matching the real resolution characteristics.

H is generated in the discrete Fourier domain by the inverse transformation of a Hamming-windowed response function with $f_{c,v} = f_{s,v}/2r_v$ as vertical, $f_{c,h} = f_{s,h}/2r_h$ as horizontal corner frequency, and $f_{s,v}$ as the vertical, $f_{s,h}$ as the horizontal spatial sampling frequency, respectively. The resolution parameters r_v, r_h are manually chosen for each dataset.

Applying the filtered random signals we retrieve two independent random circular Gaussian processes

$$\boldsymbol{x_1} = (a + jb)/\sqrt{2} \tag{17}$$

$$\boldsymbol{x_2} = (c + jd)/\sqrt{2} \tag{18}$$

Using the synthesized denoised scene reflectivity $\hat{\beta}$ and interferometric phase $\hat{\phi_{12}}$ the respeckled InSAR SLCs, coupled by their renoised interferometric phase, can be approximated (Sica et al., 2021) as

$$\boldsymbol{u_{\hat{1},r}} = \boldsymbol{x_1}\sqrt{\hat{\beta}} \tag{19}$$

Table 1
Encoder and ResNet body architecture.

Layers	No. out channels
$C_{7\times7}$, ReLU	64
$C^{\downarrow2}_{3\times3}$, SPADE normalization, ReLU	128
$C^{\downarrow2}_{3\times3}$, SPADE normalization, ReLU	256
$C^{\downarrow2}_{3\times3}$, SPADE normalization, ReLU	512
$C^{\downarrow2}_{3\times3}$, SPADE normalization, ReLU	1024
9 × ResNet Block	1024

Table 2
Decoder architecture.

Layers	No. out channels
Nearest Neighbor ↑ 2, ResNet block	512
Nearest Neighbor ↑ 2, ResNet block	256
Nearest Neighbor ↑ 2, ResNet block	128
Nearest Neighbor ↑ 2, ResNet block	64
$C_{3\times3}$, Sigmoid ($\hat{\beta}'$.$|\hat{\gamma_{12}}|$), Tanh ($\hat{\phi_{12}}$)	N_o

$$\boldsymbol{u_{\hat{2},r}} = \boldsymbol{x_1}\sqrt{\hat{\beta}}|\hat{\gamma_{12}}|\exp\left(-j\hat{\phi_{12}}\right) + \boldsymbol{x_2}\sqrt{\hat{\beta}\left(1 - |\hat{\gamma_{12}}|^2\right)} \tag{20}$$

3.3. Architecture

To implement the multitask architecture for our generative model, the decoder stage of the original generator is split into three individual decoders for $\hat{\beta}$, $\hat{\phi_{12}}$, $|\hat{\gamma_{12}}|$. All three decoders are fed with information generated by a common encoder network followed by a ResNet body that encodes the two inputs: DEM raster and semantic land cover data. As in the original architecture, the ResNet architecture is included to improve the expression and flow of information through the layers of the body (He et al., 2016). Each ResNet block in the generator network is composed from two subsequent convolutional layers with ReLU activations and SPADE normalization layers, and a residual skip connection.

Although the discriminator network from the reference GAN architecture is omitted we still use the term "generator" as a synonym for our entire multitask generative model for the rest of this paper.

Based on the positive experience in Baier et al. (2021), the semantic land cover data is not concatenated and introduced directly at the generator input, but is fed into the network by replacing the batch normalization layers (Odena et al., 2016) with SPADE normalization layers (Park et al., 2019) that present the land cover information to each layer of the network. SPADE layers are designed to transform a semantic mask into tensors that are modulated to pass information between subsequent layers of a CNN. In this way, semantic data can be combined with higher-level structural information and preserved throughout the network. In our experiments, introducing semantic information with SPADE layers leads to improved results, especially for the reflectivity and coherence channels, and to a lesser extent for the phase channel, but is retained as a replacement for batch normalization layers (Ioffe and Szegedy, 2015) in the encoder and all decoders.

The encoder directly follows the architecture of Baier et al. (2021): DEM input images are downsampled through a cascade of stepped convolution layers. Within the stages, feature maps are normalized by SPADE layers and fed through ReLU activations, finally feeding a 9-stage ResNet body.

All decoders are fed from the same output of the ResNet body. The upsampling stages in the decoders use nearest-neighbor interpolation to avoid checkerboard artifacts in the feature maps. Each upsampling layer is followed by a ResNet block with SPADE normalization that synthesizes the feature maps to output images at their original resolution (see Tables 1 and 2).

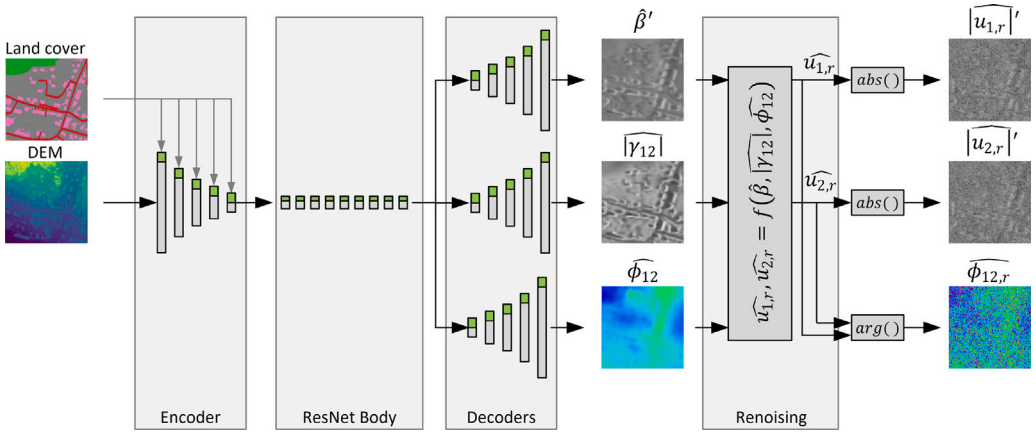

Fig. 2. Principal synthesis architecture using a common encoder and ResNet body feeding four separate generator networks, generating scene reflectivity $\hat{\beta}$, coherence magnitude estimate $|\hat{\gamma_{12}}|$, and interferometric phase $\hat{\phi_{12}}$. DEM information is presented at the input, semantic land cover information is fed indirectly to SPADE normalization layers at all stages of the networks. After generating those channels renoised versions of the complex-valued SLCs $\hat{\boldsymbol{u}}_{1,r}$, $\hat{\boldsymbol{u}}_{2,r}$ are produced as a function of $\hat{\beta}$ for the speckle characteristics and $|\hat{\gamma_{12}}|$ for the phase noise. Their normalized magnitudes $|\hat{\boldsymbol{u}}_{1,r}|'$, $|\hat{\boldsymbol{u}}_{2,r}|'$ and interferometric phase $\hat{\phi_{12}}$ are shown in this diagram.

To get to the complex-valued, renoised SLCs $\hat{\boldsymbol{u}}_{1,r}, \hat{\boldsymbol{u}}_{2,r}$ the output channels of the multitask generator are used in the renoising step: Applying the outputs to the model as given by Eqs. (19) and (20) both complex SAR SLCs can be synthesized. The SLCs are respeckled and are phase-coupled by their noisy interferometric phase

$$\hat{\phi_{12,r}} = \arg\left(\hat{\boldsymbol{u}}_{1,r} \cdot \hat{\boldsymbol{u}}_{2,r}^{*}\right) \tag{21}$$

3.4. Loss functions

For the multitask generator different loss configurations were evaluated, and finally, for all four decoder networks, we found a common multitask loss term $\mathcal{L}_{MT}$, which combines the classical L1 and L2 loss terms. All decoders use the same hyperparameters $\lambda_{L1}, \lambda_{L2}$. The loss term is applied individually to the four decoder outputs.

With y as the real and $\hat{y}$ as the synthesized output over N pixels, the L1 loss is defined as

$$\mathcal{L}_{L1} = \frac{1}{N}\sum_{N}|\hat{y} - y| \tag{22}$$

and the L2 loss is given as

$$\mathcal{L}_{L2} = \frac{1}{N}\sum_{N}(\hat{y} - y)^2 \tag{23}$$

which are combined to the multitask loss term

$$\begin{aligned}\mathcal{L}_{MT} &= \lambda_{L1}\mathcal{L}_{L1} + \lambda_{L2}\mathcal{L}_{L2} \\ &= \lambda_{L1}\frac{1}{N}\sum_{N}|\hat{y} - y| + \lambda_{L2}\frac{1}{N}\sum_{N}(\hat{y} - y)^2\end{aligned} \tag{24}$$

To refine the phase synthesis, an additional, dedicated loss term $\mathcal{L}_{\phi_{DEM}}$ with its own hyperparameter $\lambda_{\phi_{DEM}}$ was included in the experiments to further constrain the phase decoder with height information from the input DEM. The scene height and (unwrapped) interferometric phase are closely related by the *phase-to-height sensitivity* as it was shown in Eq. (6). Therefore, a new loss term is formed by assuming that their statistical distributions are similar for a given scene patch. To apply this loss, a virtual phase band ϕ_{DEM} is generated from the DEM information z by matching the means and standard deviations of both distributions.

Constructing the loss, we generate a virtual continuous (unwrapped) DEM phase band $\tilde{\phi}_{DEM}$ by estimating the circular mean $\mu_{\phi_{12}}$ and circular angular standard deviation $\sigma_{\phi_{12}}$ of the phase and the mean μ_z and standard deviation σ_z of the DEM information as given by

$$\tilde{\phi}_{DEM} = (z - \mu_z)\frac{\sigma_{\phi_{12}}}{\sigma_z} + \mu_{\phi_{12}} \tag{25}$$

with the circular mean (Fisher, 1995) defined as

$$\mu_{\phi_{12}} = \arg\left(\sum_{N_v}\sum_{N_h}\exp\left(j\phi_{12}\right)\right) = \arg\left(R_{\phi_{12}}\right), \tag{26}$$

with N_v as the number of samples in vertical, N_h as the number of samples in horizontal direction, respectively.

Starting from (26), $\bar{R}_{\phi_{12}}$ is defined as the mean resultant length of the complex phasor $R_{\phi_{12}}$, associated with the mean direction of ϕ_{12}, and is found as

$$\bar{R}_{\phi_{12}} = \frac{|R_{\phi_{12}}|}{N_v N_h}. \tag{27}$$

The circular standard deviation is given by

$$\sigma_{\phi_{12}} = \sqrt{-2\ln\bar{R}_{\phi_{12}}}. \tag{28}$$

For distributions with large $\bar{R}_{\phi_{12}}$ the circular standard deviation can be approximated to

$$\sigma_{\phi_{12}} \approx \sqrt{2\left(1 - \bar{R}_{\phi_{12}}\right)}, \tag{29}$$

sometimes also referred to as angular deviation (Zar, 2014), which in experimentation practice led to more stable results and, as a bonus, is computationally less expensive. Introducing the circular statistic expressions into (25) the virtual continuous DEM phase band therefore can be retrieved as

$$\begin{aligned}\tilde{\phi}_{DEM} &= (z - \mu_z)\frac{\sigma_{\phi_{12}}}{\sigma_z} + \mu_{\phi_{12}} \\ &= \frac{z - \mu_z}{\sigma_z}\sqrt{2\left(1 - \frac{|R_{\phi_{12}}|}{N_v N_h}\right)} + \arg\left(R_{\phi_{12}}\right)\end{aligned} \tag{30}$$

Generally, DEM height differences are expected to exceed the *height-of-ambiguity* as given by (7) of an InSAR acquisition. To account for that, the continuous phase $\tilde{\phi}_{DEM}$ is re-wrapped to a virtual wrapped DEM phase band

$$\phi_{DEM} = \arg\left(\exp\left(j\tilde{\phi}_{DEM}\right)\right). \tag{31}$$

Table 3
SAR datasets.

	S-1 Bavaria	GeoNRW TDX
Platform	Sentinel-1	TanDEM-X
Acquisition dates	Mar 29, Apr 4, 2021	Feb 28, 2020
Mode	IW	Stripmap
Polarization	VV	HH
Pass	Ascending	Descending
Spatial sampling	30 m × 30 m	1 m × 1 m

Combining generated phase $\hat{\phi_{12}}$ and wrapped virtual DEM phase ϕ_{DEM} with an L2 criterion, the dedicated phase loss

$$\mathcal{L}_{\phi_{DEM}} = \frac{1}{N}\sum_{N}\left(\hat{\phi_{12}} - \phi_{DEM}\right)^2 \tag{32}$$

is formed.

Injecting Eq. (32) to Eq. (24), the full loss term for the phase decoder results in the following

$$\begin{aligned}
&\mathcal{L}_{MT_\phi} = \lambda_{L1}\mathcal{L}_{L1} + \lambda_{L2}\mathcal{L}_{L2} + \lambda_{\phi_{DEM}}\mathcal{L}_{\phi_{DEM}} \\
&= \lambda_{L1}\frac{1}{N}\sum_{N}\left|\hat{\phi_{12}} - \phi_{12}\right| + \lambda_{L2}\frac{1}{N}\sum_{N}\left(\hat{\phi_{12}} - \phi_{12}\right)^2 + \\
&\qquad \lambda_{\phi_{DEM}}\frac{1}{N}\sum_{N}\left(\hat{\phi_{12}} - \phi_{DEM}\right)^2 \\
&= \lambda_{L1}\frac{1}{N}\sum_{N}\left|\hat{\phi_{12}} - \phi_{12}\right| + \lambda_{L2}\frac{1}{N}\sum_{N}\left(\hat{\phi_{12}} - \phi_{12}\right)^2 + \\
&\qquad \lambda_{\phi_{DEM}}\frac{1}{N}\sum_{N}\Bigg(\hat{\phi_{12}} - \\
&\arg\left(\exp\left(j\left[\frac{z-\mu_z}{\sigma_z}\sqrt{2\left(1-\frac{|R_{\phi_{12}}|}{N_v N_h}\right)} + \arg\left(R_{\phi_{12}}\right)\right]\right)\right)\Bigg)^2
\end{aligned} \tag{33}$$

4. Datasets

In this section, we present the materials, including the datasets used and the preprocessing applied. In the present investigation, we used the following two datasets:

- Sentinel-1 Bavaria dataset
- Tandem-X GeoNRW dataset

The experiments were based on both a medium-resolution Sentinel-1 and a high-resolution Tandem-X-CoSSC dataset, demonstrating the generality of our approach (see Table 3).

4.1. Sentinel-1 Bavaria dataset

For training and testing the multitask generator model with medium resolution C-band SAR data, a repeat-pass Sentinel-1 interferogram on SLCs taken on Mar 29 and Apr 4, 2021 of a predominantly rural area of the German state of Bavaria is augmented with Digital Elevation Model (DEM) and semantic land cover data. Height information is derived from the Copernicus DEM as a Digital Surface Model (DSM) with 30 m ground resolution (GLO-30). In addition to bare topographic height, DSMs represent the "populated" surface of the Earth, including not only topography but also height information from buildings, infrastructure, and vegetation. Semantic land cover information is derived from the ESA WorldCover map at 10 m resolution (Zanaga et al., 2021).

The DSM information is encoded in meters above sea level, while the land cover data is encoded according to the level 1 classes of the ESA WorldCover map, which include different vegetation types, forest, cropland, urban structures, permanent water bodies, and snow and ice. The classes are encoded as discrete information and have been translated from a color map into an RGB image.

The SLC images u_1, u_2 form the basis for the slant-range interferogram v, which is added to the dataset in a terrain-corrected form. The SLCs were acquired by the Sentinel-1A/B SAR instruments on March 29 and April 4, 2021, respectively, in IW (interferometric wide-swath) mode. The area of interest extends from 11°E to 13°E, beginning in an area north of Munich, Germany, and is covered by the IW1 and IW2 subswaths.

For both SLC images, the VV polarization is chosen: therefore, both surface scattering and (to some extent) double-bounce dihedral scattering are expected in these images.

Based on the SLC images, several pre-processing steps have to be performed in the ESA SNAP 8.0 environment to obtain geo-coded real-sensor interferogram patches v in ground coordinates: The slant-range real-sensor interferogram v is generated with u_1, u_2 co-registration and back-geocoding (SRTM 3s), forming the raw interferogram with coherence estimation and multilooking to a spatial sampling of about 70 m × 70 m in slant range and azimuth and applying a subsequent TOPSAR deburst and subswath merge operation. Finally, the real and imaginary parts of the interferogram are terrain corrected (SRTM 3s).

To build the final dataset, DEM information is selected at random positions in the area of interest for image patches of 1000 *times* 1000 pixels. The corresponding image patches for the interferogram channels and land cover information are selected and reprojected onto the same image grid.

For the renoising operation a single-look version of the Sentinel-1 Bavaria dataset was generated in parallel with, apart from the multilook operation, the same processing steps and parameters. To compare the renoising results with real-sensor data this dataset was used.

4.2. TanDEM-X GeoNRW dataset

To generalize the experiments with the multitask generator to high resolution X-band SAR data, a TanDEM-X bistatic interferogram in and around the area of Dortmund, Germany, is combined with corresponding DEM and land cover information from the GeoNRW (Baier et al., 2020) dataset. GeoNRW already contains both DEM and land cover information in a 1 m × 1 m spatial sampling and in patches of 1000 × 1000 pixels.

Although GeoNRW covers a large number of cities in the German state of North Rhine-Westphalia, to demonstrate the basic approach, we only consider the cities of Dortmund and Hagen with bistatic interferometric data from TanDEM-X. The resulting dataset with limited regional coverage will be referred to as GeoNRW TDX in this paper. With the availability of TanDEM-X data with a wider regional coverage, GeoNRW TDX can always be extended in a future work.

Covering the area of Dortmund, Germany, a TanDEM-X bistatic interferogram is processed from a Coregistered Single-Look Slant range Complex (CoSSC) product acquired in stripmap (SM) mode on February 28, 2020. The acquisition was performed in HH polarization and in single-pol SM mode, with a spatial resolution of 3 m × 3 m in slant range and azimuth.

To preserve the spatial resolution, no multilooking is applied to the two SLC images u_1, u_2 in the GeoNRW TDX dataset. As a preprocessing step, the flat-Earth and topographic phases are removed from the CoSSC stack. The magnitude information is included "as is" in the dataset without further processing prior to geocoding, while the interferometric phase ϕ_{12} and coherence $|\gamma_{12}|$ are estimated by using the CNN-based estimator Phi-Net (Sica et al., 2021), which provides the best approximation for noise- and distortion-free phase and coherence images.

The resulting maps are geocoded in SNAP 8.0 and combined with DEM and land cover information from GeoNRW, covering the cities of Dortmund and Hagen. As a final step, the geocoded CoSSC interferogram components are reprojected onto the grid of the GeoNRW patches.

Table 4
Training and renoising parameters for both datasets and training stages, with N_t as the number of training patches, m, M as parameters for log-scaling and normalization, resolution parameters r_v, r_h, and different λ values as weights for their respective loss terms. For experiments without $\mathcal{L}_{\phi_{DEM}}$ applied to the phase decoder, $\lambda_{\phi_{DEM}} = 0$.

	S-1 Bavaria	GeoNRW TDX
N_t	500	542
Mini-batch	8	8
m_β	1.0	3.0
M_β	9.3	16.0
r_v	1.0	1.7
r_h	1.0	1.7
Main stage		
Epochs	150	150
λ_{L1}	100	100
λ_{L2}	1	1
$\lambda_{\phi_{DEM}}$	1	3
Consolidation stage		
Epochs	20	20
λ_{L1}	10	10
λ_{L2}	1	1
$\lambda_{\phi_{DEM}}$	0.2	0.3

5. Experiments

In this section, we show the results of our investigation on the synthesis of complex interferometric data from spatial metadata such as the Digital Elevation Model and Land Cover.

The experiments for both datasets were run on an Intel(R) Xeon(R) Platinum 8168 2.7 GHz CPU and an Nvidia Tesla V100-SXM3 GPU with 32 GB of RAM. The networks were implemented using the PyTorch 1.10.2 framework. Training was performed for 150 epochs in the main and for 20 epochs in the consolidation phase, with a mini-batch size of 8 for both datasets. We experimentally found the best balance between the different losses. All experiments were generated at a learning rate of $\alpha = 10^{-4}$, and Adam optimization with $\beta_1 = 0$, $\beta_2 = 0.9$. Table 4 summarizes the training and network parameters that were applied to achieve the results as they are presented in this paper.

5.1. Training and test

For the training data set, $N_t = 500$ patches (1000×1000 pixels) were selected from the Sentinel-1 Bavaria scene at random locations. For the GeoNRW TDX dataset, $N_t = 542$ patches (1000×1000 pixels) were selected for training and distributed over the area of the cities of Dortmund and Hagen according to their UTM coordinates. From these patches, further random sub-patches of $N_v \times N_h = 256 \times 256$ pixels were selected during training for both datasets to match the operational dimensions of the encoder and decoder networks.

For the Sentinel-1 Bavaria dataset, patches were drawn from the same scene at random positions and used for testing. For the GeoNRW TDX dataset, a similar approach was chosen, with patches selected both from the cities of Dortmund and Hagen that were used to validate the approach based on high-resolution InSAR data.

The network training was split into a *main stage* with full loss hyperparameters λ and a subsequent *consolidation stage* with attenuated hyperparameters. Whereas in the main training stage basic network convergence is achieved, the consolidation stage is predominantly meant to further improve precision and the synthesis of fine-structural details in all decoders. This strategy is similar to a training regime with a decaying learning rate, however, allows a finer control of the different loss terms contributions.

The multitask loss terms $\mathcal{L}_{MT}$ for the magnitude and coherence magnitude decoders and $\mathcal{L}_{MT_\phi}$ for the phase decoder are individually applied to the decoder outputs. Loss hyperparameters λ_{L1}, λ_{L2} in both training stages were found experimentally. Besides the typical L2 (MSE) loss, which minimizes the root mean square error between real and synthesized patches and is inherently stable, the L1 loss term generally is more robust to outliers in the training examples. Moreover, in our experiments, the L1 loss also enhanced fine structural details in the generated patches and, eventually, turned out to be the loss term having a higher influence. Both lambda hyperparameters were tuned accordingly for training stability and the desired result. Once a stable setup for λ_{L1} and λ_{L2} had been retrieved, $\lambda_{\phi_{DEM}}$ was optimized.

As an addition and to conclude the experiments, a small qualitative ablation study on the individual effects of DEM and land cover input information was performed on the GeoNRW TDX dataset, with either land cover information presented or removed from the input of the generator network. Other than preserving or removing land cover information from the input, the same network training strategy with $\mathcal{L}_{L1}$, $\mathcal{L}_{L2}$ and $\mathcal{L}_{\phi_{DEM}}$ losses was applied for both runs.

5.2. Dataset scaling and normalization

To match the value ranges of the decoder output activations, several strategies have been found for the channels. To match the $[0, 1]$ value range of the Sigmoid decoder output activation, scene reflectivity is logarithmically scaled and normalized according to $\beta' = \left(\ln\left(\beta + 10^{-3}\right) - m_\beta\right) / \left(M_\beta - m_\beta\right)$. Coherence magnitudes $|\gamma_{12}|$ within $[0, 1]$ correspond directly to the Sigmoid activation range and were applied to the network without further transformation. Phase values were scaled to $[-1, 1[$ by $\phi_{12} = \tilde{\phi}_{12}/\pi$ to match their Tanh output activation. Normalization values were estimated from the dataset distributions and manually tuned.

As in the original architecture, DEM/DSM input information for both datasets is applied directly as heights in meters without further scaling or normalization. During training, the network automatically adapts to an appropriate normalization of the input. Semantic land cover information is applied as a class index to the SPADE layers in their native resolutions. Since land cover data is represented in discrete classes, nearest neighbor interpolation is used for resampling to preserve the correct class information.

Moreover, to further reduce speckle contribution and noise and to train on denoised datasets the scene reflectivity β is not directly derived and applied as in Eq. (8) but denoised by a 3×3 boxcar filter before scaling and normalization.

5.3. Performance metrics

To quantify the performance of the experimental results for the multitask generator we are applying different metrics:

RMSE The Root Mean Squared Error (RMSE) as defined as

$$\mathrm{RMSE}(\hat{y}, y) = \sqrt{\frac{1}{N}\sum_{i=1}^{N} \|\hat{y}_i - y_i\|^2} \tag{34}$$

with N as number of all seen image patches and $\hat{y}$ as the synthesized, y as the real image patches.

PSNR The Peak Signal to Noise Ratio (PSNR) as defined as

$$\mathrm{PSNR}(\hat{y}, y) = 10 \cdot \log_{10}\left(\frac{\mathrm{MAX}_y^2}{\mathrm{MSE}(\hat{y}, y)}\right) \tag{35}$$

with MAX_y as the maximum value range of image y, MSE as

$$\mathrm{MSE}(\hat{y}, y) = \frac{1}{N}\sum_{i=1}^{N} \|\hat{y}_i - y_i\|^2, \tag{36}$$

N number of all seen image patches and $\hat{y}$ as the synthesized, y as the real image patches.

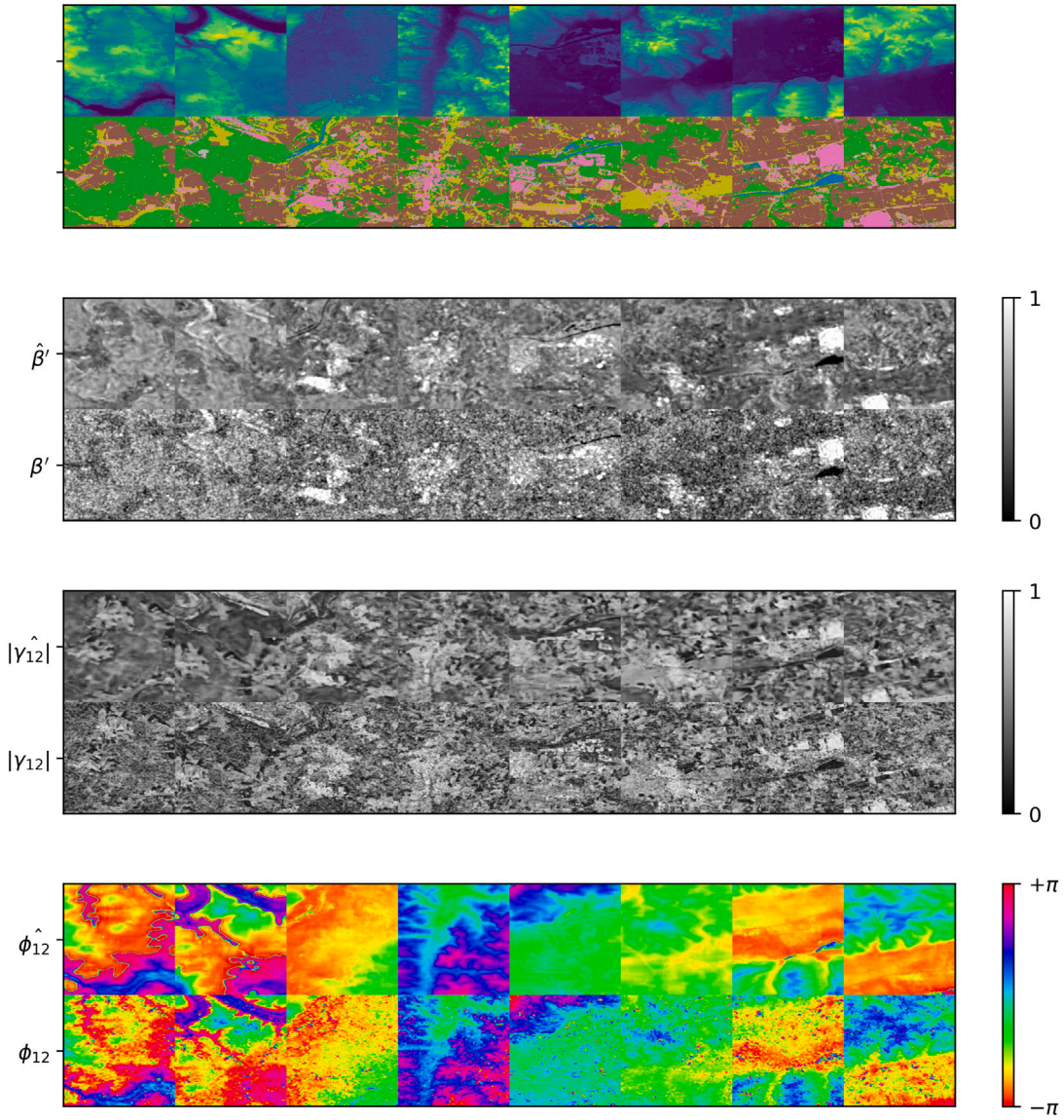

Fig. 3. Example results of the multitask network described in this paper for the Sentinel-1 Bavaria dataset, training for 150 + 20 epochs, $\mathcal{L}_{MT}$ with L1 and L2 loss only. Columns present different scene examples. Rows top to bottom: DEM input, land cover input, synthesized log-scaled and normalized reflectivity $\hat{\beta}'$, real reflectivity β', synthesized coherence magnitude $|\hat{\gamma}_{12}|$, real coherence magnitude $|\gamma_{12}|$, synthesized interferometric phase $\hat{\phi}_{12}$, real interferometric phase ϕ_{12}.

5.4. Results

In the following, the results obtained for the Sentinel-1 Bavaria dataset and the TanDEM-X GeoNRW dataset are shown separately.

We evaluated the performance of the proposed method using qualitative and quantitative metrics. The former consists of a visual assessment of the spatial features in the generated patches, while the latter consists of the use of performance metrics that can compare the generated result with the noisy real interferometric data, such as the root mean squared error (RMSE) and the peak signal-to-noise ratio (PSNR).

As the added speckle and phase noise is and cannot be coherent with the noise content of the real sensor images, which automatically results in lower performance figures compared to the originally synthesized results, we refrain from applying those performance metrics to the renoised images but present only their visual impressions in Figs. 5 and 8.

In Figs. 3 and 4, we show the generated scene reflectivity, phase, and coherence patches resulting from our network, together with the corresponding real data for the S-1 Bavaria dataset with and without the use of the additional loss term $\mathcal{L}_{\phi_{DEM}}$. Similarly, Figs. 6 and 7 show the results for the GeoNRW TDX dataset in the two cases mentioned above.

Some artefacts are observed at the $-\pi, \pi$ phase wrap in Figs. 3 and 4. These are due to the limitations of the decoder in representing instantaneous phase discontinuities. Due to the band-limited nature of the network, sharp transitions are difficult to reproduce, sometimes resulting in a slight slope at these discontinuities. A possible solution could be to feed the real and imaginary components of the complex phasor into the network separately, which warrants further investigation to improve phase wrap handling.

From a qualitative visual assessment of the synthesized maps, it can be seen that the synthetic data reflect the general trend of the real data for all target quantities, including the preservation of point scatter and sharp edges. The results show that the trained model is able to reproduce fairly similar backscatter, coherence, and phase values, including both geometric and radiometric effects typical of SAR data.

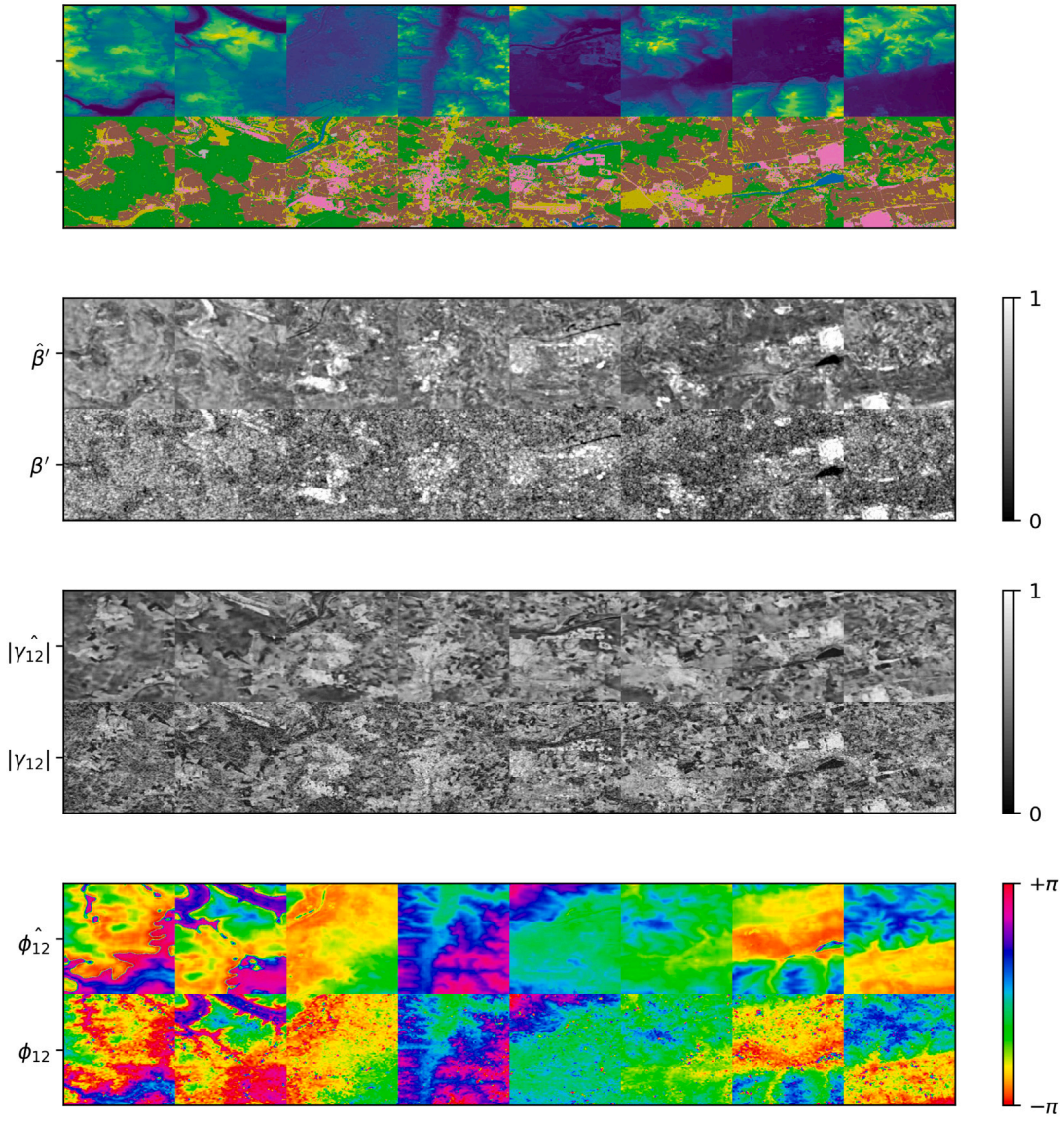

Fig. 4. Example results of the multitask network described in this paper for the Sentinel-1 Bavaria dataset, training for 150 + 20 epochs, $\mathcal{L}_{MT}$ including $\mathcal{L}_{\phi_{DEM}}$.

Table 5
Average performance metrics for the S-1 Bavaria test dataset results shown in Figs. 3 and 4, using a bare $\mathcal{L}_{MT}$ loss and a combined $\mathcal{L}_{MT_\phi}$ including the $\mathcal{L}_{\phi_{DEM}}$ loss term.

	β'	ϕ_{12}	$\lvert\gamma_{12}\rvert$
RMSE	15.837	82.608	31.220
PSNR	24.223	27.590	18.366
$\mathcal{L}_{MT} + \mathcal{L}_{\phi_{DEM}}$			
RMSE	15.947	75.015	31.323
PSNR	24.172	29.131	18.351

Table 6
Average performance metrics for the GeoNRW TDX test dataset results shown in Figs. 6 and 7, using a bare $\mathcal{L}_{MT}$ loss and a combined $\mathcal{L}_{MT_\phi}$ including the $\mathcal{L}_{\phi_{DEM}}$ loss term.

	β'	ϕ_{12}	$\lvert\gamma_{12}\rvert$
RMSE	20.495	50.263	44.504
PSNR	22.231	33.707	15.652
$\mathcal{L}_{MT} + \mathcal{L}_{\phi_{DEM}}$			
RMSE	18.045	44.089	40.379
PSNR	23.319	34.842	16.597

In the results for the GeoNRW TDX dataset in Figs. 6 and 7, geometric effects can be seen in the reflectivity and coherence maps. It is possible to observe shadow effects that simulate typical occluded targets in urban scenarios such as the one considered in this dataset. Similarly, we also observe radiometric effects such as multiple bounces, which are visible as brighter edges on a single side of the buildings, simulating the typical backscatter mechanism that occurs on the building facades exposed to the sensor illumination. The coherence maps closely follow the behavior of the reflectivity maps, as expected for urban scenarios. Clean reflectivity returns also show high coherence, while shadow areas have low coherence values. Similarly, phase values

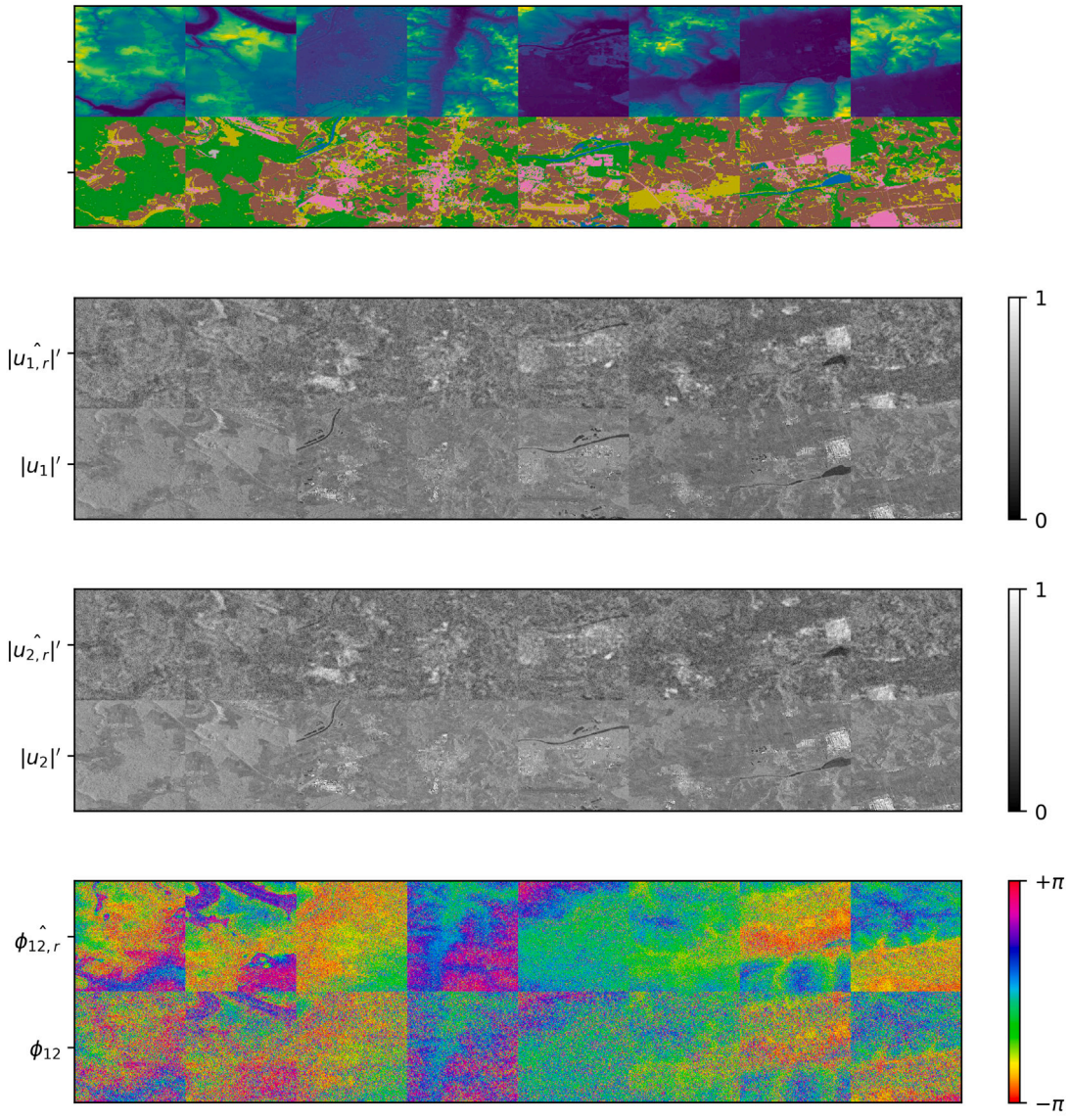

Fig. 5. Renoised results of the multitask network for the Sentinel-1 Bavaria dataset, decomposed to the interferogram SLC magnitudes and their interferometric phase, training for 150 + 20 epochs, $\mathcal{L}_{MT}$ including $\mathcal{L}_{\phi_{DEM}}$. Rows top to bottom: DEM input, land cover input, synthesized log-scaled and normalized SLC magnitude $|\hat{\boldsymbol{u}_1}|'$, real SLC magnitude $|\boldsymbol{u}_1|'$, synthesized SLC magnitude $|\hat{\boldsymbol{u}_2}|'$, real SLC magnitude $|\boldsymbol{u}_2|'$, synthesized interferometric phase $\hat{\phi_{12}}$, real interferometric phase ϕ_{12}.

show robust behavior on average across the image and follow a smooth variation of DEM values.

In this regard, the comparison of Figs. 6 and 7, with the latter considering the introduction of the $\mathcal{L}_{\phi_{DEM}}$, shows how this additional loss greatly increases the performance of our algorithm for interferometric phase generation. Nevertheless, small phase variations on small regions, such as buildings, are not fully synthesized. This is not surprising since, on average, only a few buildings show this behavior likely due to the fact that, in urban scenarios, phase centers can vary randomly due to the overlap of different backscattered signals within the same resolution cell. Since this information is usually not described in urban scenarios and is not provided in the input to the network, a mismatch between input features and reference data affects the ability of the network to reproduce this phenomenon.

Comparably good results were obtained for the Sentinel-1 Bavaria dataset, demonstrating the generalizability of the algorithm to medium resolution data. Similar to the previous experiments, we observe an excellent simulation of strong scatterers and multi-bounces over urban areas. In addition, we observe how well specular reflections over water are generated, exactly as observed in the real data and as expected for such a land cover class. The synthesized coherence maps closely follow the expected trends by correctly simulating coherence averages and preserving spatial detail. These generated coherence values match very well with the expected values given the land cover class. Finally, we observe that the synthesized phase closely follows the DEM variations and fairly matches the real data with good resolution preservation.

These observations are also supported by quantitative metrics. Table 5 presents the results for the RMSE and PSNR metrics for the Sentinel-1 Bavaria dataset, while Table 6 includes those for the GeoNRW TDX dataset. On the RMSE metric, it is evident that the introduction of the $\mathcal{L}_{\phi_{DEM}}$ term brought a significant reduction in the mean error for the interferometric phase ϕ_{12} in both datasets. Interestingly, although this loss term is applied to the simulated interferometric phase, according to the multitask learning construct, the

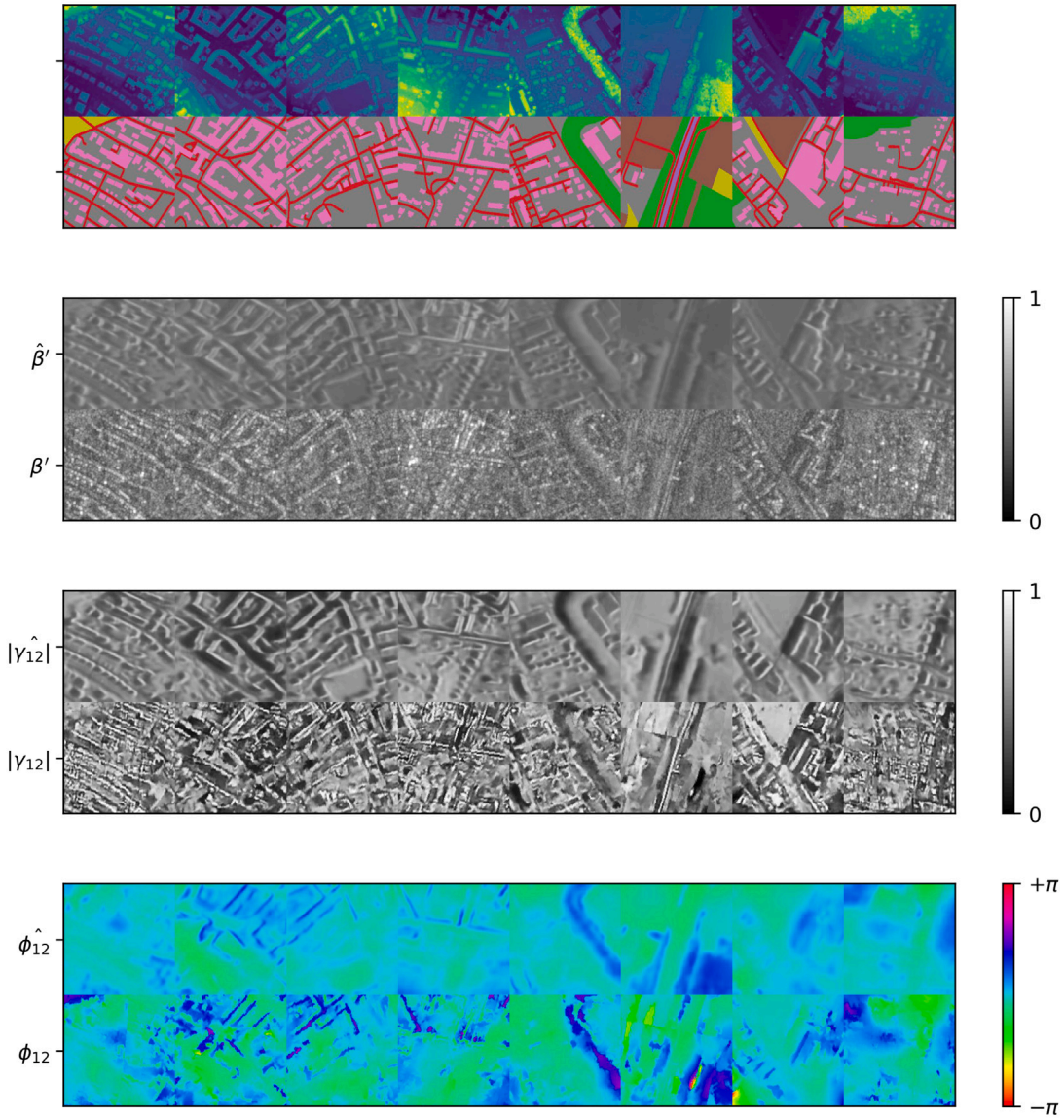

Fig. 6. Example results for the GeoNRW TDX dataset, training for 150 + 20 epochs, $\mathcal{L}_{MT}$ with L1 and L2 loss only.

network optimization takes place for all the branches of the network and not only for the one dedicated to the generation of the phase. As a consequence, the RMSE and PSNR values for the reflectivity branch β and especially for the coherence magnitude estimate $|\gamma_{12}|$ are significantly improved for the GeoNRW TDX dataset. On the Sentinel-1 Bavaria dataset, this effect seems to be particularly pronounced on interferometric phase as well, but here at a slight cost on the quality of reflectivity and coherence.

In order to estimate the individual influences of DEM and land cover information on the synthesis quality, we additionally performed an ablation study on the GeoNRW dataset, in which we omit the land cover information. As shown in Fig. 9, the reconstruction is degraded compared to the proposed solution. However, basic illumination and shadowing effects are still present in the reflectivity and coherence magnitude channels, while fine structural details and strong backscatter returns from roads or buildings seem to be mainly brought by the additional information from the land cover data.

6. Discussion

With the presented experiments, we have shown that the multitask learning strategy and, in particular, the proposed encoder–decoder network architecture can achieve excellent results for the synthesis of InSAR data, both in terms of accuracy and generalization between different SAR sensor modes and sensor operating wavelengths. Every single output branch presents high accuracy and all the main geometric properties of InSAR data are preserved, as it is possible to observe the geometric effects typical for InSAR data in the reflectivity and coherence maps of Figs. 6 and 7. Here, the presence of shadow effects as well as radiometric effects indicates that the network is able to correctly infer from the land cover and the DEM the right underlying physical scattering mechanism typical for occluded targets or multiple bounces. Similarly, the decoder branch dedicated to interferometric phase synthesis shows results that closely follow the DEM variations, implying that the network is able to encode the input DEM variation into a phase variation and therefore has implicitly learned the InSAR

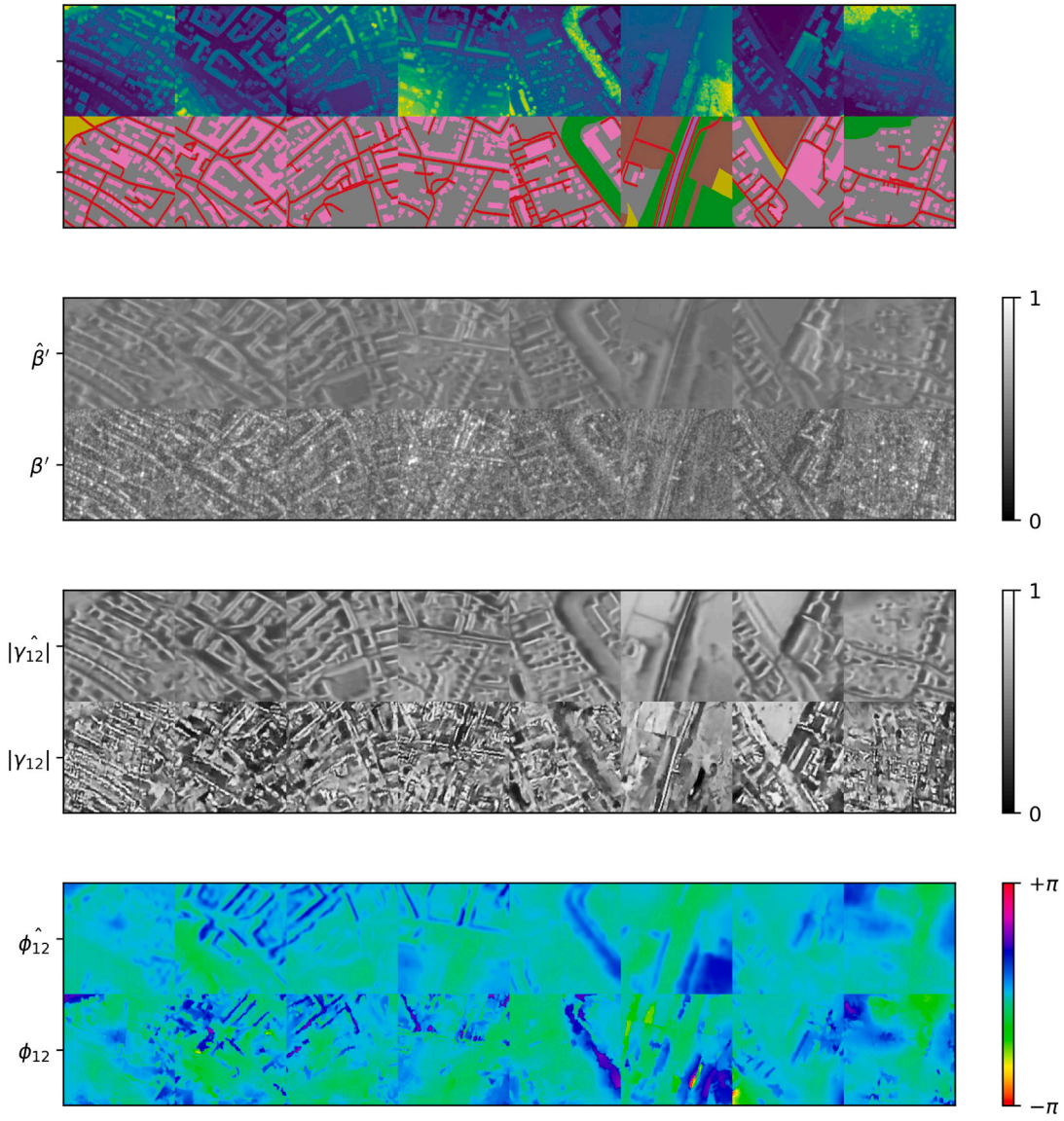

Fig. 7. Example results of the multitask network described in this paper for the GeoNRW TDX dataset, training for 150 + 20 epochs, $\mathcal{L}_{MT}$ including $\mathcal{L}_{\phi_{DEM}}$.

phase-height model of Eq. (6). This behavior indicates that the encoder–decoder network, in combination with the proposed datasets, is a successful choice for InSAR data generation.

We also provide evidence for the importance of multitask learning for our goal. We performed experiments with and without the $\mathcal{L}_{\phi_{DEM}}$ term, which is primarily intended to improve phase synthesis. The results show that not only the phase synthesis is improved, as expected, but also the whole network delivers better results. The reason is that while the loss is applied only to the output of the phase decoder, all network weights are updated according to this loss. Thus, on the one hand, we prove the validity of our proposed multitask learning approach, and on the other hand, we can deduce that the use of the $\mathcal{L}_{\phi_{DEM}}$ term improves the ability of the network to learn the DEM dependence and thus the InSAR phase-height model of Eq. (6). At the same time, this experiment tells us more about the relative importance between input features. In fact, without the $\mathcal{L}_{\phi_{DEM}}$ term, while the phase synthesis is distorted, the reflectivity and coherence synthesis still work as desired. This is due to the fact that the reflectivity and coherence are highly correlated with land cover, which is used as a shortcut proxy to get a good result. We have shown further evidence of this behavior with an ablation study in which the land cover information was not used. From the results shown in Fig. 9 it is clear how the DEM encodes information about the geometric effects typical of an InSAR image. However, additional details can only be generated by using the land cover information, which provides further information about the scattering mechanism and decorrelation effects. This indicates that land cover is initially the dominant input feature, but it cannot explain the whole InSAR acquisition model, which can be learned by using the $\mathcal{L}_{\phi_{DEM}}$ term instead. Consequently, when using this additional term, the DEM input feature takes on more weight with respect to the land cover and becomes of paramount importance for the performance of the proposed architecture.

Finally, based on the comparative analysis of the results derived from the two different datasets, it can be deduced that the proposed multitask model exhibits remarkable generalizability across different data types characterized by diverse resolutions, operative bands, and acquisition geometries. This result confirms the effectiveness and robustness of the proposed methodology.

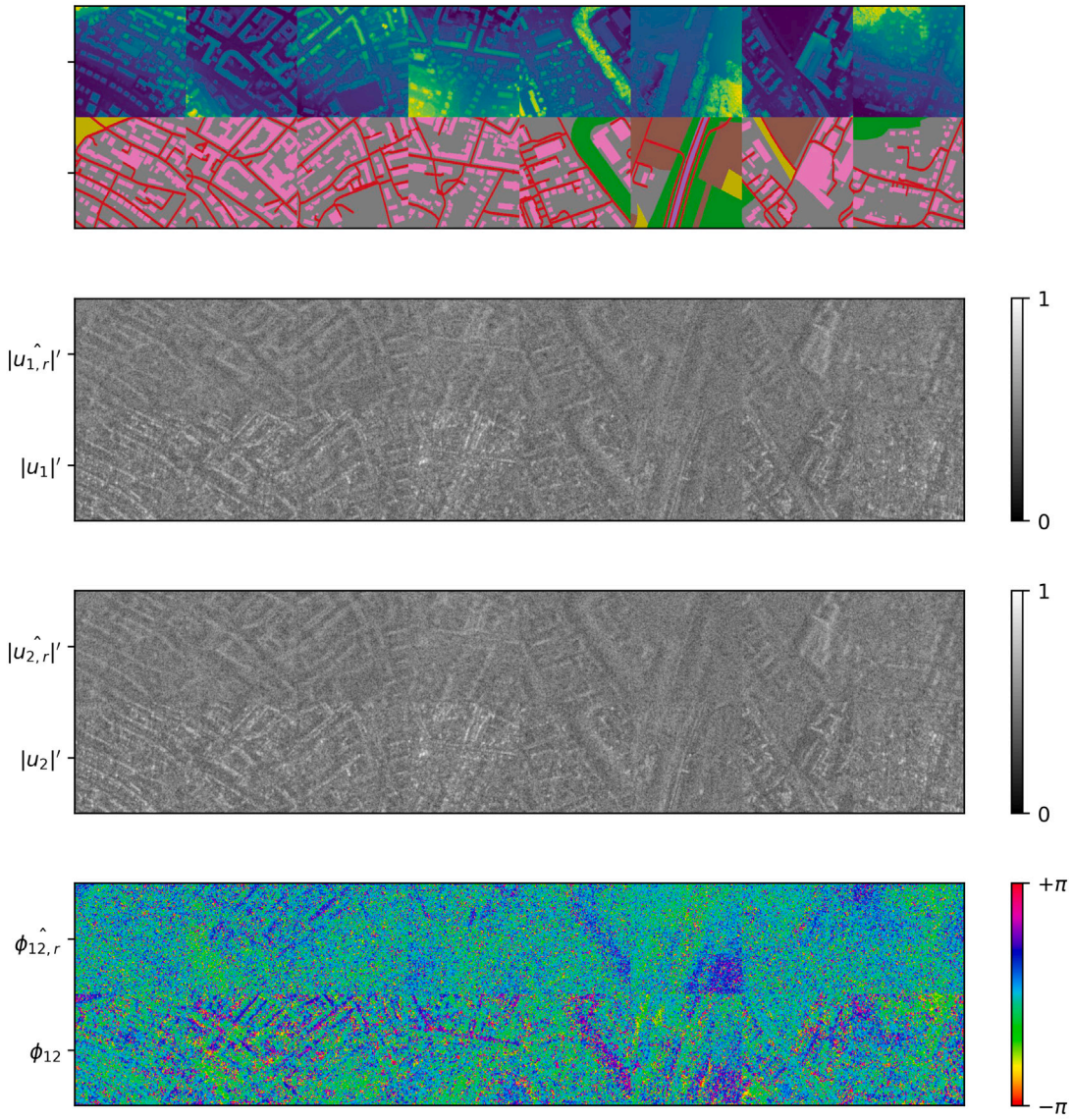

Fig. 8. Renoised results of the multitask network for the GeoNRW TDX dataset, decomposed to the interferogram SLC magnitudes and their interferometric phase, training for 150 + 20 epochs, $\mathcal{L}_{MT}$ including $\mathcal{L}_{\phi_{DEM}}$.

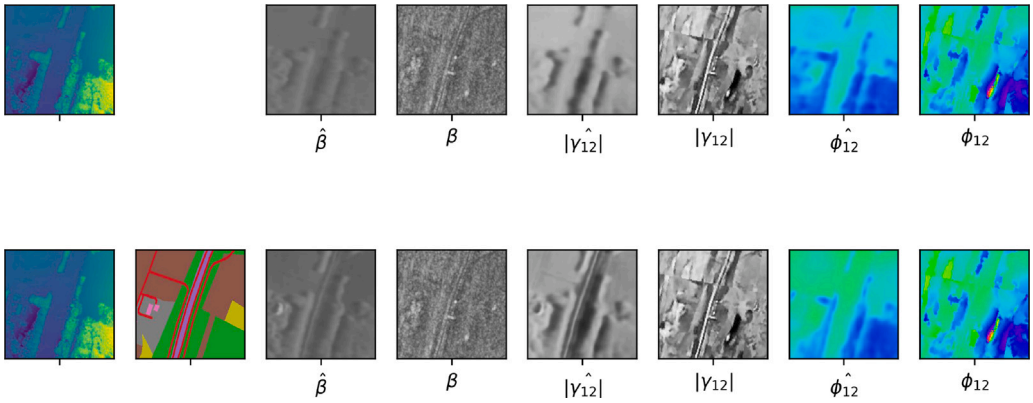

Fig. 9. Results of an additional ablation study for synthesis quality without (top row) and with (bottom row) land cover information. From left to right: DEM and land cover input information, the synthesized and real patches for reflectivity, coherence magnitude and interferometric phase, respectively.

7. Summary & conclusion

In this paper, we have addressed the challenge of synthesizing complex-valued synthetic aperture radar (SAR) data using a deep neural network. Our proposed approach overcomes the challenges of handling complex data by decomposing SAR interferograms into real-valued components that are simultaneously synthesized by a multi-branch encoder–decoder network architecture. The real-valued components are then combined to reconstruct the final complex-valued interferogram. We conducted experiments on medium-resolution repeat-pass C-band SAR data and high-resolution single-pass X-band SAR data and obtained promising results, demonstrating the general feasibility of our approach. We envision its application in various fields, including algorithm development, sensor evaluation, and understanding of SAR data characteristics. Further research can explore optimization strategies and extend the framework to additional SAR data modalities, allowing for more comprehensive simulations. By decomposing the interferogram into physically interpretable real-valued components, our framework is well suited for future use in physics-informed modeling for network optimization and design.

CRediT authorship contribution statement

Philipp Sibler: Writing – original draft, Visualization, Validation, Software, Methodology, Investigation, Data curation. **Francescopaolo Sica:** Writing – review & editing, Writing – original draft, Supervision, Methodology, Investigation. **Michael Schmitt:** Writing – review & editing, Supervision, Resources, Project administration, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Abady, L., Horváth, J., Tondi, B., Delp, E.J., Barni, M., 2022. Manipulation and generation of synthetic satellite images using deep learning models. J. Appl. Remote Sens. 16 (04), 046504. http://dx.doi.org/10.1117/1.JRS.16.046504.

Addabbo, P., Bernardi, M.L., Biondi, F., Cimitile, M., Clemente, C., Fiscante, N., Giunta, G., Orlando, D., Yan, L., 2023. Super-resolution of synthetic aperture radar complex data by deep-learning. IEEE Access 11, 23647–23658. http://dx.doi.org/10.1109/ACCESS.2023.3251565.

Asiyabi, R.M., Datcu, M., Anghel, A., Nies, H., 2023. Complex-valued end-to-end deep network with coherency preservation for complex-valued SAR data reconstruction and classification. IEEE Trans. Geosci. Remote Sens. 61, 5206417. http://dx.doi.org/10.1109/TGRS.2023.3267185.

Baier, G., Deschemps, A., Schmitt, M., Yokoya, N., 2020. GeoNRW. http://dx.doi.org/10.21227/s5xq-b822.

Baier, G., Deschemps, A., Schmitt, M., Yokoya, N., 2021. Synthesizing optical and SAR imagery from land cover maps and auxiliary raster data. IEEE Trans. Geosci. Remote Sens. 60, 4701312. http://dx.doi.org/10.1109/TGRS.2021.3068532.

Bamler, R., Hartl, P., 1998. Synthetic aperture radar interferometry. Inverse Problems 14 (4), R1–R54. http://dx.doi.org/10.1088/0266-5611/14/4/001.

Brosch, T., Neumann, C., 2021. Automatic target recognition on high resolution SAR images with deep learning domain adaptation. In: 2021 21st International Radar Symposium. IRS, IEEE, pp. 1–6. http://dx.doi.org/10.23919/IRS51887.2021.9466210.

Cao, C., Cao, Z., Cui, Z., 2020. LDGAN: A synthetic aperture radar image generation method for automatic target recognition. IEEE Trans. Geosci. Remote Sens. 58 (5), 3495–3508. http://dx.doi.org/10.1109/TGRS.2019.2957453.

Deledalle, C.-A., Denis, L., Tupin, F., 2011. NL-InSAR: Nonlocal interferogram estimation. IEEE Trans. Geosci. Remote Sens. 49 (4), 1441–1452. http://dx.doi.org/10.1109/tgrs.2010.2076376.

Fisher, N.I., 1995. Statistical Analysis of Circular Data. Cambridge University Press.

Fornaro, G., Pascazio, V., 2014. SAR interferometry and tomography: Theory and applications. In: Academic Press Library in Signal Processing. Elsevier, pp. 1043–1117. http://dx.doi.org/10.1016/b978-0-12-396500-4.00020-x.

Fuentes Reyes, M., Auer, S., Merkle, N., Henry, C., Schmitt, M., 2019. SAR-to-optical image translation based on conditional generative adversarial networks—Optimization, opportunities and limits. Remote Sens. 11 (17), 2067. http://dx.doi.org/10.3390/rs11172067.

Goodfellow, I.J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial networks. pp. 1–9. http://dx.doi.org/10.1145/3422622, arXiv:1406.2661.

Goodman, J.W., 1975. Statistical properties of laser speckle patterns. In: Dainty, J.C. (Ed.), Laser Speckle and Related Phenomena. no. 9, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 9–75. http://dx.doi.org/10.1007/bfb0111436.

Guo, J., Lei, B., Ding, C., Zhang, Y., 2017. Synthetic aperture radar image synthesis by using generative adversarial nets. IEEE Geosci. Remote Sens. Lett. 14 (7), 1111–1115. http://dx.doi.org/10.1109/LGRS.2017.2699196.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Identity mappings in deep residual networks. In: Lecture Notes in Computer Science. Springer International Publishing, pp. 630–645. http://dx.doi.org/10.1007/978-3-319-46493-0_38.

Hirose, A., 2012. Complex-Valued Neural Networks. In: Studies in Computational Intelligence, vol. 400, Springer Berlin Heidelberg, Berlin, Heidelberg, http://dx.doi.org/10.1007/978-3-642-27632-3.

Howe, J., Pula, K., Reite, A.A., 2019. Conditional generative adversarial networks for data augmentation and adaptation in remotely sensed imagery. In: Zelinski, M.E., Taha, T.M., Howe, J., Awwal, A.A., Iftekharuddin, K.M. (Eds.), Applications of Machine Learning. SPIE, p. 111390G. http://dx.doi.org/10.1117/12.2529586.

Hughes, L.H., Schmitt, M., Zhu, X.X., 2018. Mining hard negative samples for SAR-optical image matching using generative adversarial networks. Remote Sens. 10 (10), 1552. http://dx.doi.org/10.3390/rs10101552.

Ioffe, S., Szegedy, C., 2015. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: Proc. of 32nd International Conference on International Conference on Machine Learning. 37, pp. 448 – 456.

Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A., 2016. Image-to-image translation with conditional adversarial networks. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition. pp. 5967–5976. http://dx.doi.org/10.1109/CVPR.2017.632.

Merkle, N., Auer, S., Mueller, R., Reinartz, P., 2018. Exploring the potential of conditional adversarial networks for optical and SAR image matching. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (6), 1811–1820. http://dx.doi.org/10.1109/JSTARS.2018.2803212.

Mou, L., Schmitt, M., Wang, Y., Zhu, X.X., 2017. A CNN for the identification of corresponding patches in SAR and optical imagery of urban scenes. In: 2017 Joint Urban Remote Sensing Event. JURSE, IEEE, http://dx.doi.org/10.1109/JURSE.2017.7924548.

Odena, A., Dumoulin, V., Olah, C., 2016. Deconvolution and checkerboard artifacts. Distill http://dx.doi.org/10.23915/distill.00003.

Park, T., Liu, M.Y., Wang, T.C., Zhu, J.Y., 2019. Semantic image synthesis with spatially-adaptive normalization. In: Proc. of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2332–2341. http://dx.doi.org/10.1109/CVPR.2019.00244.

Shi, X., Zhou, F., Yang, S., Zhang, Z., Su, T., 2019. Automatic target recognition for synthetic aperture radar images based on super-resolution generative adversarial network and deep convolutional neural network. Remote Sens. 11 (2), 135. http://dx.doi.org/10.3390/rs11020135.

Sibler, P., Wang, Y., Auer, S., Ali, S.M., Zhu, X.X., 2021. Generative adversarial networks for synthesizing InSAR patches. In: EUSAR 2021; 13th European Conference on Synthetic Aperture Radar. pp. 1–6.

Sica, F., Cozzolino, D., Zhu, X.X., Verdoliva, L., Poggi, G., 2018. InSAR-BM3D: A nonlocal filter for SAR interferometric phase restoration. IEEE Trans. Geosci. Remote Sens. 56 (6), 3456–3467. http://dx.doi.org/10.1109/tgrs.2018.2800087.

Sica, F., Gobbi, G., Rizzoli, P., Bruzzone, L., 2021. Φ-Net: Deep residual learning for InSAR parameters estimation. IEEE Trans. Geosci. Remote Sens. 59 (5), 3917–3941. http://dx.doi.org/10.1109/TGRS.2020.3020427.

Song, Q., Xu, F., Zhu, X.X., Jin, Y.-Q., 2022. Learning to generate SAR images with adversarial autoencoder. IEEE Trans. Geosci. Remote Sens. 60, 5210015. http://dx.doi.org/10.1109/TGRS.2021.3086817.

Wang, T.-C., Liu, M.-Y., Zhu, J.-Y., Tao, A., Kautz, J., Catanzaro, B., 2018. High-resolution image synthesis and semantic manipulation with conditional GANs. In: Proc. of IEEE Conference on Computer Vision and Pattern Recognition. IEEE, pp. 8798–8807. http://dx.doi.org/10.1109/cvpr.2018.00917.

Zanaga, D., Van De Kerchove, R., De Keersmaecker, W., Souverijns, N., Brockmann, C., Quast, R., Wevers, J., Grosu, A., Paccini, A., Vergnaud, S., Cartus, O., Santoro, M., Fritz, S., Georgieva, I., Lesiv, M., Carter, S., Herold, M., Li, L., Tsendbazar, N.-E., Ramoino, F., Arino, O., 2021. ESA WorldCover 10 m 2020 v100. http://dx.doi.org/10.5281/zenodo.5571936.

Zar, J.H., 2014. Biostatistical Analysis, fifth ed. Pearson Education Limited.