IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 13 January 2023, accepted 24 February 2023, date of publication 2 March 2023, date of current version 13 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3251565

RESEARCH ARTICLE

Super-Resolution of Synthetic Aperture Radar Complex Data by Deep-Learning

PIA ADDABBO[1], (Senior Member, IEEE), MARIO LUCA BERNARDI[2], (Member, IEEE), FILIPPO BIONDI[3], (Member, IEEE), MARTA CIMITILE[4], (Member, IEEE), CARMINE CLEMENTE[5], (Senior Member, IEEE), NICOMINO FISCANTE[6], (Student Member, IEEE), GAETANO GIUNTA[6], (Senior Member, IEEE), DANILO ORLANDO[7], (Senior Member, IEEE), AND LINJIE YAN[8]

[1]Università degli Studi "Giustino Fortunato", 82100 Benevento, Italy
[2]Department of Engineering, Università degli Studi del Sannio, 82100 Benevento, Italy
[3]Italian Ministry of Defense, 00187 Rome, Italy
[4]Department of Law and Economics, Unitelma Sapienza University, 00161 Rome, Italy
[5]Department of Electronic and Electrical Engineering, University of Strathclyde, G1 1XW Glasgow, U.K.
[6]Industrial, Electronic and Mechanical Engineering Department, Università degli Studi Roma Tre, 00146 Rome, Italy
[7]Università degli Studi "Niccolò Cusano," 00166 Rome, Italy
[8]Institute of Acoustics, Chinese Academy of Sciences, Beijing 100045, China

Corresponding author: Gaetano Giunta (gaetano.giunta@uniroma3.it)

ABSTRACT One of the greatest limitations of Synthetic Aperture Radar imagery is the capability to obtain an arbitrarily high spatial resolution. Indeed, despite optical sensors, this capability is not just limited by the sensor technology. Instead, improving the SAR spatial resolution requires large transmitted bandwidth and relatively long synthetic apertures that for regulatory and practical reasons are impossible to be met. This issue gets particularly relevant when dealing with Stripmap mode acquisitions and with low carrier frequency sensors (where relatively large bandwidth signals are more difficult to be transmitted). To overcome this limitation, in this paper a deep learning based framework is proposed to enhance the spatial resolution of low-resolution SAR images while retaining the complex image accuracy. Results on simulated and real SAR data demonstrate the effectiveness of the proposed framework.

INDEX TERMS SAR, super-resolution, deep learning, CNN, COSMO-skymed.

I. INTRODUCTION

Synthetic Aperture Radar (SAR) imagery has become an important Earth observation technique to obtain a detailed insight of physical and shape characteristics of targets and scenes. As the technology and its applications grow up, there is a growing demand for SAR images with finer spatial and temporal resolutions. On one hand, time resolution can be addressed with a larger number of sensors deployed and, hence, able to provide more frequent observations, on the other hand, the spatial resolution challenge has intrinsic limitations. Indeed, in order to obtain a fine range resolution it is required to transmit relatively large bandwidths, which is not always possible because of frequency allocation or hardware constraints. As for the cross-range (or azimuth) resolution, it depends on the overall synthetic aperture length which cannot be made arbitrarily long [1]. It is thus necessary to identify data processing based solutions to solve the SAR spatial resolution problem with the aim to extract finer spatial details. In recent years a number of solutions have been proposed exploring sparsity, compressive sensing and deep learning (DL). As an example in [2], a recovery solution for SAR Single Look Complex (SLC) images exploits spectrum extrapolation to achieve super-resolution while mitigating noncoherent electromagnetic noise covering only the higher frequency spectrum. In the last decade,

The associate editor coordinating the review of this manuscript and approving it for publication was Gerardo Di Martino.

it becomes also popular to investigate compressive sensing to achieve super-resolution in SAR, such as in [3] where a super-resolution reconstruction method was proposed by combining compressive sensing with a multi-dictionary. The method performs well when a gradient-based optimization is used to learn the observation matrix. However, the performance assessment does not investigate the capability to preserve the phase information. Furthermore, this type of methods often suffers noisy environments and have off-grid issues.

Another interesting family of approaches is based on DL. Actually, such techniques find their application in a very wide field; just to name a few, DL has been succesfully apllied to modelling of pedestrian movements [4], [5] or gait recognition [6], [7]. Focusing on radar [8], [9], [10] and specifically on SAR [11], [12], [13], DL exploits large training datasets in order to infer patterns that can be then translated into filters, the latter can then be used to process the radar image with the aim to refine its resolution. Indeed, DL has been used to successfully reconstruct high resolution optical image from its low resolution sample. In [14], authors use a frequency domain based scheme to reconstruct the high-resolution image at various frequency bands. Further, authors propose a method that incorporates the wavelet transform and the recursive Res-Net. The wavelet transform is applied to the low-resolution image to divide it into various frequency components. To validate the effectiveness of the proposed method, extensive experiments are performed using the NWPU-RESISC45 data set, and the results demonstrate that the proposed method outperforms several state-of-the-art methods in terms of both objective evaluation and subjective perspective. Recently, in [15], the authors proposed a novel Convolutional Neural Network (CNN) based technique that exploits both spatial and temporal correlations to combine multiple images. This novel framework integrates the spatial registration task directly inside the CNN, relying on a single CNN with three main stages. Forced by the success on optical images, DL has been investigated also for SAR super-resolution. For example in [16], DL and dictionary are combined together with the result of providing good visual super resolution results. Another example is the framework proposed in [17], where the authors combine the advantages of multiple-image fusion with learning the low-to-high resolution mapping using deep neural networks. A more complex framework is proposed in [18], where a multi-component alternating direction method of multipliers is combined with DL in order to derive a processing framework to super-resolve SAR images. The proposed solution exploits a hybrid echo model to overcome the point target model limitations and formulate the SAR imaging as a constrained optimization problem. Such an approach could be very promising once the method is extended to off-grid scenarios and with un-paired data training. When available also polarimetric information can support the development of super-resolution methods, as in [19], where the authors propose a full-polarimetric SAR image super-resolution reconstruction method combining a CNN and residual compensation. The advantages of the deep CNN for nonlinear model fitting are exploited in order to perform super-resolution reconstruction on low-resolution full-polarimetric SAR images, and then a residual compensation is applied to the network reconstruction results. Finally, it is worth mentioning the approach introduced in [20], in which the authors propose a method based on DL to realize the reconstruction of SAR images while introducing the use of the Structural Similarity Index Measure (SSIM) index into the loss function, so that the reconstructed SAR image is improved both in subjective visual and in objective evaluation indicators.

In [21], an extension of the Deep CNN with residual net Skip Connection and Network (DCSCN) model described in [22] and [23] is proposed to address the SAR super resolution challenge. The proposed DC2SCN fully convolutional neural network extends the DCSCN model by accounting for 2 channels, for the real and the imaginary part of the SAR SLC, respectively. This paper completes [21] by providing the details of the DL-based framework devised to improve the spatial resolution. More importantly, we explain the motivations behind the proposed methodology and demonstrate it through an extensive experimental validation that comprises both synthetic as well as real-recorded data. Specifically, the proposed network extends the simpler DCSCN model, introduced in [22] and [23] for image super resolution, and is composed by a feature extraction stage accounting for 2 different channels, one for the real and another for the imaginary part of the SAR SLC image. This design choice is of primary importance in SAR applications since data are complex and besides amplitude information, also phase information is restored by the proposed network. The illustrative examples not only show that the method can provide an improved spatial resolution, but it also retains the phase information, a fundamental property if more advanced SAR analyses are required (e.g. interferometry).

The remainder of the paper is organized as follows: in Section II, the proposed DC2SCN framework is introduced, by formalizing the problem at hand and the proposed architecture. Section III describes the assessment criteria and quantitatively discusses the performance of the proposed framework on both real COSMO-Skymed and simulated SAR data.

II. THE PROPOSED METHODOLOGY

In this section, we describe the two DC2SCN channels as an extension of the DCSCN model described in [22] and [23]. The novelty of the proposed network lies in its two-channel architecture that allows us to handle complex data (and, hence, the phase information). In particular, some parts of the network are able to jointly learn the real and the imaginary part separately while other parts are joint and are able to learn the relationships between them leading to a much more precise internal representation model of the signal.

The overall network consists of a combination of a feature extraction and reconstruction tasks. The feature

extraction part of the network learns, from the complex SAR image, a wide range of features used by the subsequent reconstruction part to rebuild the input SAR image at a higher resolution.

The proposed DC2SCN variant, shown in Figure 1, exploits two main channels, instead of a single one, for both feature extraction and reconstruction of each SAR image component: one for real part and the other for imaginary part of the signal (recall that the processed SAR data are complex).

The feature extraction part is comprised, for each channel, of a cascade of N_F sets of 3×3 CNN, bias and Parametric Rectified Linear Units (ReLU). The original image is used as an input of the model allowing the network to efficiently grasp the features. Skip connections (as already proposed in [24] and [25]) are used to optimize the number of filters of each CNN layer and send extracted features at each level to the subsequent reconstruction part of the network.

Looking at the reconstruction part, in the case of data up-sampling, the transposed convolutional layer proposed in [26] is generally used. However, the transposed convolutional layer is characterized by a limited reconstruction ability since it can learn up-sampling kernels. To improve the reconstruction performance, it is necessary to introduce heavy computation. For this reason, in this study, the reconstruction part of the network is formed for each channel by a 1×1 CNN layer in parallel with two CNN blocks (a 1×1 CNN block and a 3×3 CNN block, each one followed by a bias and a parametric ReLU units) that are shared by the real and imaginary channels as shown at the center of the figure. This shared portion of the network allows

(*i*) for a decrease of the computational load by reducing the dimensions of the previous layers with a low information loss;

(*ii*) the entire network to learn relationships across real and imaginary parts of the signal thus preserving the phase information in the course of the high resolution image reconstruction.

For each channel an estimation of the up-sampled original image is obtained also resorting to bicubic interpolation. Notice that the input layer of the reconstruction network is characterized by large dimensions since all the features are included. For this reason a parallelized 1×1 CNN [27] is used to reduce the dimension before generating the final images ($\mathcal{H}_X$ and $\mathcal{H}_Y$). It also enhances the final representation by including an additional nonlinearity.

A. FORMALIZATION

Let us consider a low resolution SAR image composed of two input layers composed by the real and the imaginary part of $X^L + jY^L$, where j is the imaginary unit and the superscript "L" means low resolution. The first and only pre-processing performed on the original image is the upscaling to the desired higher resolution. Bicubic interpolation [28] is also a convolutional operation implemented as a convolutional layer. The difference in this case concerns the output size of this layer that is larger than the input one (the stride is fractional). To exploit the optimized implementations available in cuda CNN, this layer is implemented using a CNN layer but excluded from the learning process. Let be $X^u + jY^u$ such upscaled image (superscript "u" means upscaled resolution). The aim of the network is to recover an image $\mathcal{H}(X^u, Y^u; \Theta)$, with Θ, the network parameters, that is as faithful as possible to the ground-truth high resolution image, namely $X^H + jY^H$. In the following, $X^u + jY^u$ will be still called, for simplicity, low resolution SAR image even though has the same resolution as $X^H + jY^H$ (superscript "H" means high resolution). The network has to learn the $\mathcal{H}$ function; such a task is accomplished, as highlighted above, by cascading the sequence of feature extraction and reconstruction steps.

1) FEATURE EXTRACTION

Formally, the first CNN layer of each channel of the network is expressed as

$$F_{1X}(X^u, Y^u) = \max(0, W_{1X} \cdot X + B_{1X})$$
$$F_{1Y}(X^u, Y^u) = \max(0, W_{1Y} \cdot Y + B_{1Y})$$

where:

- W_{1X} and W_{1Y} represent the filters of the two channels,
- B_{1X} and B_{2X} are the corresponding biases.

More precisely, W_{1X} and W_{1Y} correspond both to the dimensions of the first filters (of size $f_1 \times f_1$); such parameters (including the biases B_{1X} and B_{2X}) are determined during the training process. The output is composed of o_1 feature maps and this also applies to biases whose elements are associated position-wise to each filter. As shown in Figure 1, we use Parametric ReLU as activation functions to the filter responses. This output is propagated in the same way through the subsequent layers and, at each layer, the output is concatenated with previous using skip connections, as shown in the figure; for example for the case of seven layers, we have $N_F = 7$. Obviously, it is possible to add multiple convolutional layers to increase the nonlinearity. However, this also increases the model size ($o_k \times f_k \times f_k \times o_k$ parameters for one layer) and thus requires higher time to train the model. In this experimentation, we used two networks of 20 and 40 total layers (hereafter called L20 and L40), respectively.

2) RECONSTRUCTION

In traditional reconstruction methods, the predicted overlapping high-resolution patches are averaged over the upscaled version to generate the high resolution final image. This process can be implemented as a filter applied on a set of generated feature maps (each region is a flattened vector representation of a high-resolution patch). For this reason, to build the final high-resolution image, the following convolutional layers are used, one per channel,

$$O_X(X^u, Y^u) = \max(0, W_{OX} \cdot X + B_{OX}) + F_e(X^u, Y^u)$$
$$O_Y(X^u, Y^u) = \max(0, W_{OY} \cdot Y + B_{OY}) + F_e(X^u, Y^u)$$

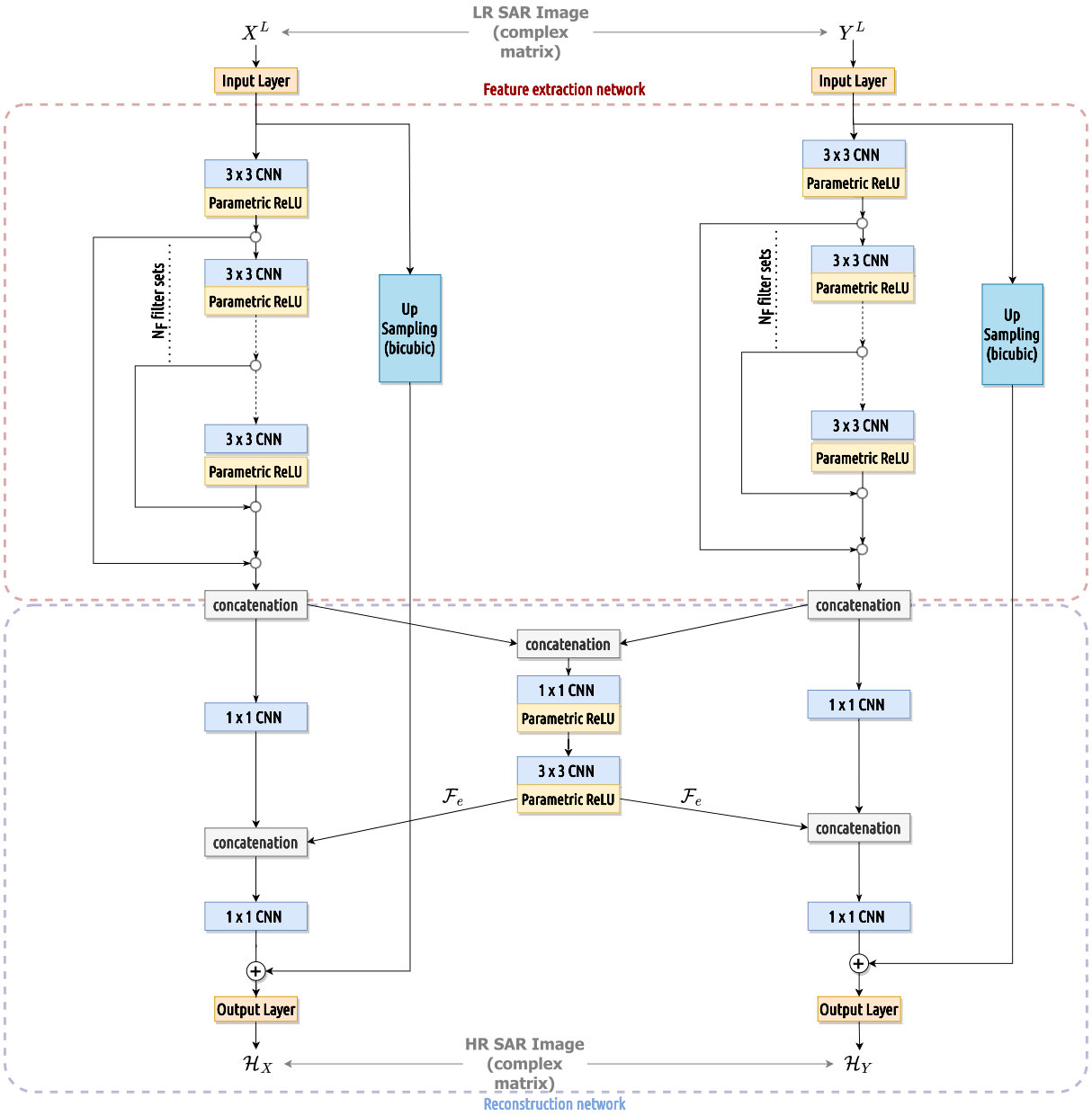


FIGURE 1. The proposed architecture.

where the subscript "O" means output and:

- F_e is the output of the part of the network shared across X and Y channels;
- W_{OX} and W_{OY} are the learnable filters of size $f_o \times f_o$;
- B_{OX} and B_{OY} are the learnable biases vectors.

B. LOSS FUNCTION AND TRAINING

To learn the end-to-end function $\mathcal{H}$, we need to estimate the network parameters $\Theta = \{(W_{iX}, W_{iY}) \mid i \in \{1, k\}\}$ where k is the number of layers. This can be performed by minimizing a loss function between the reconstructed images $\mathcal{H}(X^u, Y^u; \Theta)$ and the related ground truth high resolution images (i.e., $X^H + jY^H$). Since the convolutional neural networks do not hinder the adoption of other kinds of loss functions (if only the loss functions are derivable), given a set of n high-resolution images $\{X_h^u + jY_h^u\}$ and their corresponding low resolution images $\{X_h^L + jY_h^L\}$, we used a hybrid loss function based on the Mean Squared Error (MSE) and the structural SSIM [29] as defined in:

$$\mathcal{L}_{\text{hybrid}} = \frac{1}{n}\sum_{h=1}^{n} ||\mathcal{H}(X_h^u, Y_h^u; \Theta) - (X_h^H + jY_h^H)||^2 + \mathcal{L}_{\text{ssim}}$$

where n is the number of images exploited for training. The part of loss using MSE favors higher Peak Signal-to-Noise-Ratio (PSNR) that is a widely-used metric to evaluate image reconstruction quality from an objective point of

view. The second part of the loss is based on SSIM to assign higher weights to the fuzzy boundary. Specifically, two corresponding $N \times N$ patches are cropped from the reconstructed SAR image components R_X and R_Y and from the ground truth G_X and G_Y, which can be indicated, for each channel, with $r = \{r_j \,|j = 1, \ldots, N^2\}$ and $g = \{g_k \,|k = 1, \ldots, N^2\}$ respectively. The $\mathcal{L}_{\text{ms-ssim}}$ part of the overall loss for r and g can be expressed as:

$$\mathcal{L}_{\text{ssim}} = 1 - \left(\frac{2\mu_r\mu_g + K_1}{\mu_p^2 + \mu_g^2 + K_1}\right)\left(\frac{2\sigma_{rg} + K_2}{\sigma_r^2 + \sigma_g^2 + K_2}\right)$$

where μ_r, μ_g, σ_r, σ_g are the mean, respectively, and the standard deviation of r and g, and σ_{rg} is their covariance. Finally, K_1 and K_2 are small constants selected to avoid dividing by zero in some circumstances (in our experimentation $K_1 = 0.01^2$ and $K_2 = 0.03^2$).

III. EXPERIMENTS AND RESULTS

In what follows, we present some numerical examples to illustrate the behavior of the proposed deep network architecture.

All the single classifiers have been developed using PyTorch,[1] Tensorflow,[2] and Keras,[3] which are three open-source neural network libraries with Python bindings (used to implement the classifiers).

For this experimentation the following two workstations have been used:

- AMD Ryzen Threadripper 3960X 24-Core, with 128GB of RAM and two GPU NVIDIA RTX 3090 (with 24GB of RAM);
- Intel Core i9 9940X (14 cores), with 64GB of RAM and four GPU NVIDIA Tesla T4 (with 16GB of RAM).

To quantitatively evaluate the performance in terms of the quality of the reconstructed super-resoluted image, hereafter indicated as $\mathbf{I_r} \in \mathbb{C}^{M\times N}$, with respect to the original full-resoluted one, namely $\mathbf{I_f} \in \mathbb{C}^{M\times N}$, we use the following metrics:

- the Mean Absolute Error (MAE) expressed as

$$\text{MAE} = \frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}[|\mathbf{I_f}(m,n)| - |\mathbf{I_r}(m,n)|];$$

- the Root Mean Square Error (RMSE) expressed as

$$\text{RMSE} = \sqrt{\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}[|\mathbf{I_f}(m,n)| - |\mathbf{I_r}(m,n)|]^2};$$

- the PSNR, expressed in dB, defined as the ratio of the maximum pixel intensity to the power of the distortion,

$$\text{PSNR} = 10\log_{10}\frac{\max(|\mathbf{I_r}|^2)}{\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}[|\mathbf{I_f}(m,n)| - |\mathbf{I_r}(m,n)|]^2};$$

- the SSIM, a widely used perceptual image quality metric, that can be expressed as

$$\text{SSIM} = l \cdot c \cdot s,$$

with l, c, s the luminance, the contrast and the structural changes as defined in [29].

Moreover, to evaluate the ability to identify a weak target from a nearby strong one, the Peak Side Lobe Ratio (PSLR), considering an azimuth or a range section of the image, expressed in dB, is calculated as

$$\text{PSLR} = 10\log_{10}\frac{m_0}{m_1},$$

with m_0 the peak intensity of the greatest sidelobe and m_1 the peak intensity of the mainlobe.

Another metric of interest for isolated bright point targets is the spatial resolution measured as the distance between the points with intensities 3 dB below the maximum intensity of the main lobe peak in the azimuth and/or range sections [30].

A real recorded dataset is used to train the networks and is discussed in detail in Subsection III-A. The first part of the analysis (Subsection III-B) is aimed at showing the super resolution behavior of the proposed architecture by using synthetic data adhering to the nominal design assumptions, whereas, in the second part (Subsections III-C and III-D) the performance are assessed exploiting real SAR data.

A. REAL DATASET

The data used in the experiments are SAR SLC images sensed by COSMO-SkyMed (CSK) in the StripMap (SM) acquisition mode [31]. The SM mode implements approximately a spatial resolution of 3×3 m in ground coordinates. The dataset that we take into account is made by 10 scenes of different sizes acquired over Northern Italy. Since each whole scene consists of wide geographical areas containing millions of pixels, the images are divided into blocks of size 512×512 pixels, hereafter referred to as tiles. Precisely, the whole dataset is divided into a total of 20239 tiles. Figure 5 shows the entire first scene that covers the river Po valley in Northern Italy. This scene covers an area of about 50×65 km and is made by 1344 tiles.

We trained the network employing degraded resolution SAR images. Spatial resolution degradation was done in both the range and azimuth directions. In this context, we focused the images by employing adapted filters that consider half-band, for both range and azimuth. The same images, that we focused at half spatial resolution, were also focused at maximum resolution, namely using all the available chirp and Doppler bandwidth. The maximum resolution images were used as ground truth during the training phase of the neural network.

Precisely, we selected 16191 tiles for training and validation and 4048 tiles for test. This experimental setup allowed for an extensive training of the network as well as a computational efficiency.

[1] https://pytorch.org/
[2] https://www.tensorflow.org/
[3] https://keras.io/

TABLE 1. Hyper-parameters optimized during assessment.

Hyper-paramenter	Values
Number of layers	[20, 40]
Learning rate	[0.0001, 0.001, 0.01, 0.1, 0.5{]
Dropout	[0.15, 0.20, 0.25]

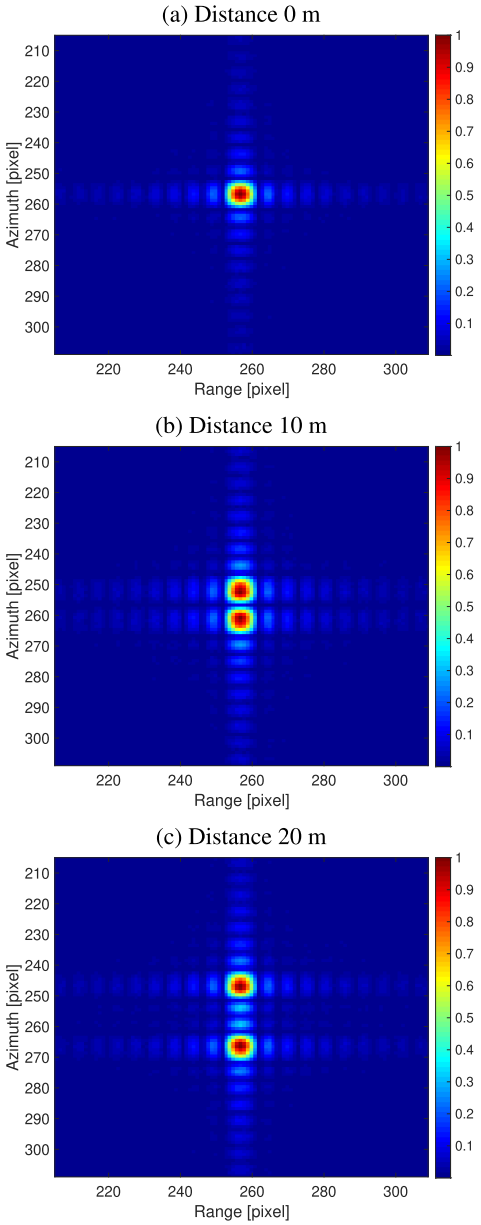


FIGURE 2. Simulated targets images with −25 dB noise power (original resolution).

Finally during the training process we optimize the most relevant hyper-parameters of the network reported in Table 1. Specifically, for each network of a given number of layers (20 and 40), we find the values of dropout and learning rate leading to the best performing model.

B. PERFORMANCE EVALUATION ON SIMULATED DATA

The simulation setup is designed to match the same characteristics of the CSK SAR sensor. We simulate a tile

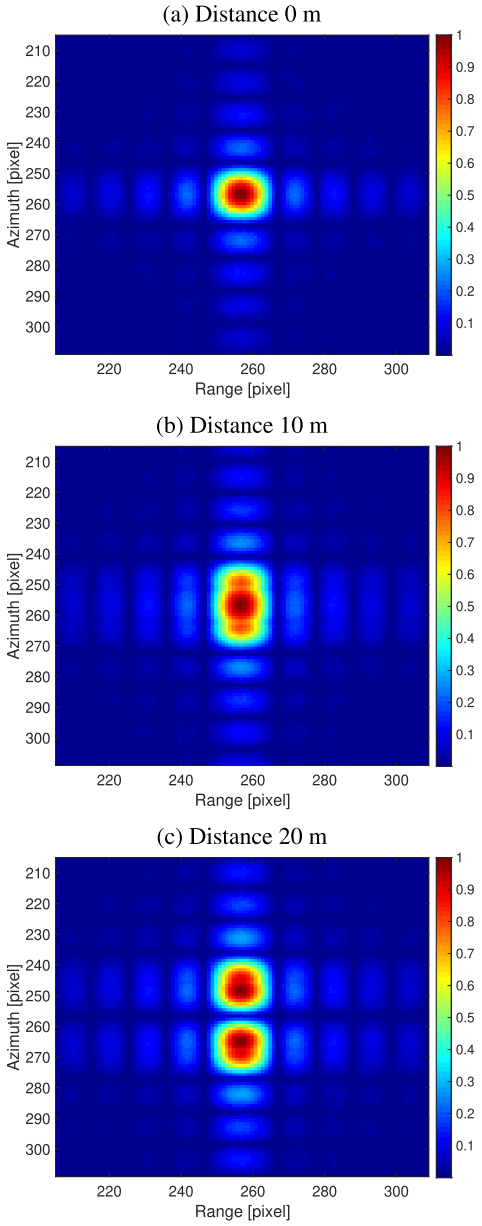


FIGURE 3. Simulated targets images with −25 dB noise power (degraded resolution).

with two sinc spread functions point targets with equal power intensity and a gaussian noise is added.

Three different cases are considered for the simulation with different distances between the two targets, namely, 0 m, 10 m and 20 m, at the three instantaneous PSNR levels of 5 dB, 15 dB and 25 dB. In Figures 2 and 3, the simulated normalized images with −25 dB of noise power and at the three distances in the subplots are shown, with the original and the degraded resolutions, respectively. As expected, at 0 m a single bright target is visible. At 10 m, it is interesting to notice that only with the full resoluted image, it is possible to discriminate the two targets. At 20 m with both the original and the degraded resolutions, the two targets are clearly separated.

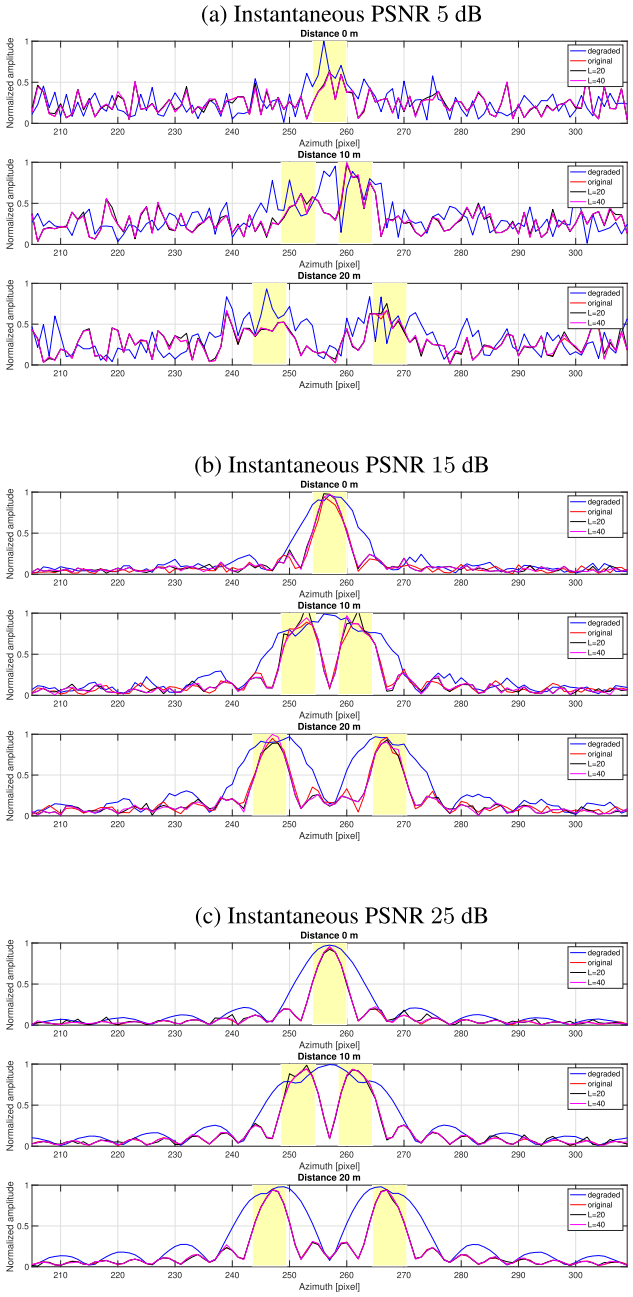


FIGURE 4. Simulated azimuth targets' sections.

In Figure 4, the central azimuth sections are shown for instantaneous PSNR, defined as the peak signal power over the noise power, equal to 5 dB, 15 dB and 25 dB at the top, center and at the bottom, respectively. The original full-resoluted, the degraded, and the reconstructed with L20 and L40 sections are plotted with red, blue, black, and violet lines, respectively. The yellow box, identify the peaks locations, and in the case where the distance is 0 m, the peaks are overlapped.

The results show that the L20 and L40 configurations are able to reconstruct the original resolution as well as to discriminate the two targets for both 10 and 20 m of distances.

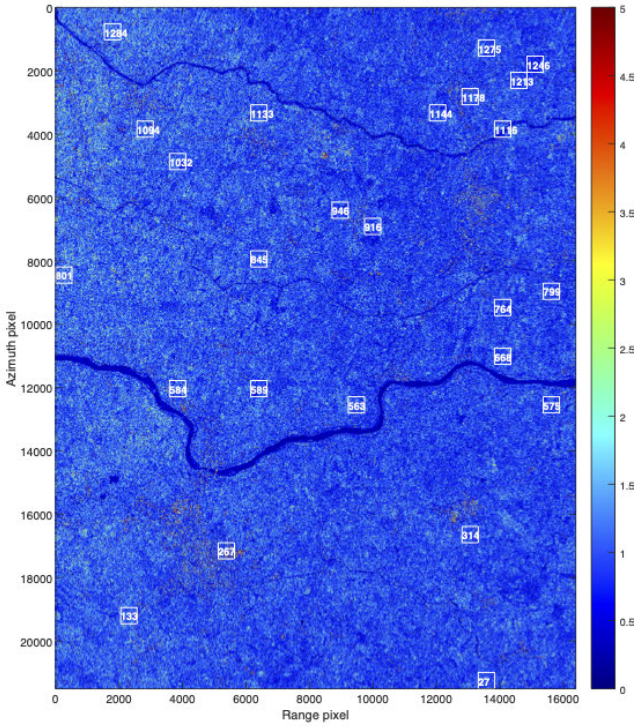


FIGURE 5. Full scene: white boxes indicate the 25 tiles analyzed in Table 2. The intensity image is represented with a zoomed color scale from 0 to 5 for image enhancement.

C. PERFORMANCE EVALUATION ON REAL DATA

The 25 tiles highlighted in the boxes in Figure 5 are here considered for assessing the performance of the proposed architecture on real recorded data. Particularly, in Table 2, the PSNR [dB], the RMSE, the MAE and the SSIM are evaluated using both L20 and L40 configurations which are compared with the corresponding degraded images used as input. This is done to evaluate the gain obtained through the reconstructed resolution of the two networks. The Root Mean Squared (RMS) values of these metrics are also computed and reported as the last row of the table.

The experimental results show an increase in the values of PSNR and SSMI for all the tiles and for both L20 and L40 configurations. The RMS PSNR gains of 13.97 dB and 12.54 dB with respect to degraded image are obtained by L40 and L20, respectively. As for the SSIM, the RMS value for the degraded image is 0.17 whereas 0.60 and 0.67 are obtained with L40 and L20, respectively. Both the RMSE and MAE metrics also confirm the superiority of L40 with respect to L20 with slightly lower values.

In Figure 6, the tile 267 of the full scene in Figure 5 is shown representing a mixed urban and vegetated area. Particularly, from the top sub-image to the bottom one, the original full-resoluted image, the degraded image and the reconstructed images with L20 and L40 are shown. From visual inspection, it is possible to notice a smoothing effect on the degraded resoluted images whereas for both the reconstructed resolution image finer details of the urban area are more visible.

TABLE 2. PSNR [dB], RMSE, MAE, SSIM for real data: degraded, L20 and L40.

Tile	Degraded				Reconstructed L20				Reconstructed L40			
NUMBER	PSNR	RMSE	MAE	SSIM	PSNR	RMSE	MAE	SSIM	PSNR	RMSE	MAE	SSIM
27	29.6079	0.7125	1.0826	0.1701	39.7836	0.5307	0.6579	0.5825	41.4470	0.4500	0.5457	0.6598
133	25.8905	0.8332	1.1953	0.1681	39.6861	0.5288	0.6575	0.6319	40.9619	0.4484	0.5460	0.7072
267	28.6224	2.2048	1.9856	0.2558	49.9019	0.5317	0.6578	0.7387	51.3356	0.4509	0.5469	0.7937
314	24.6694	0.7103	1.0778	0.1738	35.8558	0.5313	0.6587	0.5802	37.0988	0.4497	0.5472	0.6602
563	19.3600	0.7061	1.0887	0.1631	33.2897	0.5275	0.6582	0.5866	34.7124	0.4498	0.5467	0.6643
575	31.0196	0.6500	1.0140	0.1595	33.4646	0.5298	0.6582	0.5496	34.9106	0.4488	0.5457	0.6334
584	24.2356	0.8379	1.2567	0.1631	36.8599	0.5295	0.6601	0.6582	38.2199	0.4486	0.5463	0.7301
589	24.2564	0.6628	1.0168	0.1637	35.8477	0.5322	0.6587	0.5462	37.2345	0.4502	0.5469	0.6299
668	20.0254	0.6842	1.0635	0.1632	30.6031	0.5293	0.6572	0.5765	32.0997	0.4500	0.5470	0.6556
764	16.5774	0.6509	1.0175	0.1678	26.3496	0.5301	0.6580	0.5594	28.0454	0.4496	0.5467	0.6432
799	24.4627	0.6728	1.0318	0.1654	33.8356	0.5296	0.6577	0.5622	35.6740	0.4503	0.5466	0.6433
801	30.1402	0.8700	1.2963	0.1505	45.2202	0.5266	0.6583	0.6736	46.6197	0.4477	0.5454	0.7426
845	16.9878	0.7048	1.1156	0.1632	30.0363	0.5280	0.6574	0.6011	32.0957	0.4484	0.5464	0.6784
916	27.6395	0.7746	1.0927	0.1708	41.6875	0.5295	0.6570	0.5830	43.1880	0.4502	0.5459	0.6623
946	26.0363	0.6914	1.0205	0.1674	39.5531	0.5305	0.6578	0.5494	41.1147	0.4502	0.5462	0.6295
1032	20.8259	0.7107	1.0752	0.1591	31.0058	0.5329	0.6584	0.5685	32.4057	0.4529	0.5470	0.6437
1094	28.7567	0.9027	1.3007	0.1694	44.0202	0.5291	0.6578	0.6725	45.3567	0.4488	0.5464	0.7410
1116	23.9181	0.7457	1.0322	0.1764	37.3402	0.5376	0.6590	0.5390	38.6867	0.4541	0.5454	0.6128
1133	30.0263	1.0219	1.2104	0.1771	48.0891	0.5299	0.6585	0.6312	49.5094	0.4491	0.5458	0.7047
1144	20.9290	0.6877	1.0697	0.1568	30.9397	0.5290	0.6569	0.5728	32.1756	0.4481	0.5449	0.6528
1178	36.8083	0.9967	1.2190	0.1836	50.9925	0.5291	0.6582	0.6293	52.3661	0.4499	0.5471	0.7027
1213	24.3621	0.6744	1.0396	0.1595	36.7461	0.5287	0.6580	0.5624	38.3441	0.4483	0.5457	0.6458
1246	16.0980	0.5163	0.8082	0.1642	26.7157	0.5348	0.6575	0.4280	27.6323	0.4531	0.5466	0.5177
1275	16.8501	0.6959	1.0871	0.1681	28.3132	0.5285	0.6579	0.5842	29.6994	0.4470	0.5454	0.6664
1284	18.7144	0.8759	1.3426	0.1510	32.4554	0.5280	0.6598	0.6897	34.1152	0.4471	0.5465	0.7584
RMS value	**24.83**	**0.86**	**1.16**	**0.17**	**37.37**	**0.53**	**0.66**	**0.60**	**38.80**	**0.45**	**0.55**	**0.67**

In Figure 7, the histograms of the phase errors for the tile 267 are shown. From figure's inspection, it is clear that the super-resolution of the image produces a phase error correction with respect to the degraded image, for both L20 and L40 with most of the errors concentrated between 0 and $\pi/8$ (near 56% and 63% for L20 and L40, respectively).

Moreover, a zoom plot on a bright target of the degraded image is shown at the top in Figure 8. At the center of the same figure, the interpolated azimuth sections for the four images are shown. In this case, the PSLR values are -12.9809 dB, -5.9038 dB, -13.0155 dB and -13.0461 dB for the original full-resoluted image, the degraded image and the reconstructed images with L20 and L40, respectively. In terms of 3 dB main lobe amplitude, 40 interpolated azimuth pixels (corresponding to 4 original azimuth pixels) for the degraded are found whereas 20 pixels are obtained for both L20 and L40 which coincide with the full resolution of the original image. At the bottom of Figure 8, the phase values are reported. Interestingly, it can be noticed that the reconstructed phases show a very good agreement with respect to the original one. On the contrary, the degraded phase shows a phase step of π in the correspondence of the main lobe position (azimuth pixel 234).

These results confirm that the proposed architecture is able to super-resolve as well as to reconstruct both the module and the phase of the SAR complex data.

D. INTERFEROMETRIC ANALYSIS ON REAL DATA

In this subsection, the real data experimental analysis is enriched to demonstrate the robustness of the proposed architecture using a different real SAR dataset. The selected scenario used, here, is composed by a mixed urban/desertic area over Dubai. More precisely, a couple of CSK Spotlight (SP) images acquired over the same area using two different temporal observations and, at fixed spatial baseline, is used to demonstrate the capability of the interferometric phase's preservation.

The training network, used here, is the same as the previous experiments. Please notice that the data type (SP in place of SM) from CSK used for testing here is different and also the scenario is different (Po valley in place of Dubai). This is done to prove the robustness of the architecture to data type changes.

The interferogram using the half spatial resolution images (degraded as in the previous subsections and used as input for reconstruction) is reported in Figure 9(b), whereas the interferogram generated using the images reconstructed by the L40 network is represented in Figure 9(c). Figure 9(a) shows the phase interferogram generated using the original full resoluted images, reported for comparison purposes.

Finally, from visual inspection, it is possible to appreciate that, when the degraded images are used to generate interferogram, some fine details on the relative phase information are lost; on the contrary, the reconstructed L40 interferogram is very compatible with the original one.

In Figure 10, we report the phase line and phase gradient at range pixel 450 and with azimuth ranging between 1 and 350 pixels of the previous interferograms (original, degraded and reconstructed L40). From Figure 10(a), it is possible to observe that the reconstructed profile phase closely follows the slow phase changes. While using the degraded profile we would not be able to reconstruct the phase information, the reconstructed profile still contains the original information.

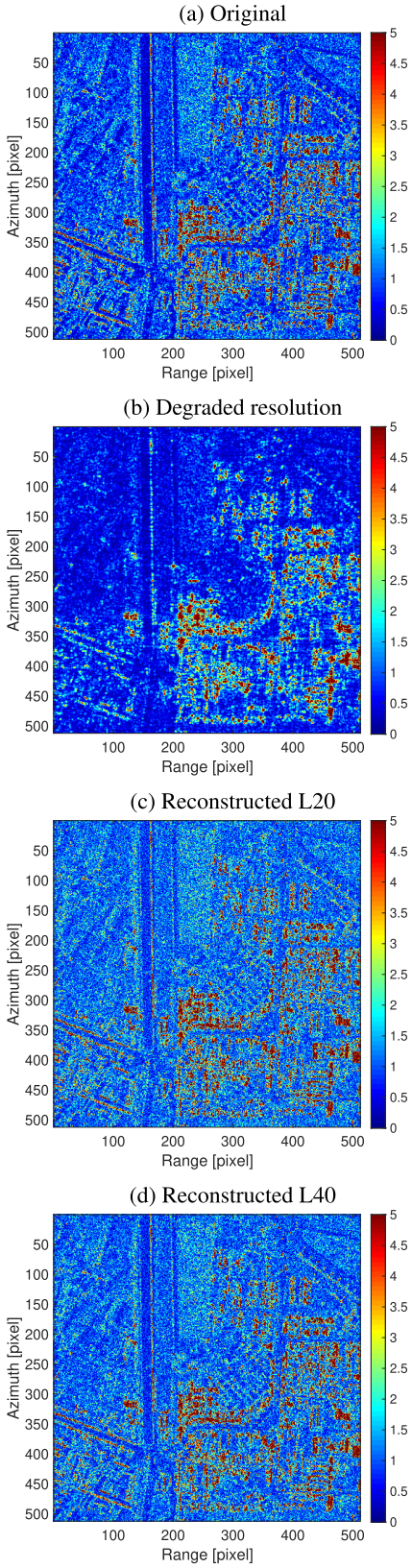


FIGURE 6. Tile 267 of the full scene in Figure 5.

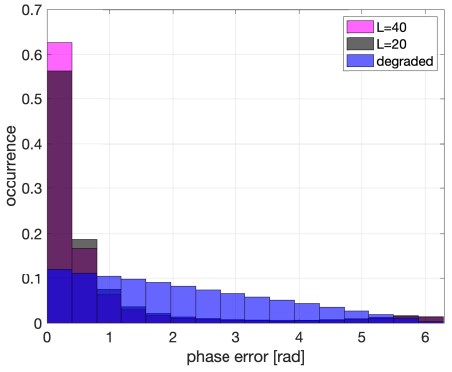


FIGURE 7. Histograms of the phase errors of tile 267.

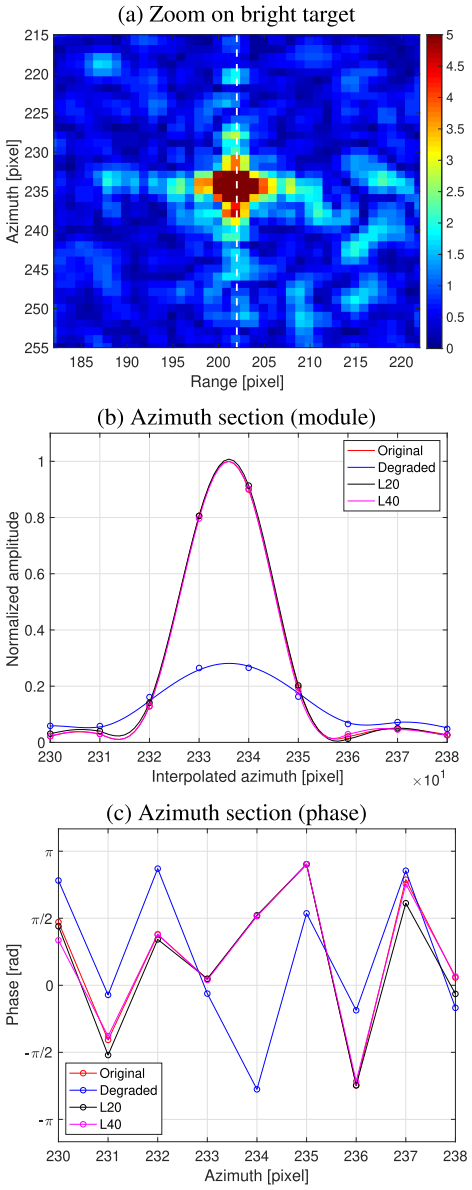


FIGURE 8. Bright target in tile 267.

This is quite evident by looking at Figure 10(b) where we show the phase gradient plots. It is clear that the degraded gradient phase profile (brown dashed line) exhibits a variation range greater than the reconstructed one (dotted black line).

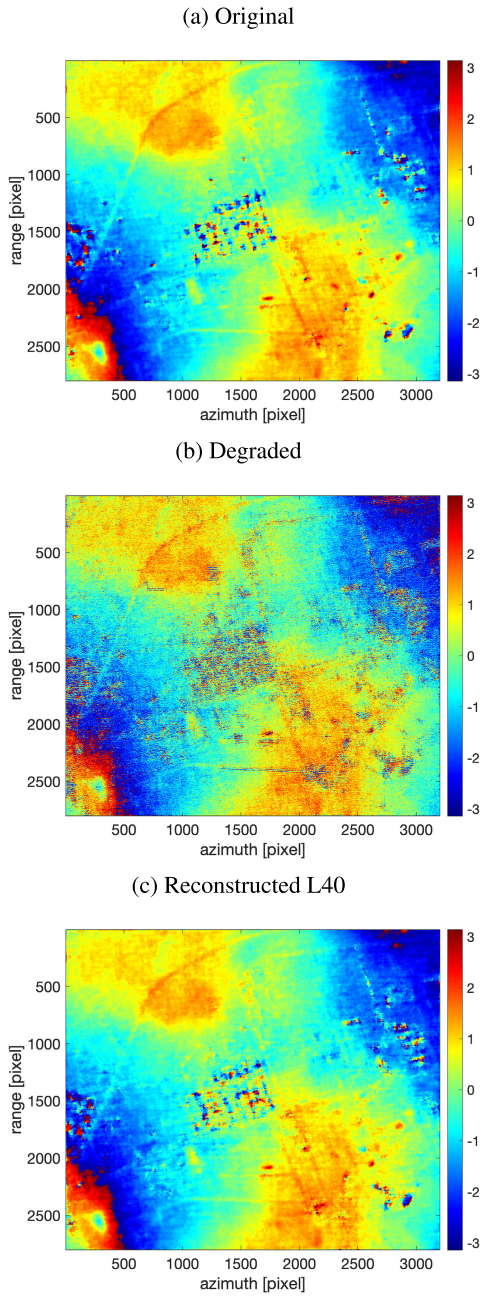


FIGURE 9. Phase interferograms using two different temporal observations and at fixed spatial baseline.

The RMSE values, computed over 350 samples, with respect the original phase, say ϕ_o, and defined as

$$\sqrt{\frac{1}{350}\sum_{n=1}^{350}[\phi_o(n)-\phi_k(n)]^2}, k \in \{d, s\}$$

for the degraded phase, ϕ_d, and the L40 reconstructed phase, ϕ_s, are 0.3383 and 0.0863, respectively. Thus, the improvement with respect to the degraded image is about one order of magnitude.

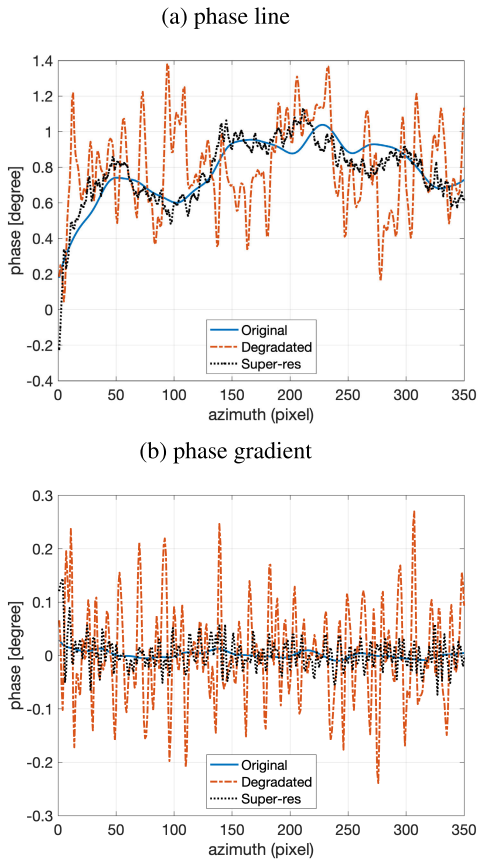


FIGURE 10. Phase line and phase gradient: range line 450, azimuth interval [1,350]. Phases are from the interferograms in Figure 9 (original phase in blue line, degraded phase in brown dashed line and reconstructed L40 phase in black dotted line).

IV. CONCLUSION

In this paper a novel convolutional neural network architecture, called DC2SCN, is introduced to reconstruct super-resolution SLC SAR images. It is important to highlight that, unlike existing works dealing with optical images or amplitude data, the proposed framework processes complex data, namely both real and imaginary parts, in order to reconstruct not only the amplitude but also the phase information that is of primary importance in advanced SAR applications such as interferometry. To this end, the proposed method employs two channels to deal with both real and imaginary parts of the SAR image and combines a feature extraction stage with a reconstruction stage in order to rebuild a super-resolved SAR image. At the analysis stage, both simulated and real data have been used to assess quantitatively the effectiveness of the proposed framework. In particular, on simulated data it has been shown how the framework is able to enhance image resolution in different noise conditions, while real COSMO-skymed SAR data have been used to extensively assess the performance of the network. Different figures of merit have been used to asses the performance over rural and anthropized areas, all confirming the capability of the proposed framework to

reliably create super-resolution images. Finally, the capability to preserve phase information has been also demonstrated, thus remarkably, the output of the proposed processing chain can be used for advanced SAR techniques.

REFERENCES

[1] I. Cumming and F. Wong, *Digital Processing of Synthetic Aperture Radar Data: Algorithms and Implementation*. Norwood, MA, USA: Artech House, 2005.

[2] F. Biondi, "Recovery of partially corrupted SAR images by super-resolution based on spectrum extrapolation," *IEEE Geosci. Remote Sens. Lett.*, vol. 14, no. 2, pp. 139–143, Feb. 2017.

[3] C. He, L. Liu, L. Xu, M. Liu, and M. Liao, "Learning based compressed sensing for SAR image super-resolution," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 5, no. 4, pp. 1272–1281, Aug. 2012.

[4] A. K.-F. Lui, Y.-H. Chan, and M.-F. Leung, "Modelling of pedestrian movements near an amenity in walkways of public buildings," in *Proc. 8th Int. Conf. Control, Autom. Robot. (ICCAR)*, Apr. 2022, pp. 394–400.

[5] A. K.-F. Lui, Y.-H. Chan, and M.-F. Leung, "Modelling of destinations for data-driven pedestrian trajectory prediction in public buildings," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2021, pp. 1709–1717.

[6] P. Addabbo, M. L. Bernardi, F. Biondi, M. Cimitile, C. Clemente, and D. Orlando, "Gait recognition using FMCW radar and temporal convolutional deep neural networks," in *Proc. IEEE 7th Int. Workshop Metrol. Aerosp. (MetroAeroSpace)*, Jun. 2020, pp. 171–175.

[7] P. Addabbo, M. L. Bernardi, F. Biondi, M. Cimitile, C. Clemente, and D. Orlando, "Temporal convolutional neural networks for radar micro-Doppler based gait recognition," *Sensors*, vol. 21, no. 2, p. 381, Jan. 2021.

[8] Z. Geng, H. Yan, J. Zhang, and D. Zhu, "Deep-learning for radar: A survey," *IEEE Access*, vol. 9, pp. 141800–141818, 2021.

[9] A. Davoli, G. Guerzoni, and G. M. Vitetta, "Machine learning and deep learning techniques for colocated MIMO radars: A tutorial overview," *IEEE Access*, vol. 9, pp. 33704–33755, 2021.

[10] J. Fuchs, M. Gardill, M. Lübke, A. Dubey, and F. Lurz, "A machine learning perspective on automotive radar direction of arrival estimation," *IEEE Access*, vol. 10, pp. 6775–6797, 2022.

[11] W. Zhang, Y. Zhu, and Q. Fu, "Semi-supervised deep transfer learning-based on adversarial feature learning for label limited SAR target recognition," *IEEE Access*, vol. 7, pp. 152412–152420, 2019.

[12] K. Wang, Q. Qiao, G. Zhang, and Y. Xu, "Few-shot SAR target recognition based on deep kernel learning," *IEEE Access*, vol. 10, pp. 89534–89544, 2022.

[13] A. Passah, S. N. Sur, B. Paul, and D. Kandar, "SAR image classification: A comprehensive study and analysis," *IEEE Access*, vol. 10, pp. 20385–20399, 2022.

[14] W. Ma, Z. Pan, J. Guo, and B. Lei, "Achieving super-resolution remote sensing images via the wavelet transform combined with the recursive res-net," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3512–3527, Jun. 2019.

[15] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "DeepSUM: Deep neural network for super-resolution of unregistered multitemporal images," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 5, pp. 3644–3656, May 2020.

[16] L. Zhao, Q. Sun, and Z. Zhang, "Single image super-resolution based on deep learning features and dictionary model," *IEEE Access*, vol. 5, pp. 17126–17135, 2017.

[17] M. Kawulok, P. Benecki, S. Piechaczek, K. Hrynczenko, D. Kostrzewa, and J. Nalepa, "Deep learning for multiple-image super-resolution," *IEEE Geosci. Remote Sens. Lett.*, vol. 17, no. 6, pp. 1062–1066, Jun. 2020.

[18] Y. Wei, Y. Li, Z. Ding, Y. Wang, T. Zeng, and T. Long, "SAR parametric super-resolution image reconstruction methods based on ADMM and deep neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 12, pp. 10197–10212, Dec. 2021.

[19] L. Lin, J. Li, Q. Yuan, and H. Shen, "Polarimetric SAR image super-resolution VIA deep convolutional neural network," in *Proc. IEEE Int. Geosci. Remote Sens. Symp. (IGARSS)*, Jul. 2019, pp. 3205–3208.

[20] Z. Luo, J. Yu, and Z. Liu, "The super-resolution reconstruction of SAR image based on the improved FSRCNN," *J. Eng.*, vol. 2019, no. 19, pp. 5975–5978, Oct. 2019.

[21] P. Addabbo, M. L. Bernardi, F. Biondi, M. Cimitile, C. Clemente, N. Fiscante, G. Giunta, and D. Orlando, "Super-resolution of synthetic aperture radar complex data by deep-learning," in *Proc. IEEE 9th Int. Workshop Metrol. Aerosp. (MetroAeroSpace)*, Jun. 2022, pp. 237–241.

[22] K. Peng and D. Ma, "Tree-structure CNN for automated theorem proving," in *Neural Information Processing*, D. Liu, S. Xie, Y. Li, D. Zhao, and E. M. El-Alfy, Eds. Springer, 2017, pp. 3–12. [Online]. Available: https://www.springerprofessional.de/tree-structure-cnn-for-automated-theorem-proving/15202012

[23] J. Yamanaka, S. Kuwashima, and T. Kurita, "Fast and accurate image super resolution by deep CNN with skip connection and network in network," in *Proc. Int. Conf. Neural Inf. Process. (ICONIP)*, 2017, pp. 217–225.

[24] L. Aversano, M. L. Bernardi, M. Cimitile, M. Iammarino, and C. Verdone, "An enhanced UNet variant for effective lung cancer detection," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Padua, Italy, Jul. 2022, pp. 1–8.

[25] L. Aversano, M. L. Bernardi, M. Cimitile, and R. Pecori, "Deep neural networks ensemble to detect COVID-19 from CT scans," *Pattern Recognit.*, vol. 120, Dec. 2021, Art. no. 108135.

[26] M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2528–2535.

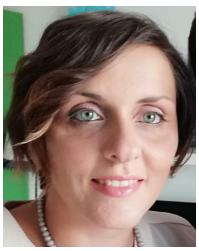
[27] M. Lin, Q. Chen, and S. Yan, "Network in network," *CoRR*, vol. abs/1312.4400, pp. 1–14, Dec. 2014.

[28] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. ECCV*, 2016, pp. 391–407.

[29] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[30] A. Martinez and J. L. Marchand, "SAR image quality assessment," Tech. Rep., 1993. [Online]. Available: https://www.researchgate.net/publication/39195289_SAR_Image_Quality_Assessment

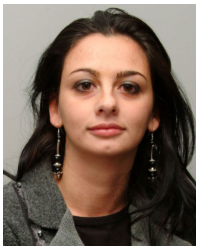
[31] *COSMO-SkyMed SAR Products Handbook*, Italian Space Agency (ASI)-COSMO-SkyMed Mission, Rome, Italy. [Online]. Available: https://earth.esa.int/eogateway/documents/20142/37627/Cosmo-SkyMed-Product-Handbook.pdf



PIA ADDABBO (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in telecommunication engineering and the Ph.D. degree in information engineering from the Università degli Studi del Sannio, Benevento, Italy, in 2005, 2008, and 2012, respectively. She is currently an Associate Professor with Giustino Fortunato University, Benevento. She is the coauthor of many scientific publications in international journals and conferences. Her research interests include statistical signal processing applied to radar target recognition, global navigation satellite system reflectometry, and hyperspectral unmixing. She is currently an Associate Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING, IEEE ACCESS, and *Scientific Reports*.

MARIO LUCA BERNARDI (Member, IEEE) received the Laurea degree in computer science engineering from the University of Naples "Federico II," Italy, in 2003, and the Ph.D. degree in information engineering from the University of Sannio, in 2007. Since 2003, he has been a Researcher in software engineering with the University of Sannio. He has authored more than 70 articles published in journals and conference proceedings. His main research interests include software maintenance and testing, software development, and architectural design, with a particular interest in data mining, artificial intelligence, and computational intelligence. He has served and is serving in the program and organizing committees of conferences. He was a referee of main journals and magazines. He was involved in several projects on tasks related to software engineering, software maintenance, quality assurance, and artificial intelligence applied in different contexts and domains, for both the industry and services sectors.

FILIPPO BIONDI (Member, IEEE) was born in L'Aquila, Italy, in 1974. He received the B.S. degree in information engineering and the M.S. degree in telecommunication engineering from Salento University, Lecce, Italy, in 2009 and 2011, respectively, and the Ph.D. degree in telecommunication engineering from the University of L'Aquila, L'Aquila, in 2015. His research interests include the development of signal processing and statistical signal processing techniques on radar equipment. In particular, he is working on synthetic aperture radar (SAR), interferometry, SAR-polarimetry, multichromatic analysis, and tomography, also using a convex optimization approach. From April 2010 to April 2011, he was a Guest Scientist with the Nation Research Council (ISSIA-CNR), Bari, Italy, and the Italian Space Agency, Geodesy Space Center, Matera, Italy. He is currently an Associate Editor of IEEE TRANSACTIONS ON GEOSCIENCE AND REMOTE SENSING.



MARTA CIMITILE (Member, IEEE) received the Ph.D. degree in computer science from the Department of Computer Science, University of Bari, in 2008. She is currently an Associate Professor with Unitelma Sapienza University, Rome, Italy. She received the Italian Scientific Qualification for Full Professor position in computer science engineering, in 2021. In the last years, she was involved in several industrial and research projects. She is a Founding Member of the spin-off of the University of Bari, named Software Engineering Research and Practices s.r.l. She has published more than 110 articles in international conferences and journals. Her main research interests include convolutional neural networks and process and data mining. She has served in the program and organizing committees of several international conferences. She is on the editorial board of *Artificial Intelligence Review*. She is an Associate Editor of IEEE ACCESS.

CARMINE CLEMENTE (Senior Member, IEEE) received the Laurea (B.Sc.) (cum laude) and Laurea Specialistica (M.Sc.) (cum laude) degrees in telecommunications engineering from the Universita degli Studi del Sannio, Benevento, Italy, in 2006 and 2009, respectively, and the Ph.D. degree in electronic and electrical engineering from the University of Strathclyde, Glasgow, U.K., in 2012. He is currently a Reader with the Department of Electronic and Electrical Engineering, University of Strathclyde, working on advanced radar signal processing algorithms, micro-doppler analysis, SAR image processing, and advanced radar applications. He directs the Sensor Signal Processing and Security Laboratories, University of Strathclyde, where his group's research interests include micro-Doppler signature analysis and extraction, waveform design, passive and multistatic radars, automatic target recognition, electronic surveillance, space situation awareness, ballistic missile defense, automotive radars, applications of radars in industry 4.0, and agritech and statistical signal processing. He is the Chair of the Advanced Signal Processing Focus Group of the Electromagnetic Sensing Interest Group—the U.K. Radar Society. He is an Associate Editor of the IEEE TRANSACTIONS ON AEROSPACE AND ELECTRONIC SYSTEMS.

NICOMINO FISCANTE (Student Member, IEEE) was born in Benevento, Italy, in January 1982. He received the M.Sc. degree (cum laude) in telecommunication engineering from the University of Sannio, Benevento, in 2007. From January to June 2007, he was a Trainee with the Antenna Group, The Netherlands Organization for Applied Scientific Research (TNO Defense, Security and Safety), The Hague, The Netherlands. From 2008 to 2012, he was with the Mediterranean Agency for Remote Sensing and Environmental Control (MARSec), Benevento. From 2012 to 2014, he was with Advanced Research Technology Spa, Passignano sul Trasimeno, Italy. Since 2014, he has been a Project Manager with Geoslab (currently GeneGIS GI). He has vast work experience both nationally and internationally, thanks to the collaboration with important companies, research centers, universities and institutions. His research interests include remote sensing, GIS, data processing, TLC and EM systems, mobile mapping systems, and renewable resources.

GAETANO GIUNTA (Senior Member, IEEE) received the degree in electronic engineering from the University of Pisa, Italy, in 1985, and the Ph.D. degree in information and communication engineering from the University of Rome La Sapienza, Italy, in 1990. Since 1989, he has been a Research Fellow with the Signal Processing Laboratory (LTS), EPFL, Lausanne, Switzerland. In 1992, he was an Assistant Professor with the INFO-COM Department, University of Rome La Sapienza. From 2001 to 2005, he was with the Third University of Rome as an Associate Professor, where he has been a Full Professor of telecommunications, since 2005. His research interests include signal processing for mobile communications, image communications, and security. He was a member of the IEEE Societies of Communications, Signal Processing, and Vehicular Technology. He has served as a Reviewer for several IEEE TRANSACTIONS, IET (formerly IEE) proceedings, and EURASIP journals, and a TPC Member for several international conferences and symposia in the above mentioned fields.

DANILO ORLANDO (Senior Member, IEEE) was born in Gagliano del Capo, Italy, in August 9, 1978. He received the Dr.Eng. degree (Hons.) in computer engineering and the Ph.D. degree in information engineering from the University of Salento (formerly University of Lecce), Lecce, Italy, in 2004 and 2008, respectively. From July 2007 to July 2010, he was with the University of Cassino, Cassino, Italy, where he was engaged in a research project on algorithms for track-before-detect of multiple targets in uncertain scenarios. From September to November 2009, he was a Visiting Scientist with the NATO Undersea Research Centre, La Spezia, Italy. From September 2011 to April 2015, he was with Elettronica S.p.A., as a System Analyst in the field of electronic warfare. In May 2015, he joined Università degli Studi Niccolò Cusano, Rome, Italy, where he is currently an Associate Professor. In 2007, he held visiting positions with the Department of Avionics and Systems, ENSICA (currently Institute Supérieur de l'Aéronautique et de l'Espace, ISAE), Toulouse, France, and from 2017 to 2019, he was with the Chinese Academy of Sciences, Beijing, China. He is the author or coauthor of more than 150 scientific publications in international journals, conferences, and books. His main research interests include statistical signal processing with more emphasis on adaptive detection and tracking of multiple targets in multisensor scenarios. He was a Senior Area Editor of IEEE TRANSACTIONS ON SIGNAL PROCESSING, and currently an Associate Editor of IEEE OPEN JOURNAL ON SIGNAL PROCESSING, *EURASIP Journal on Advances in Signal Processing*, and *Remote Sensing* (MDPI).

LINJIE YAN received the B.E. degree in communication engineering from the Shandong University of Science and Technology, in 2016, and the Ph.D. degree in signal and information processing from the Institute of Acoustics, Chinese Academy of Sciences, Beijing, China, in 2021. She is currently a Postdoctoral Researcher with the Institute of Acoustics, Chinese Academy of Sciences, Beijing. Her research interests include statistical signal processing, with more emphasis on adaptive sonar and radar signal processing.

Open Access funding provided by 'Università degli Studi Roma Tre' within the CRUI CARE Agreement