Autoencoder for On-Board InSAR Phase Denoising and Compression

Lorenzo Bruno Garavelli[a], Luca Dell'Amore[a], Nicola Gollin[a], Michele Martone[a], and Paola Rizzoli[a]
[a]Microwaves and Radar Institute, German Aerospace Centre (DLR), 82234 Weßling, Germany

Abstract

In the last decades, SAR missions have been proposed in the framework of interplanetary exploration, featuring the on-board generation of higher-level products, such as SAR interferograms. However, the limited down-link capacity of the satellites has posed significant challenges, thus motivating the development of efficient on-board compression strategies. In this work, we present a novel deep-learning approach, through the implementation of a Convolutional AutoEncoder (CAE), which allows for the joint denoising and compression of the interferometric (InSAR) phase. The proposed network is trained and tested using synthetic datasets, derived starting from real TanDEM-X observations and assuming corresponding InSAR acquisition geometries and underlying topography. Results are assessed against a combination of boxcar filtering and JPEG 2000 compression, which reflects one of the possible strategies reported in the literature. In particular, we focus on three different performance metrics, i.e. denoising capability, data volume reduction and preservation of high-resolution details, thus showing the enhanced flexibility of the proposed methodology with respect to a state-of-the-art baseline method.

1 Introduction

Synthetic Aperture Radar Interferometry (InSAR) is a key technology in Remote Sensing, providing unparalleled sensitivity from space for multiple applications, above all Digital Elevation Model (DEM) generation and deformation monitoring. This capability relies on the phase difference between two SAR acquisitions, acquired from two different positions or at different times. However, as SAR imaging is degraded by several noise sources, denoising strategies are crucial in order to properly estimate InSAR parameters. Beside the standard approach, i.e., boxcar filtering [1], consisting of an averaging operation over a sliding window, several strategies can be found in the literature, ranging from non-local filters [2, 3, 4, 5] to Convolutional Neural Networks (CNN) [6, 7, 8]. In the framework of interplanetary exploration, SAR has already been exploited for several missions, especially on Venus, such as Venera 15 & 16 [9] and Magellan [10], paving the way to more recent ones, like VERITAS [11]. Here, on-board InSAR processing has been proposed in the context of a global Digital Elevation Model (DEM) generation, with the interferometric phase being of primary interest [12]. However, given current technologies and the limited down-link capacity of the satellites, the definition of effective on-board compression strategies has become a crucial aspect for such missions.

In this sense, we propose a deep-learning-based technique to jointly denoise and compress the interferometric phase on-board. While Convolutional AutoEncoders (CAE) are a well-established tool in computer vision and have been already successfully applied to SAR backscatter, e.g. in SAR despeckling [13], their application to InSAR phase is far less studied. Therefore, this work explores the use of a CAE to handle the wrapped interferometric phase, with the goal of achieving joint denoising and compression within a single end-to-end trainable pipeline. The network is trained in a supervised manner using a synthetic dataset, derived by exploiting real TanDEM-X data together with corresponding InSAR geometries and underlying topography. Such an architecture is naturally divided in two parts: an encoder and a decoder. By placing the former on-board the satellite and the latter on ground, data compression is achieved, as the transmitted latent space is significantly smaller in size than the original data. The proposed architecture is designed in a parametric way, thus allowing for hyper-parameter tuning. A family of models is trained targeting three different performance metrics: denoising capability, compression ratio and high-resolution detail preservation. Results are then compared to a more traditional method, i.e., a combination of boxcar filtering [1] and JPEG 2000 compression [14], which will be referred to as the baseline approach in this study.

This paper is structured as follows: in Section 2 we present the proposed methodology, from dataset generation to testing; in Section 3 we present and discuss the results; conclusions and outlook on future activities are drawn in Section 4.

2 Methodology

In this section we present the framework developed in the current work. First, we address the dataset generation part. Then, we introduce the baseline/comparison method, consisting of boxcar filtering and JPEG 2000 compression. After that, the proposed Convolutional AutoEncoder architecture is presented, together with the corresponding hyper-parameter tuning. Finally, we discuss the considered performance metrics: denoising capability, compression ratio and high-resolution detail preservation.

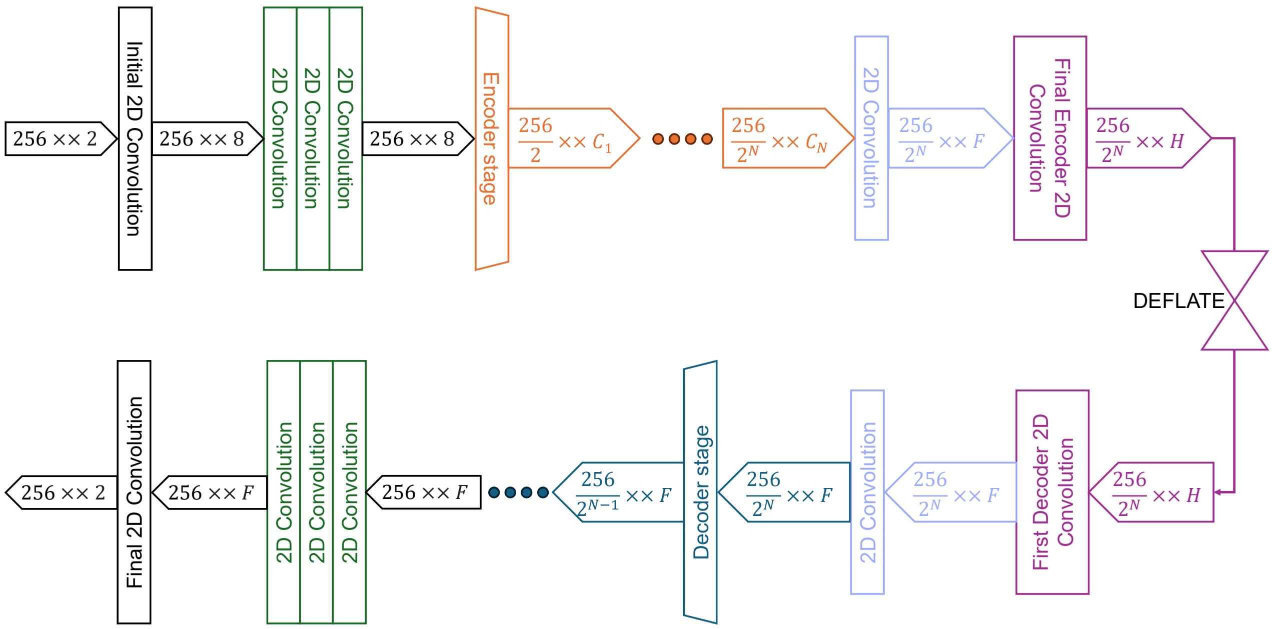


Figure 1 Proposed autoencoder architecture. The upper part corresponds to the encoder part, while the lower one to the decoder. Each layer is a block, with arrows indicating the dimensionality of the input and output tensors. Note that the encoder/decoder stages consist of 2D convolutions followed by a pooling/up-sampling operation, which reduces/increases the output spatial dimensionality.

2.1 Dataset

In this work we rely entirely on synthetic interferograms, generated starting from real TanDEM-X quantities and by considering the actual InSAR geometries together with the corresponding topography on ground. The complex interferogram Γ is given by:

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

where z_i, for $i \in [1, 2]$, are the single-look-complex (SLC) pair and * denotes the complex conjugate operator. The single SLCs can be defined as:

$$z_i = c_i + n_i \tag{2}$$

where c_i is the noise-free contribution and n_i the additive noise term. Both c_i and n_i are modelled as complex Gaussian random processes taking advantage of the Cholesky decomposition of the covariance matrix, similarly as it was done in [3, 4, 5, 6].

2.2 Baseline method

Below, we introduce the baseline method used for comparison purposes.

As far as phase denoising is concerned, we employ a boxcar filter, consisting of an averaging operation over a sliding square window. Note that denoising is achieved at the cost of spatial resolution, thus directly impacting the obtained compression rate as well. In particular, different windows are considered in order to investigate different trade-offs between denoising and compression performance.

Moreover, we compress the denoised phase by exploiting the JPEG 2000 standard, being the basis of the recommendation on Image Data Compression of the Consultative Committee for Space Data Systems (CCSDS) [14]. Multiple JPEG 2000 outputs, with different compression ratios, are produced, ranging from lossless to one fifth of the uncompressed size.

2.3 Autoencoder architecture and hyper-parameter tuning

Below, we describe the proposed Convolutional AutoEncoder (CAE) architecture. Note that it is defined in a parametric way, so that different trade-offs between denoising, compression and high-resolution detail preservation can be investigated. The autoencoder peculiarity consists in the separability of the encoder and decoder parts, thus allowing, e.g., for placing the former on-board a satellite and the latter on ground. Note that, as the encoder reduces the spatial dimensions of the input, while the decoder reconstructs it in the output by exploiting standard up-sampling techniques, the original high-resolution spatial details are partially lost.

Figure 1 shows the considered parametric architecture: each layer is represented by a block, with the arrows in between reporting the dimensionality of the flowing tensors; the encoder is displayed on the upper part of the figure, while the decoder on the lower part. The encoder output is called latent space; it is a compressed representation of the original input data and corresponds to the signal that has to be down-linked to ground. Note that a further compression is achieved using the DEFLATE algorithm [15].

All layers contain at least a bi-dimensional convolution, which is followed by a pooling operation in the encoder stages and by an up-sampling in the decoder ones. The parameters that control the generated models are the following:

1. N is the number of encoder stages, which is also equal to the decoder ones, giving the autoencoder symmetry;
2. C_k is the number of output channels of the k^{th} convolutional block of the encoder. It is computed as $C_k = \lfloor MC_{k-1} \rfloor$, with M being a settable hyper-parameter.
3. F is the number of channels in output from the encoder's penultimate convolution, which is the same throughout the decoder.
4. H is the number of channels of the latent space.

Note that N is proportional to the amount of pooling operations, thus controlling the spatial dimensionality of the latent space. On the other hand, H controls the features space, as it sets the number of output channels. Therefore, for a patch of n_{az} pixels in azimuth and n_{rg} pixels in range, the compression ratio, except for the DEFLATE contribution, is fixed for given N and H, with the latent space shape being: $\left[\frac{n_{\text{az}}}{2^N} \quad \frac{n_{\text{rg}}}{2^N} \quad H\right]$.

2.4 Training

The utilized dataset comprises 75367 patches with size equal to [256 × 256]. Note that 80% of the patches are used for training, while the remaining 20% for validation. The considered loss function is defined as:

$$\mathcal{L} = \sum_i h(\rho_i) \left|\phi_{\text{err}\,i}\right| + l_2 \sum_k w_k^2 \tag{3}$$

with ϕ_{err} being the phase error, $h(\rho)$ a shifted logistic function dependent on the coherence magnitude ρ and $|\cdot|$ the absolute value operator. Also, w_k indicates the network's weights while l_2 is a L2-regularisation factor, which is empirically set to 10^{-7}. The phase error ϕ_{err} is defined as:

$$\phi_{\text{err}} = \angle\left(\frac{\Gamma_{\text{out}}}{\Gamma_{\text{ref}}}\right) \tag{4}$$

where Γ_{ref} and Γ_{out} correspond to the desired noise-free and noisy complex interferograms, respectively, while $\angle$ indicates the phase operator.

2.5 Performance metrics

In order to assess the performance of the proposed methodology, we consider three different performance aspects: denoising capability, compression ratio and preservation of high-resolution details. A detailed explanation of the utilised metrics is reported in the following list:

1. The Standard Deviation of the Phase Error ($\boldsymbol{\sigma}_{\boldsymbol{\phi},\text{err}}$) across all test scenes is used as indicator for the denoising performance. It is defined as:

$$\sigma_{\phi,\text{err}} = \sqrt{\frac{\sum_{i=1}^{m}\left(\phi_{\text{err}\,i} - \bar{\phi}_{\text{err}}\right)^2}{m-1}} \tag{5}$$

with $\bar{\phi}_{\text{err}}$ being the mean phase error and m the total number of pixels from the whole test dataset.

2. The achieved compression ratio is evaluated using the Bits Per Original Pixel (**BPOP**) metric. This represents the number of bits used by the compressed latent space with respect to the number of pixels in the input data.
3. In order to assess the amount of **preserved high-resolution details** in the autoencoder output, a custom procedure is utilised. This involves Fourier transforming the desired noise-free reference phase and applying ellipsoidal Gaussian filters, which remove increasing fractions of the frequency content. After an inverse Fourier transform, the filtered version of the reference with the lowest $\sigma_{\phi,\text{err}}$ with respect to the decoder reconstruction determines the percentage of preserved spatial details. Note that this procedure produces consistent results with the baseline method, where the expected preserved resolution is inversely proportional to the boxcar window size.

3 Results

As explained in Section 2.5, we assess the performance of the proposed method in terms of denoising capability, data volume reduction and high-resolution detail preservation. Some results are shown in Figure 2. First, we focus on the baseline performance, for all considered boxcar window sizes; as expected, a better denoising is achieved at the cost of spatial resolution. Indeed, when increasing the window size, the phase error decreases, while, on the other hand, less resolution is preserved. Also, bigger windows lead to higher compression ratios. The four points corresponding to each boxcar filter refer to different JPEG 2000 compression ratios, ranging from lossless to one fifth of the uncompressed data size, as explained in Section 2.2. As far as the proposed methodology is concerned, we show here a subset of the trained models, divided in three groups, each of which is tuned on a specific performance metric. Note that we also provide an additional group of models which targets good compromises between all three performance measures. All these 24 models are efficient according to the Pareto definition, meaning that no other model enhances any performance metric while matching the remaining ones. It is worth mentioning that the proposed autoencoder allows for a higher overall flexibility compared to the baseline method, since we can now explore the whole performance space by tuning specific hyper-parameters during training. For example, a better denoising is expected at the cost of data volume reduction (yellow squares); also, in these cases, the corresponding models are among the best at spatial details preservation, eclipsed only by those optimised for this specific task (purple squares). On the other hand, the best hi-res details preservation is achieved at the cost of phase denoising capability, similarly to, e.g., a [3 × 3] boxcar filter. Moreover, the CAE allows for an overall good trade-off between the three performance measures, even if tuning the network for the compression task

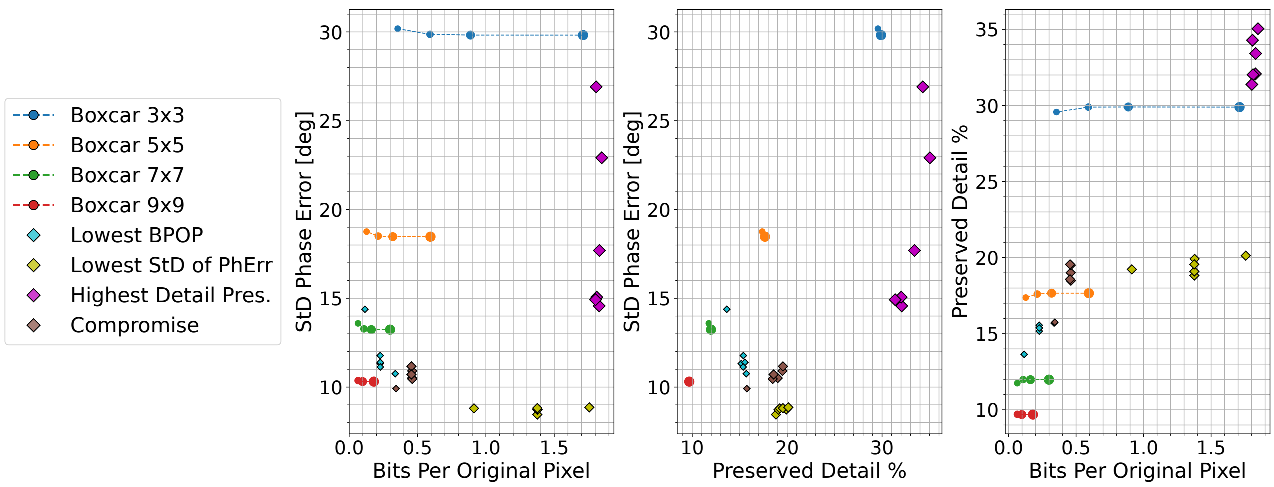


Figure 2 Graphical representation of the proposed autoencoder performance with respect to the three investigated performance metrics, i.e., Standard Deviation of the Phase Error (lower is better), Bits Per Original Pixel (lower is better) and Preserved high-resolution Details percentage (higher is better). The squares represent four sets of trained autoencoder models: three are tuned in order to optimise a specific performance metric, while the last one refers to the best performance trade-offs. As comparison, circles indicate the baseline method, with different colors assigned to different boxcar window sizes. Moving on the dashed boxcar lines, different JPEG 2000 compression ratios are encountered: 1 (lossless), 2, 3, 5.

only (cyan squares). Finally, the compromise solutions (brown squares), show the best performance trade-offs and are valid competitors to the considered baseline approach.

In order to take a closer look at the InSAR phase estimated by the decoder, we select the top models from the four groups in Figure 2, i.e., cyan, yellow, purple and brown squares. For each of them, the corresponding InSAR phase reconstruction (zoom-in) is provided in Figure 3, along with the original synthetic noise-free and noisy phases. Moreover, the figure shows two baseline methods as well, i.e., boxcar [5 × 5] and [7 × 7], with a JPEG 2000 compression ratio of 2. Note that the corresponding denoising performance is consistent with the above discussion of Figure 2.

4 Conclusion and outlook

In this paper we proposed a novel deep-learning-based methodology for the joint denoising and compression of the interferometric (InSAR) phase, in the framework of on-board InSAR processing for future interplanetary missions. The method relies on the definition of a Convolutional AutoEncoder (CAE), which allows for placing the encoder on-board the satellite and the decoder on ground. Synthetic interferograms, generated starting from real TanDEM-X observations and assuming corresponding InSAR geometries and underlying topography, were used throughout training and testing. Also, we conducted a hyper-parameter tuning on different settings, such as network depth and number of channels/features in the resulting latent space. In order to assess the effectiveness of the proposed strategy, we investigated three different performance metrics: denoising capability, data volume reduction and preservation of high-resolution spatial details. Moreover, we compared the achieved results to a more traditional state-of-the-art procedure, consisting of a combination of boxcar filtering and JPEG 2000 compression. The presented results look promising and have shown already a higher overall flexibility with respect to the baseline method, in terms of performance scalability, even if they do not excel in compression.

Further investigations shall compare the computational costs of the proposed CAE-based approach against the baseline method, as a first step towards implementation. Moreover, a complete assessment on final generated DEMs will be required. Lastly, the proposed CAE could be extended to encode the interferometric coherence as well, thus enhancing the corresponding compression gain relative to the most common baseline strategies.

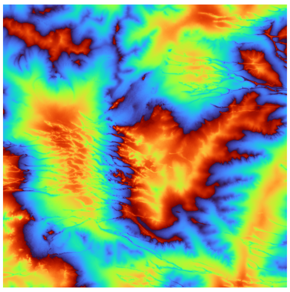
5 Literature

[1] R. Bamler and P. Hartl, ‘Synthetic aperture radar interferometry’, Feb. 1998, doi: 10.1088/0266-5611/14/4/001.

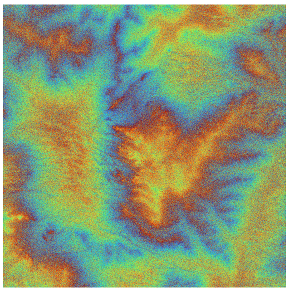
[2] G. Xu, Y. Gao, J. Li, and M. Xing, ‘InSAR Phase Denoising: A Review of Current Technologies and Future Directions’, IEEE Geoscience and Remote Sensing Magazine, vol. 8, no. 2, pp. 64–82, June 2020, doi: 10.1109/MGRS.2019.2955120.

[3] C.-A. Deledalle, L. Denis, F. Tupin, A. Reigber, and M. Jäger, ‘NL-SAR: a unified Non-Local framework for resolution-preserving (Pol)(In)SAR denoising’, IEEE Transactions on Geoscience and Remote Sensing, vol. 53, no. 4, pp. 2021–2038, Apr. 2015, doi: 10.1109/TGRS.2014.2352555.

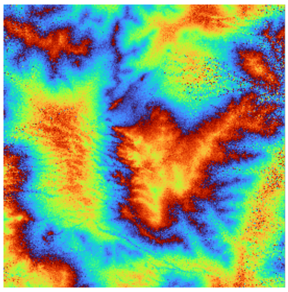
[4] F. Sica, D. Cozzolino, X. X. Zhu, L. Verdoliva, and G. Poggi, ‘InSAR-BM3D: A Nonlocal Filter for SAR In-



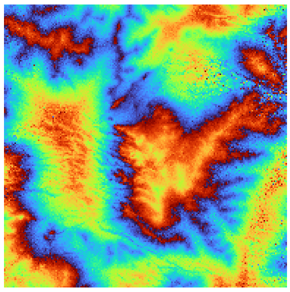
(a) Noise-free reference



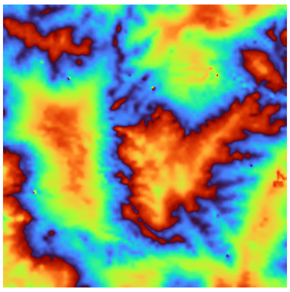
(b) Noisy input



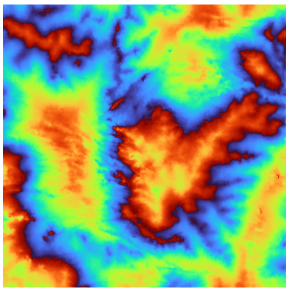
(c) 5×5 Boxcar JPEG2000 r2



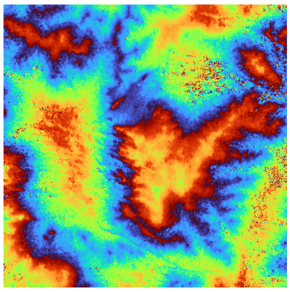
(d) 7×7 Boxcar JPEG2000 r2



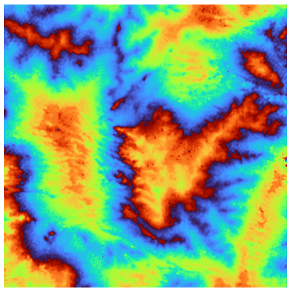
(e) Smallest latent space



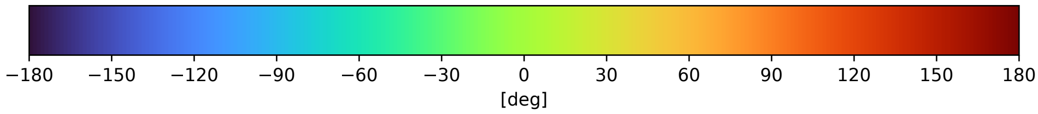
(f) Best denoising



(g) Best hi-res details preservation



(h) Compromise



(i) Phase color-bar

Figure 3 InSAR phase reconstruction (zoom-in) from the top models of the four presented groups in Figure 2, i.e., cyan (e), yellow (f), purple (g) and brown (h) squares. The original synthetic noise-free and noisy phases are shown as well (a-b) together with two baseline methods (c-d), i.e., boxcar $[5 \times 5]$ and $[7 \times 7]$ with a JPEG2000 compression ratio of 2.

terferometric Phase Restoration', IEEE Trans. Geosci. Remote Sensing, vol. 56, no. 6, pp. 3456–3467, June 2018, doi: 10.1109/TGRS.2018.2800087.

[5] F. Sica, D. Cozzolino, L. Verdoliva, and G. Poggi, 'The Offset-Compensated Nonlocal Filtering of Interferometric Phase', Remote Sensing, vol. 10, no. 9, p. 1359, Aug. 2018, doi: 10.3390/rs10091359.

[6] F. Sica, G. Gobbi, P. Rizzoli, and L. Bruzzone, 'Φ-Net: Deep Residual Learning for InSAR Parameters Estimation', IEEE Trans. Geosci. Remote Sensing, vol. 59, no. 5, pp. 3917–3941, May 2021, doi: 10.1109/TGRS.2020.3020427

[7] X. Sun, A. Zimmer, S. Mukherjee, N. K. Kottayil, P. Ghuman, and I. Cheng, 'DeepInSAR—A Deep Learning Framework for SAR Interferometric Phase Restoration and Coherence Estimation', Remote Sensing, vol. 12, no. 14, p. 2340, July 2020, doi: 10.3390/rs12142340.

[8] B. Liu, L. Wu, H. Hao, and J. Dong, 'Interferometric phase image denoising method via residual learning', J. Electron. Imag., vol. 30, no. 02, Mar. 2021, doi: 10.1117/1.JEI.30.2.023013.

[9] A. F. Bogomolov, N. V. Zherikhin, and G. A. Sokolov, 'Venera 15 and 16: Synthesized aperture radar in orbit around Venus', Radiophys Quantum Electron, vol. 28, no. 3, pp. 169–181, Mar. 1985, doi: 10.1007/BF01035473.

[10] R. S. Saunders, G. H. Pettengill, R. E. Arvidson, W. L. Sjogren, W. T. K. Johnson, and L. Pieri, 'The Magellan Venus Radar Mapping Mission', Journal of Geophysical Research: Solid Earth, vol. 95, no. B6, pp. 8339–8355, 1990, doi: 10.1029/JB095iB06p08339.

[11] S. Smrekar et al., 'VERITAS (Venus Emissivity, Radio Science, InSAR, Topography, and Spectroscopy): A Discovery Mission', in 2022 IEEE Aerospace Conference (AERO), Mar. 2022, pp. 1–20. doi: 10.1109/AERO53065.2022.9843269.

[12] M. Younis et al., 'Venus Interferometric Synthetic Aperture Radar Instrument Performance and Options', in IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium, Pasadena, CA, USA: IEEE, July 2023, pp. 4181–4183. doi: 10.1109/IGARSS52108.2023.10282968.

[13] E. Dalsasso, L. Denis, and F. Tupin, 'As If by Magic: Self-Supervised Training of Deep Despeckling Networks With MERLIN', IEEE Trans. Geosci. Remote Sensing, vol. 60, pp. 1–13, 2022, doi: 10.1109/TGRS.2021.3128621.

[14] Image Data Compression, BLUE BOOK 122.0-B-2, 2017

[15] L. P. Deutsch, 'DEFLATE Compressed Data Format Specification version 1.3', Internet Engineering Task Force, Request for Comments RFC 1951, May 1996. doi: 10.17487/RFC1951.