SUPER-RESOLUTION OF SENTINEL-1 IMAGERY USING AN ENHANCED ATTENTION NETWORK AND REAL GROUND TRUTH DATA

Christian Ayala & Juan Francisco Amieva
Tracasa Instrumental
Calle Cabárceno 6, 31621 Sarriguren, Spain
{cayala, jfamieva}@itracasa.es

Mikel Galar
Institute of Smart Cities (ISC), Public University of Navarre (UPNA)
Arrosadia Campus, 31006 Pamplona, Spain
{mikel.galar}@unavarra.es

ABSTRACT

Active imaging systems, particularly Synthetic Aperture Radar (SAR), offer notable advantages such as the ability to operate in diverse weather conditions and provide day and night observations of Earth's surface. These attributes are especially valuable when monitoring regions consistently obscured by clouds, as seen in Northern Europe. One of the most recognized SAR constellations is Sentinel-1 (S1), known for providing imagery freely to the community. Despite this accessibility, problems arise due to the inherent limitations of the spatial resolution of S1 and the presence of speckle noise, which makes the data difficult to interpret. Although there are several commercial SAR satellites offering on-demand high-resolution data, their high costs hinder their use among remote sensing experts. Motivated by the outlined advantages and limitations, this paper introduces a novel deep learning-based methodology aimed at simultaneously reducing speckle noise and enhancing the spatial resolution of S1 data. Contrary to previous works that rely on a high-resolution satellite as ground truth (typically TerraSAR-X), we propose to use the same satellite in another operational mode as ground truth. Accordingly, the proposed method focuses on enhancing the spatial resolution of S1 Interferometric Wide Swath mode products from 10 to 5 m GSD by leveraging S1 Stripmap mode as the ground truth for training the model. As a result, super-resolved images duplicated the input spatial resolution, closing the gap between S1 and commercial SAR satellites.

1 INTRODUCTION

Synthetic Aperture Radar (SAR) imagery is a fundamental source of information for monitoring targets on the Earth's surface due to its ability to penetrate clouds and operate during both daytime and nighttime, regardless of weather conditions Torres et al. (2012). However, due to its nature, SAR data is strongly affected by speckle noise, which results from the coherent summation of echoes generated by elementary scatterers within the same resolution cell Lee et al. (1994). Consequently, SAR images become challenging to interpret and utilize for both remote sensing experts and automated tools.

Nowadays, there is a great deal of SAR constellations operated by both public and private companies. While public SAR constellations provide extensive global coverage at the expense of spatial resolution, private ones offer very high-resolution products with shorter revisit times, making interpretation easier Curlander & McDonough (1991). However, the prohibitive costs associated with private constellations hinder remote sensing experts from utilizing them.

Sentinel-1 (S1), operated by the European Space Agency as part of the Copernicus program, stands out as one of the most distinguished SAR constellations. S1 provides data through a dual-polarization C-band SAR instrument operating in four different modes, each designed to capture data with varying spatial resolutions and coverage to meet diverse Earth observation needs Filipponi (2019). Nevertheless, despite the ease of access to its data, S1 also has some limitations, such as low spatial resolution and significant speckle noise, adding complexity to the processing and interpretation of its products.

There are many works in the literature aiming at reducing or removing speckle noise from SAR images, enhancing their visual quality, and improving the accuracy of subsequent image analysis tasks Fracastoro et al. (2021). Traditionally, the majority of these works fall into two categories: Bayesian and non-Bayesian methods Singh et al. (2021). While Bayesian methods are grounded in a probabilistic framework, allowing for modeling uncertainty and incorporating contextual information, non-Bayesian methods are often faster and simpler, relying on deterministic filtering techniques.

Nevertheless, recently, despeckling tasks have benefited from advancements in deep learning, clearly outperforming established approaches. Chierchia et al. (2017) employed a convolutional neural network trained using a residual learning strategy with synthetic SAR data. This synthetic SAR data was generated by injecting single-look speckle in amplitude format into optical images. A similar approach was followed by Lattari et al. (2019), who pre-trained a modified U-Net network using synthetic SAR data generated by simulating speckle on optical imagery, employing a Rayleigh distribution. In contrast to the aforementioned works, Dalsasso et al. (2020) pre-trained their model using, as ground truth, an image obtained by temporal and spatial filtering of a stack of multiple SAR images. Other approaches focus on preventing the appearance of artifacts while preserving important details such as edges. This is the case of Wang et al. (2017), who proposed combining standard restoration losses with Total Variation regularization to encourage smoother results. It must be noted that, although these approaches enhance the visual quality of the images by reducing speckle noise, they do not improve spatial resolution.

To increase the spatial resolution, the majority of works within SAR super-resolution literature are based on Generative Adversarial Networks (GANs) Goodfellow et al. (2014). Wang et al. (2018) trained an SRGAN generating low-resolution images by downsampling TerraSAR-X counterparts by a factor of four. Zheng et al. (2019) followed a similar approach, training their GAN-based network with pairs of TerraSAR-X and MSTAR images. Similarly, Ao et al. (2018) designed a Dialectical GAN to super-resolve S1 images by transferring the style of TerraSAR-X. However, since these methods are based on generative networks, models may easily hallucinate or generate synthetic artifacts having a negative impact on the performance of downstream tasks. Moreover, working with undersampled versions of the original input data is known to achieve suboptimal results, whereas working with image pairs coming from different satellites involves challenges due to their differences regarding corregistration, operating frequency, polarizations, swath width and so on Galar et al. (2020).

To overcome these problems, we propose to exploit the operational modes of S1, taking advantage of Stripmap (SM) mode, where S1 achieves a resolution of 5×5 m GSD, compared with the standard working mode of Interferometric Wide Swath (IW) working at 5×20 m GSD. This way, we can form pairs of images coming from the same satellite to train standard single-image super-resolution networks, avoiding generative adversarial training. Note that finding pairs of IW and SM images is not straighforward as SM mode is not available in all locations.

2 METHODOLOGY

The super-resolution model is trained in two phases: pre-training and fine-tuning. Firstly, the model is pre-trained, either with synthetically generated SAR data derived from optical images or with aggregated SAR time-series data. After pre-training, the model is fine-tuned using S1 SM data.

2.1 DATASET GENERATION

We consider synthetic SAR data for model pre-training and real SAR data for fine-tuning.

Pre-training dataset: For the pre-training phase, a multi-temporal dataset of pairs of S1 and Sentinel-2 (S2) optical images has been used. This dataset comprises a selection of 44 cities spread across the Spanish territory Ayala et al. (2021). Synthetic SAR data has been generated from both S2 optical images and S1 time-series following different strategies, as described in Section 2.3.

Fine-tuning dataset: For the fine-tuning phase, we generated a curated dataset of pairs consisting of low-resolution and high-resolution SAR data. In contrast to other works within the literature, we propose to use the same sensor, S1, in different operational modes to create these pairs. While the IW mode offers wide coverage at a fixed spatial resolution of 5×20 m GSD, the SM mode provides higher spatial resolution (5×5 m GSD), making it ideal for detailed observations of specific regions. It must be noted that the latter is only used for small islands and upon request for extraordinary events, such as emergency management. However, between 2017 and 2020, a few S1 SM acquisitions over land were made in Europe, Asia and Africa.

The fine-tuning dataset was generated by selecting pairs of S1 IW and SM images. The selection process focused on continental areas, considering SM beams ranging from S3 to S6 to match the incidence angle range of the IW products. It must be noted that only the ascending flight direction mode and dual vertical polarizations (VV+VH) have been considered to maximize coverage. To enhance diversity and prevent the overrepresentation of highly homogeneous areas, a careful visual inspection was conducted. As a result, the dataset comprises four areas of interest, covering a total area of 25,871 km^2.

2.2 Model architecture

We opted for using the Second-Order Attention Network proposed by Dai et al., for Single-Image Super Resolution of RGB images Dai et al. (2019). This architecture allows for more powerful feature expression and feature correlation learning by leveraging two elements: the second-order channel attention module and the non-locally enhanced residual group structure. In our case, we have chosen to use 20 residual groups, each comprising 10 residual blocks, based on preliminary experiments.

2.3 Model pre-training

As described in Section 1, there are different ways to pre-train models depending on whether the synthetic SAR data is generated from optical data or SAR time-series. Both strategies are considered:

From optical data: Synthetic SAR data is generated by injecting speckle and Rayleigh noise into S2 optical images, following the approaches of Chierchia et al. (2017) and Lattari et al. (2019), respectively. Specifically, noise was injected into S2 grayscale images, which were created by averaging the values of the Red, Green, and Blue bands to generate the low-resolution input image. On the other hand, the noise-free S2 grayscale images were upsampled by a factor of 2 to generate the high-resolution ground truths.

From SAR time-series: Synthetic SAR data is generated by performing multi-temporal multilooking to reduce speckle noise, following the approach of Dalsasso et al. (2020). This entails combining the intensity of images acquired at different dates, assuming that no changes occurred. In this regard, the dataset described in Section 2.1 comprising stacks of 4 S1 IW images has been used. Specifically, an arbitrary time step is selected as the low-resolution input image, while different aggregation strategies are employed to combine the 4 time steps (upsampled by a factor of 2 beforehand) to generate the high-resolution ground truths. In this study, mean and median operations are considered as aggregation strategies.

To conduct the pre-training, a consistent training schema was followed across all approaches. Specifically, models were trained for 100 epochs, taking batches of 128 24×24 low-resolution samples. For the loss function, a combination of the L_2 and Total Variation losses has been chosen as suggested in Wang et al. (2017), to prevent the appearance of artifacts while preserving important details. The loss has been minimized using the OneCycleLR scheduler Smith & Topin (2018) with a maximum learning rate of 1×10^{-3}. Instead of taking the last epoch model for fine-tuning, we opted for the one that exhibited superior performance during validation.

2.4 Model fine-tuning

After the pre-training phase, a fine-tuning step is applied. In this case, the dataset consisted of spatially and temporally coregistered pairs of S1 images corresponding to the IW and SM modes, as described in Section 2.1. To ensure radiometric normalization between images from both modes, we aligned their histograms following Galar et al. (2020). Regarding the training setup, the only deviation from the one employed in the pre-training (Section 2.3) is the adjustment of the number of epochs, which was set to 30.

3 Results and discussion

The proposed models have been compared with the baseline model that enhances resolution through the application of a bicubic interpolation algorithm, as well as the model resulting from training end-to-end with the IW and SM pairs. The performance has been assessed using commonly used metrics in super-resolution and image restoration tasks, including the Root Mean Square Error (RMSE), the Structural Similarity Index Measure (SSIM), and the Peak Signal-to-Noise Ratio (PSNR) Dalsasso et al. (2020). All the results are presented in Table 1.

As it can be observed, although all the models except the baseline perform similarly in terms of SSIM, the model pretrained by injecting Rayleigh noise into S2 optical images outperforms the others in terms of RMSE and PSNR. This may indicate that the SSIM metric is not an adequate measure to be used in SAR super-resolution scenarios.

Experiment	↑ SSIM	↓ RMSE	↑ PSNR (dB)
Baseline (bicubic)	64.61	0.5711	28.45
End-to-end	75.33	0.3419	30.68
PT. S2 + Speckle	75.31	0.3236	30.92
PT. S2 + Rayleigh	75.33	**0.2608**	**31.86**
PT. S1 + Mean	**75.36**	0.3165	31.02
PT, S1 + Median	75.32	0.3254	30.90

Table 1: Comparison between different S1 super-resolution approaches.

The quality of the super-resolution has been assessed qualitatively in Figure 1, where it is compared with the baseline bicubic interpolation algorithm and the ground truth. Additionally, a reference aerial image has been included for validation purposes. Overall, the proposed model enhances the clarity and sharpness of the image while maintaining a natural appearance, without introducing synthetic artifacts. Consequently, the image looks more visually pleasing and becomes easier to interpret.

Acknowledgments

Funded by the European Union

Funded by the European Union under Grant Agreement N.101103622. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or DG-DEFIS. Neither the European Union nor the granting authority can be held responsible for them.

M. Galar was supported by the Spanish Ministry of Science and Innovation under projects PID2020-118014RB-I00 and PID2022-136627NB-I00 (MCIN/AEI/10.13039/501100011033/FEDER, UE) and by the Public University of Navarre under project PJUPNA2023-11377.

References

Dongyang Ao, Corneliu Octavian Dumitru, Gottfried Schwarz, and Mihai Datcu. Dialectical gan for sar image translation: From sentinel-1 to terrasar-x. *Remote Sensing*, 10(10):1597, 2018.

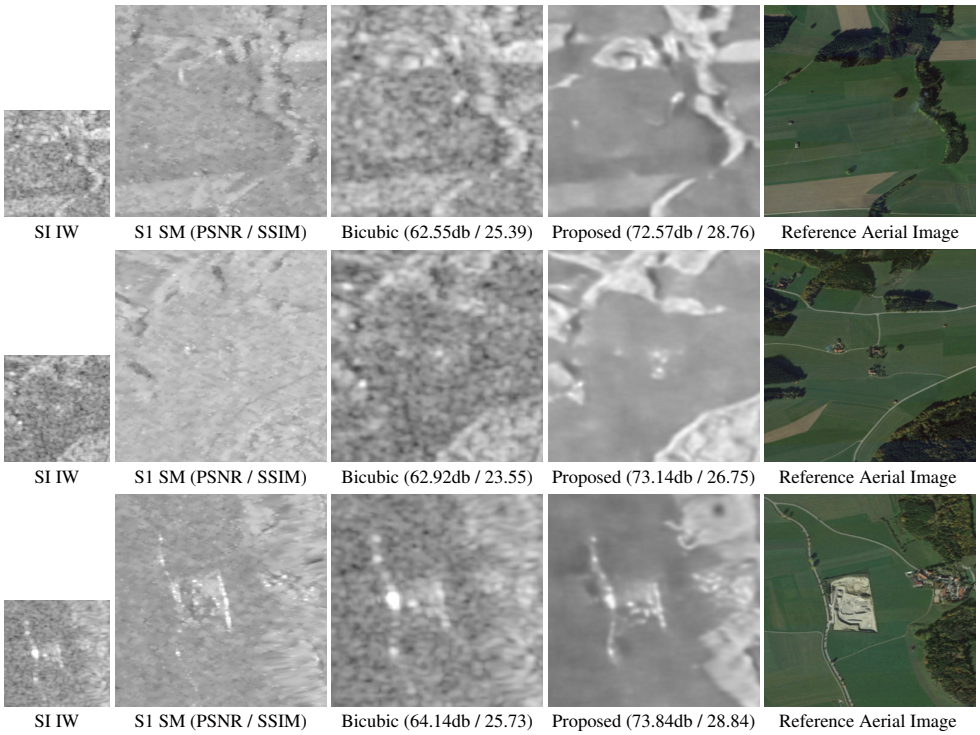

Figure 1: Visual comparison between the bicubic interpolation baseline and our proposal.

Christian Ayala, Rubén Sesma, Carlos Aranda, and Mikel Galar. A deep learning approach to an enhanced building footprint and road detection in high-resolution satellite imagery. *Remote Sensing*, 13(16), 2021.

Giovanni Chierchia, Davide Cozzolino, Giovanni Poggi, and Luisa Verdoliva. Sar image despeckling through convolutional neural networks. In *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, pp. 5438–5441. IEEE, 2017.

John C Curlander and Robert N McDonough. *Synthetic aperture radar*, volume 11. Wiley, New York, 1991.

Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 11065–11074, 2019.

Emanuele Dalsasso, Xiangli Yang, Loïc Denis, Florence Tupin, and Wen Yang. Sar image despeckling by deep neural networks: From a pre-trained model to an end-to-end training strategy. *Remote Sensing*, 12(16):2636, 2020.

Federico Filipponi. Sentinel-1 grd preprocessing workflow. In *International Electronic Conference on Remote Sensing*, pp. 11. MDPI, 2019.

Giulia Fracastoro, Enrico Magli, Giovanni Poggi, Giuseppe Scarpa, Diego Valsesia, and Luisa Verdoliva. Deep learning methods for synthetic aperture radar image despeckling: An overview of trends and perspectives. *IEEE Geoscience and Remote Sensing Magazine*, 9(2):29–51, 2021.

Mikel Galar, Rubén Sesma, Christian Ayala, Lourdes Albizua, and Carlos Aranda. Super-resolution of sentinel-2 images using convolutional neural networks and real ground truth data. *Remote Sensing*, 12(18):2941, 2020.

Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, et al. Generative adversarial networks, 2014.

Francesco Lattari, Borja Gonzalez Leon, Francesco Asaro, Alessio Rucci, Claudio Prati, and Matteo Matteucci. Deep learning for sar image despeckling. *Remote Sensing*, 11(13):1532, 2019.

Jong-Sen Lee, L Jurkevich, Piet Dewaele, Patrick Wambacq, and André Oosterlinck. Speckle filtering of synthetic aperture radar images: A review. *Remote sensing reviews*, 8(4):313–340, 1994.

Prabhishek Singh, Manoj Diwakar, Achyut Shankar, Raj Shree, and Manoj Kumar. A review on sar image and its despeckling. *Archives of Computational Methods in Engineering*, 28(7):4633–4653, Dec 2021.

Leslie N. Smith and Nicholay Topin. Super-convergence: very fast training of neural networks using large learning rates. In *Defense + Commercial Sensing*, 2018.

Ramón Torres, Paul Snoeij, Malcolm Davidson, et al. The sentinel-1 mission and its application capabilities. In *2012 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1703–1706, 2012.

Longgang Wang, Mana Zheng, Wenbo Du, Menglin Wei, and Lianlin Li. Super-resolution sar image reconstruction via generative adversarial network. In *2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE)*, pp. 1–4. IEEE, 2018.

Puyang Wang, He Zhang, and Vishal M Patel. Sar image despeckling using a convolutional neural network. *IEEE Signal Processing Letters*, 24(12):1763–1767, 2017.

Ce Zheng, Xue Jiang, Ye Zhang, Xingzhao Liu, Bin Yuan, and Zhixin Li. Self-normalizing generative adversarial network for super-resolution reconstruction of sar images. In *IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium*, pp. 1911–1914. IEEE, 2019.