# SELF-SUPERVISED LEARNING OF MULTI-MODAL COOPERATION FOR SAR DESPECKLING

Victor Gaya, Emanuele Dalsasso, Loïc Denis, Florence Tupin, Béatrice Pinel-Puysségur, Cyrielle Guérin

# SELF-SUPERVISED LEARNING OF MULTI-MODAL COOPERATION FOR SAR DESPECKLING

*Victor Gaya*[1,2]*, Emanuele Dalsasso*[4*]*, Loïc Denis*[3]*, Florence Tupin*[2]*,*
*Béatrice Pinel-Puysségur*[1]*, Cyrielle Guérin*[1]

[1]CEA, DAM, DIF, F-91297 Arpajon, France
[2]LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France
[3]UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School,
Laboratoire Hubert Curien UMR 5516, F-42023,SAINT-ETIENNE, France,
[4]CÉDRIC, Conservatoire National des Arts et Métiers, Paris, France

## ABSTRACT

Synthetic aperture radar (SAR) is a widely used modality for Earth observation, as they provide weather-independent imaging capabilities. However, interpretation of SAR images is difficult due to the speckle phenomenon: fluctuations appear in the image, which are stronger in areas with high radar reflectivity. As a result, many speckle reduction methods have been developed, with deep learning approaches standing out as particularly effective. Our article presents here a deep learning approach with two novel features: the use of an optical image to improve the restoration of a SAR image, while using a self-supervised neural network training.

***Index Terms*—** SAR, remote sensing, self-supervised, deep learning, multi-modal.

## 1. INTRODUCTION

SAR satellite imaging is widely used for various applications in mapping, environmental monitoring, and defense. However, SAR images present high intensity fluctuations limiting their exploitation due to the speckle phenomenon. Speckle results from constructive and destructive interferences of coherent electromagnetic waves backscattered within each resolution cell. Recent years have witnessed the emergence of numerous neural network-based approaches for speckle reduction. These encompass supervised, semi-supervised, and self-supervised methods, with the latter removing the necessity of providing pairs of speckled/speckle-free images. This article builds on the MERLIN approach introduced in [1], leveraging the independence of real and imaginary components for self-supervised network training. Notably, this approach has been extended in [2] to accommodate input from other dates acquired on the same orbit. The incorporation of an optical sensor image, explored in [3] using a patch-based filtering method, presents a potential avenue for further enhancing despeckling based on deep learning: this is the path followed in this paper.

## 2. SELF-SUPERVISED DESPECKLING WITH MERLIN

This section summarizes the self-supervised MERLIN approach [1], that we extend in the following section to a multi-modality context. The methodology relies on decomposing single-look complex SAR images into real and imaginary components, to exploit their statistical independence for self-supervised learning, as detailed in [1, 4].

Complex speckle in the Goodman model [5] is modeled by a random variable $s$ which follows a complex circular Gaussian distribution with an identity covariance matrix. The complex amplitude $z = a + jb$ over an area with reflectivity $r$ can be modeled by $z = s\sqrt{r}$. As a result, the distribution of $z$ is determined by:

$$\begin{aligned} \mathrm{p}_Z(z|r) = \frac{1}{\pi r}\exp\left(-\frac{|z|^2}{r}\right) &= \underbrace{\frac{1}{\pi r}\exp\left(-\frac{a^2+b^2}{r}\right)}_{\mathrm{p}(a,b|r)} \\ &= \underbrace{\frac{1}{\sqrt{\pi r}}\exp\left(-\frac{a^2}{r}\right)}_{\mathrm{p}(a|r)} \cdot \underbrace{\frac{1}{\sqrt{\pi r}}\exp\left(-\frac{b^2}{r}\right)}_{\mathrm{p}(b|r)}, \end{aligned} \quad (1)$$

which shows the statistical independence of the real and imaginary parts of the complex amplitudes.

This insight paves the way for a self-supervised neural network training for speckle reduction. In this approach, the network $f_{\boldsymbol{\theta}}$, parameterized[1] by $\boldsymbol{\theta}$, takes one component (e.g., the real part $\mathbf{a}$) as input and evaluates the restored reflectivity image quality ($\tilde{\mathbf{r}} = f_{\boldsymbol{\theta}}(\mathbf{a})$) based on the other component (e.g., the imaginary part $\mathbf{b}$) using the following loss function:

*The author performed the work while at [4]

[1]throughout the text, bold font denotes vectors and images

$$\mathcal{L}_{\text{MERLIN}}(\tilde{\mathbf{r}}, \mathbf{b}) = \sum_k -\log \mathrm{p}(b_k|\tilde{r}_k)$$
$$= \sum_k \frac{1}{2}\log\left(\tilde{r}_k\right) + \frac{b_k^2}{\tilde{r}_k}, \qquad (2)$$

The index $k$ in the equation represents the $k$-th pixel of either the estimated reflectivity image $\tilde{\mathbf{r}}$ or the imaginary part $\mathbf{b}$. In practical implementation, owing to their independence, the real and imaginary parts undergo permutation at each iteration during training. In the inference phase, the trained network is individually applied to the real and imaginary parts of the test image. The two intermediate estimations, $f_\theta(\mathbf{a})$ and $f_\theta(\mathbf{b})$, are then averaged to derive the final reflectivity estimation.

For the task of despeckling, it only necessitates a radar image. The MERLIN approach can be extended by incorporating auxiliary information to further enhance the speckle reduction, as illustrated in [2], where additional dates from the same orbit are used in a multi-temporal framework.

## 3. FUSE-MERLIN: INTRODUCTING OPTICAL DATA TO ENHANCE SAR IMAGE RESTORATION

Our proposal involves expanding the MERLIN framework to include optical data of the same scene. The objective is to collectively harness various sources of information pertaining to the same scene. When data obtained from another remote sensing modality is accessible, it can offer supplementary information to radar data, thereby assisting in SAR image despeckling, as highlighted in [3]. In this article, our focus is specifically on the use of optical imagery in this context.

Our method is based on the same architecture as MERLIN, but we use a different encoder for each modality (Fig. 1. The goal is to perform an Intermediate Fusion by combining multi-modal information at the latent space of the network. In contrast to an Early Fusion, raw images are not concatenated at input and processed jointly by a common encoder; but rather their latent representations are merged. The network comprises two independent branches designed to extract complementary descriptors from the two modalities. Concatenating these descriptors results in an enriched latent space that includes information from the auxiliary modality. The common branch of the network then reconstructs the denoised image based on this combined latent space.

While exploring fusion strategies, our investigation also considered the Early Fusion approach. Both fusion approaches lead to comparable results. Yet, the Intermediate Fusion model is more interesting because it offers more flexibility. With an encoder for each modality, it allows for greater adaptability when considering more additional modalities.

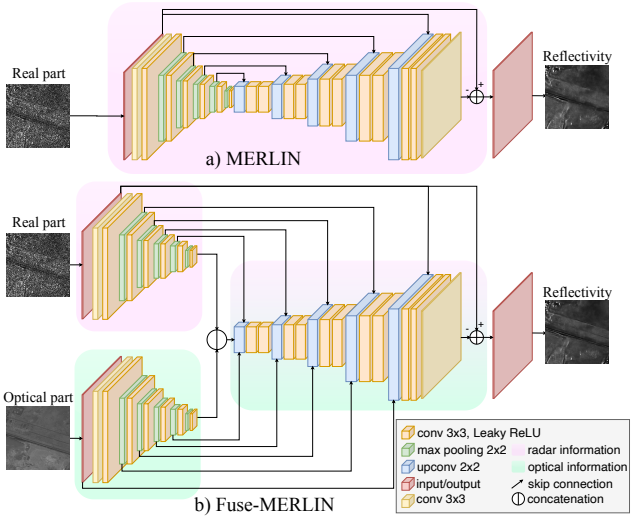


**Fig. 1**: Architectures of MERLIN, and Fuse-MERLIN.

## 4. EXPERIMENTS

### 4.1. Data Preparation

Self-supervised training offers the possibility of direct training on the image of interest (yet, a pre-trained network can also be applied). A single image for training is sufficient given the network's capacity for generalization, and the U-Net architecture is particularly well-suited for training with limited data, thanks to its incorporation of residual connections.

Nine interferometric radar images have been captured between 16/03/2019 and 30/10/2020 by the TerraSAR-X (TSX) satellite in Stripmap mode during a descending orbit. From the nine images, only one has been used for the training process, all the remaining images were used for test purposes. The accompanying optical image is a stereo panchromatic image acquired on 22/03/2019 by the Pleiades (PHR) satellite. The scene, situated in Jeddah, Saudi Arabia, encompasses both rural and urban areas, including desert and harbor. Four PHR images acquired between March and April 2019 underwent orthorectification and a Digital Surface Model (DSM) was generated through photogrammetry. Both steps were accomplished using the MICMAC software (IGN). Subsequently, the orthorectified image acquired on 22/03/2019 was projected into radar geometry using the previously computed DSM.

Given the sensor's geolocation accuracy, a refinement step for registration becomes essential. The OSCAR algorithm, as detailed in [6], was employed for image registration. This algorithm uses the optical image and the DSM to simulate radar images, which are then correlated with the radar amplitude image to estimate a geometric transformation involving rotation, scaling, and translation. The optical image was registered to the radar image with sub-pixel accuracy, effectively reducing the initial offset.

### 4.2. Implementation Details

The MERLIN framework is general and does not restrict the architecture of the neural network that can be used. In our experiments, we adopt the same U-Net network variant as used in MERLIN [1] in order to ease the comparison.

In contrast to MERLIN (Fig. 1a), in our method the network includes not only the real (or imaginary) part of the complex SAR image, but also the optical image through two independent encoders, (as illustrated in Fig. 1b). Each encoder separately produces its feature maps from the input image. The outputs of these encoders are concatenated in the latent space, which is then utilized by a common decoder. This common decoder incorporates residual connections from the two encoders, facilitating the fusion of intermediate representations of two images at different resolution scales, in line with the U-Net architecture principle.

Both MERLIN and Fuse-MERLIN underwent training using batches of 12 patches, each of size $256 \times 256$, for 30 epochs. The stride used was 32, the $9000 \times 4300$ pixels image was decomposed into 34810 partially overlapping patches.

### 4.3. Results

The validation of the proposed method involves training the network with the TSX image and the PHR optical image registered to the radar image.

Quantitative evaluation of the method would be valuable to compare MERLIN with Fuse-MERLIN. As nine SAR images are available, an arithmetic temporal mean of the intensity has been computed in order to decrease the speckle noise. However, in this desertic region, the mean intensity appeared to be still highly prone to speckle. Thus, it could not be used as a proxy for image reflectivity. In the following, only qualitative results are shown.

The results underscore the significance of incorporating a registered optical image for training. A key attribute of the model is its capacity to better preserve radar information in the reconstructed image, ensuring faithful representation of the original radar data.

Firstly, the networks were run on the first radar acquisition. Comparison between MERLIN and Fuse-MERLIN are shown on three areas in Fig. 2. For the two first areas, it is clearly visible that Fuse-MERLIN restores some details pointed by red arrows that MERLIN did not succeed to restore (circles in the first area and a thin road in the second area). One could wonder if Fuse-MERLIN adds some information derived from the optical image but absent from the SAR image. In the third area located in the harbour, there is a ship pointed by a blue arrow in the optical image that is missing in the radar image. In this case, the results of MERLIN and Fuse-MERLIN are equivalent, showing that the latter network does not add any information specific to the optical image only (no cross-modality contamination).

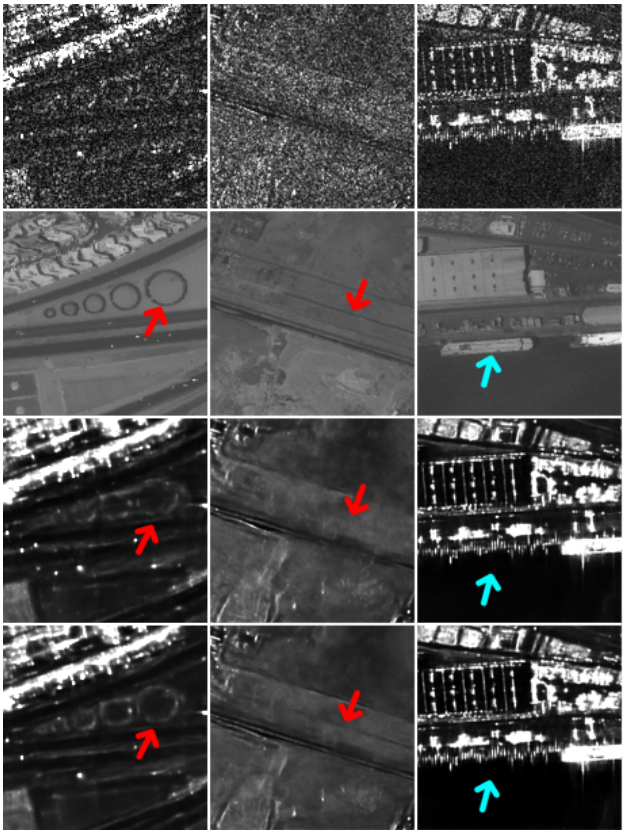

**Fig. 2**: Comparison of MERLIN and Fuse-MERLIN results. First row: original noisy amplitude image. Second row: optical image. Third row: square root of the reflectivity estimated by MERLIN. Fourth row: square root of the reflectivity estimated by Fuse-MERLIN. Each column corresponds to a different area. Red arrows show improvements of Fuse-MERLIN over MERLIN, while blue arrows indicate optical local information that was not spread into the radar image.

Secondly, MERLIN and Fuse-MERLIN were tested over the whole SAR temporal series. The square root of the temporal mean of the reflectivities estimated by each method is shown on Fig. 3 on another area. It clearly appears that for the whole time series, MERLIN does not succeed to restore the fine contrasts in the road details (subdivided in two parts) that were correctly restored by Fuse-MERLIN. It can be observed that the details of the road that were revealed by Fuse-MERLIN appear on the square root of the temporal mean of the intensities. These details are not artificial information purely driven by the optical data as they are present in the square root of the temporal mean of the radar intensities.

Tests on other acquisitions from different sensors confirmed a qualitative improvement, particularly on thin and/or low-contrasted structures. Preliminary tests also demonstrated the network's robustness to a small optical-radar misregistration of 2 pixels.

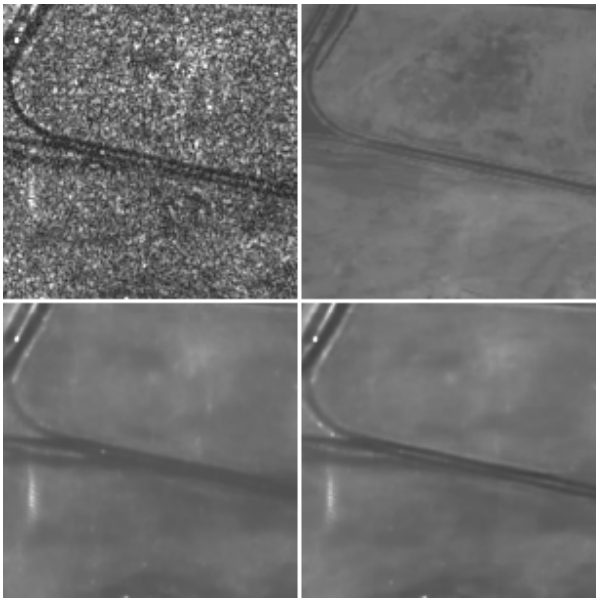

**Fig. 3**: Top left : square root of the mean of nine noisy intensities. Top right : optical image. Bottom left and bottom right respectively : square root of the mean of nine reflectivities estimated by MERLIN and Fuse-MERLIN.

## 5. CONCLUSION

The integration of optical data within the Fuse-MERLIN model has revealed notable improvements in the quality of denoised SAR images compared to the original MERLIN approach. The ability to leverage complementary information from different modalities, particularly from registered optical imagery, has demonstrated promising results in better restoring finer details and enhancing the overall accuracy of speckle reduction in SAR images.

One significant observation is the capacity of Fuse-MERLIN to preserve radar-specific information while benefiting from auxiliary data. This aspect emphasizes the model's ability to use multi-modal inputs effectively without introducing artifacts or bias derived solely from the auxiliary data. It suggests a balanced integration that enhances rather than distorts the inherent radar information.

The initial tests presented in this study showcase the promising potential of the Fuse-MERLIN approach by enhancing results in specific areas compared to MERLIN. However, it is imperative to subject this approach to a more extensive evaluation using a diverse set of radar and optical data or other modalities, characterized by varying resolutions and acquired by different sensors. Additionally, the study sites should be diversified to ensure the generalizability of the proposed method.

Currently, no quantitative validation of the algorithm's performance has been conducted. A perspective for future work involves the calculation of validation metrics to quantitatively compare the performance of different MERLIN variants and alternative despeckling algorithms. It is acknowledged that one challenge associated with unsupervised algorithms is the absence of ground truth for comparison, as highlighted in [7]. Therefore, obtaining an image for which speckle has been perfectly filtered, especially through multi-temporal filtering of a long series of acquisitions, would be essential for a robust validation process.

## 6. REFERENCES


[1] E. Dalsasso, L. Denis, and F. Tupin, "As if by magic: self-supervised training of deep despeckling networks with MERLIN," *IEEE TGRS*, vol. 60, pp. 1–13, 2021.

[2] I. Meraoumia, E. Dalsasso, L. Denis, R. Abergel, and F. Tupin, "Multi-temporal speckle reduction with self-supervised deep neural networks," *IEEE TGRS*, vol. 61, pp. 1–14, 2023.

[3] Sergio Vitale, Davide Cozzolino, Giuseppe Scarpa, Luisa Verdoliva, and Giovanni Poggi, "Guided patchwise nonlocal sar despeckling," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, pp. 6484–6498, 2018.

[4] E. Dalsasso, L. Denis, M. Muzeau, and F. Tupin, "Self-supervised training strategies for SAR image despeckling with deep neural networks," in *EUSAR*, 2022.

[5] J. W. Goodman, *Speckle phenomena in optics: theory and applications.*, Roberts and Company Publishers, 2007.

[6] B. Pinel-Puysségur, L. Maggiolo, M. Roux, N. Gasnier, D. Solarna, G. Moser, S. B. Serpico, and F. Tupin, "Experimental comparison of registration methods for multi-sensor SAR-optical data," in *2021 IEEE IGARSS*, 2021, pp. 3022–3025.

[7] A. B. Molini, D. Valsesia, G. Fracastoro, and E. Magli, "Speckle2void: Deep self-supervised SAR despeckling with blind-spot convolutional neural networks," *IEEE TGRS*, 2021.