

# Débruitage multi-modal d'images radar à synthèse d'ouverture par apprentissage profond auto-supervisé

Victor GAYA<sup>1,2</sup> Emanuele DALSSASSO<sup>4</sup> Loïc DENIS<sup>3,1</sup> Florence TUPIN<sup>1</sup> Béatrice PINEL-PUYSSÉGUR<sup>2</sup> Cyrielle GUÉRIN<sup>2</sup>

<sup>1</sup>LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France

<sup>2</sup>CEA, DAM, DIF, F-91297 Arpajon, France

<sup>3</sup>UJM-Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France

<sup>4</sup>CÉDRIC, Conservatoire National des Arts et Métiers, Paris, France

**Résumé** – L'observation de la Terre a été largement facilitée depuis de nombreuses années grâce à l'utilisation des satellites imageurs radar à synthèse d'ouverture (SAR), qui offrent des capacités d'imagerie indépendantes des conditions météorologiques. Toutefois, l'interprétation de ces images SAR est complexe en raison de la présence de bruit inhérent à l'imagerie cohérente. En effet, des fluctuations apparaissent dans les images, notamment là où la réflectivité radar est élevée. Ainsi, de nombreuses méthodes ont été développées pour réduire le bruit présent dans ces images, notamment des méthodes neuronales particulièrement efficaces. Dans cet article, nous proposons d'étudier comment l'ajout d'une donnée auxiliaire comme une image optique peut améliorer la restauration de la réflectivité dans ce cadre.

**Abstract** – Earth observation has been greatly facilitated for many years by the use of synthetic aperture radar (SAR) imaging satellites, which provide imaging capabilities independent of weather conditions. However, the interpretation of these SAR images is complex due to the presence of inherent noise. Indeed, fluctuations appear in the images, especially where the radar reflectivity is high. Thus, many methods have been developed to reduce the noise present in these images, including particularly efficient neural methods. In this paper we propose to study how the addition of an auxiliary data such as an optical image can improve the reflectivity restoration in this framework.

## 1 Introduction

L'imagerie satellitaire radar à synthèse d'ouverture (SAR) est devenue une technologie importante pour la surveillance de la surface de la Terre. Le SAR est une technique de télédétection qui utilise des ondes électromagnétiques centimétriques à métriques émises vers la surface terrestre. Les ondes rétrodiffusées par les objets de la surface sont ensuite enregistrées par le radar. En combinant les données enregistrées de multiples émissions, une antenne virtuelle (synthétique) est créée, ce qui permet d'obtenir une résolution spatiale élevée en utilisant les propriétés de phase et d'amplitude de la rétrodiffusion. L'intérêt de cette technologie est principalement le fait de n'être dépendante ni de la météo, ni du cycle diurne. Les images SAR sont par conséquent utilisées dans diverses applications telles que la cartographie, la surveillance environnementale et la défense. Cependant, ces images sont souvent soumises à des niveaux de bruit importants, qui peuvent affecter leur qualité et leur précision. Le speckle est le bruit inhérent à l'imagerie SAR car il est directement lié à la physique de l'acquisition. Il est dû aux interférences constructives et destructives des ondes électromagnétiques cohérentes réfléchies par la surface terrestre à l'intérieur d'une cellule de résolution.

De nombreuses approches de réduction de speckle à base de réseaux de neurones ont été proposées dans les dernières années. Il existe des approches supervisées, semi-supervisées [1] ou auto-supervisées [5], ces dernières permettant de s'affranchir de bases de données d'images de référence. Dans

cet article nous nous appuyons en particulier sur l'approche MERLIN proposée dans [2] qui exploite l'indépendance des parties réelles et imaginaires pour réaliser l'entraînement auto-supervisé du réseau. Elle a été étendue dans [4] pour prendre en compte d'autres dates en entrée, acquises sur la même orbite. Dans cet article, nous proposons d'étudier comment la prise en compte d'une image issue d'un capteur optique pourrait améliorer l'estimation de la réflectivité.

## 2 MERLIN

Nous rappelons dans cette partie l'approche auto-supervisée MERLIN [2] que nous allons étendre au contexte multi-modalité. Elle repose sur la séparation de la donnée complexe en partie réelle et partie imaginaire : celles-ci étant statistiquement indépendantes, les conditions idéales pour un apprentissage auto-supervisé sont vérifiées [1].

Le modèle de Goodman [3] décrit le speckle complexe comme une variable aléatoire  $s$  distribuée selon une loi circulaire gaussienne complexe de matrice de covariance identité. Dans la suite, les images sont notées en caractères gras. Le signal  $z = a + jb$  mesuré sur une zone de réflectivité  $R$  est défini comme  $z = s\sqrt{R}$ . Ainsi, la distribution de  $z$  est donnée par :

$$P_Z(z) = \frac{1}{\pi R} \exp\left(-\frac{|z|^2}{R}\right) = \frac{1}{\pi R} \exp\left(-\frac{a^2 + b^2}{R}\right). \quad (1)$$

Sur la base du modèle de speckle complexe, [2] a montré qu’une image SAR complexe  $z$  peut être décomposée en chaque pixel en deux composantes indépendantes et identiquement distribuées  $a$  et  $b \sim \mathcal{N}(0, R/2)$ , chacune contenant la moitié de l’information d’une image d’intensité ( $I = a^2 + b^2$ ). En effet, le signal que l’on cherche à reconstruire sera le même entre parties réelle et imaginaire, seul le speckle sera différent. Cela implique qu’un réseau neuronal peut être entraîné de manière auto-supervisée pour la réduction du speckle comme suit : le réseau  $f_\theta$  paramétré par  $\theta$  prend en entrée une composante (p.ex. la partie réelle  $a$ ) et évalue la qualité de l’image de réflectivité restaurée ( $\tilde{R} = f_\theta(a)$ ) par rapport à l’autre composante (p.ex. la partie imaginaire  $b$ ). Pour ce faire, on maximise la log-vraisemblance de  $b$  par rapport à la sortie du réseau  $\tilde{R}$ , ce qui revient à minimiser la fonction de coût suivante :

$$\begin{aligned} \mathcal{L}_{\text{MERLIN}}(\tilde{R}, b) &= \sum_k -\log p(b_k | \tilde{R}_k) \\ &= \sum_k \frac{1}{2} \log \left( \tilde{R}_k \right) + \frac{b_k^2}{\tilde{R}_k}, \end{aligned} \quad (2)$$

où l’indice  $k$  indique le  $k$ -ième pixel de l’image de réflectivité estimée  $\tilde{R}$  ou de la partie imaginaire  $b$ . En pratique, en raison de leur indépendance, les parties réelle et imaginaire sont permutées à chaque itération pendant l’entraînement. Au moment de l’inférence, le réseau entraîné est appliqué séparément sur la partie réelle et imaginaire de l’image de test : les deux estimations intermédiaires  $f_\theta(a)$  et  $f_\theta(b)$  sont ensuite moyennées pour obtenir l’estimation finale de la réflectivité.

Cette méthode est non seulement très performante, mais aussi extrêmement flexible : en effet, pour apprendre la tâche de débruitage, elle ne nécessite qu’une image radar. Par conséquent, l’approche MERLIN peut naturellement être étendue afin d’incorporer une information auxiliaire pour améliorer la réduction du speckle, comme dans [4] où d’autres données prises sur la même orbite sont utilisées.

### 3 Fuse-MERLIN : introduction d’une donnée optique pour améliorer la restauration d’images SAR

Nous proposons d’étendre le formalisme de MERLIN pour prendre en compte une donnée optique disponible sur la scène. L’objectif est d’exploiter conjointement les différentes sources d’information sur une même scène : lorsqu’une donnée acquise avec une autre modalité de télédétection est disponible, celle-ci peut apporter une information complémentaire au radar, facilitant le débruitage de l’image SAR [7]. Dans cet article, nous considérons le cas de l’imagerie optique.

#### 3.1 Stratégies de fusion

La fusion amont (voir Fig. 1b) consiste à concaténer les images en entrée du réseau, afin que le réseau puisse commencer à construire ses cartes de caractéristiques en utilisant les informations contenues dans les deux modalités.

La fusion intermédiaire (voir Fig. 1c) combine l’information multi-modale au niveau de l’espace latent du réseau. À la différence de la fusion amont, ce ne sont pas les images brutes qui sont concaténées, mais leur représentation latente. Le réseau

sera donc constitué de deux branches indépendantes visant à extraire de deux modalités des descripteurs complémentaires. La concaténation de ces descripteurs permet la création d’un espace latent plus riche grâce à l’information extraite de la modalité auxiliaire. C’est en partant de cet espace latent que la branche commune du réseau reconstruira l’image débruitée.

Dans la suite nous noterons Fuse-MERLIN early (FME) et Fuse-MERLIN intermediate (FMI), respectivement les versions du modèle en fusion amont et en fusion intermédiaire.

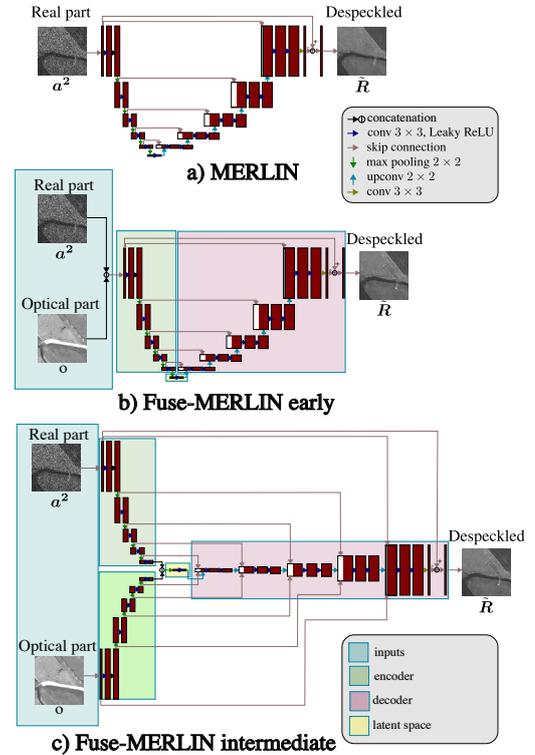


FIGURE 1 : Architectures de MERLIN, FME et FMI.

#### 3.2 Pré-traitement des données

La nécessité d’un pré-traitement des données optiques en amont du réseau de débruitage multi-modal se pose. En effet, on souhaite que le réseau soit insensible aux changements de radiométrie de l’image optique dans ses cartes de caractéristiques, pour ne pas biaiser la reconstruction de l’image SAR. Plusieurs approches ont été étudiées dès l’entraînement pour imposer l’indépendance à la radiométrie optique : modification de l’image optique par normalisation ou égalisation d’histogramme ; filtrage de l’image optique pour détecter les contours (p.ex. filtre de Sobel). Ces tests ayant donné des résultats similaires aux versions utilisant directement l’image optique, seule cette dernière sera utilisée en entrée des réseaux par la suite.

## 4 Expériences

#### 4.1 Préparation des données

L’entraînement auto-supervisé offre l’avantage de s’entraîner directement sur l’image à débruiter. L’utilisation d’une seule image pour l’entraînement est justifiée grâce à la capacité du réseau à généraliser. L’architecture UNet est adaptée à

un entraînement avec peu de données grâce à l'utilisation de connexions résiduelles.

L'image radar utilisée dans cette étude a été acquise le 10/07/2022 par le satellite TerraSAR-X (TSX) en mode Spotlight sur une orbite descendante. L'image optique est une image panchromatique stéréo acquise le 08/07/2022 par le satellite Pléiades Néo (PNEO). La scène, située à Brétigny-sur-Orge, France, est constituée de zones rurales et urbaines, ainsi que de pistes d'aéroport. L'image PNEO a été orthorectifiée et un MNS (Modèle Numérique de Surface) a été calculé par photogrammétrie. Les deux traitements ont été générés par la chaîne MICMAC (IGN). L'image orthorectifiée a été projetée en géométrie radar en utilisant le MNS calculé précédemment.

La précision de géolocalisation des capteurs nécessite une étape d'affinage du recalage. L'algorithme OSCAR décrit dans [6] a été utilisé pour recalrer les images. Cet algorithme utilise l'image optique et le MNS afin de simuler des images radar, qui sont alors corrélées à l'image d'amplitude radar, permettant d'estimer une transformation géométrique (rotation, homothétie et translation). L'image optique a été recalée sur l'image radar à une précision de l'ordre du pixel alors que le décalage initial était supérieur à 200 pixels.

## 4.2 Détails de l'implémentation

Le cadre MERLIN est agnostique à l'architecture du réseau neuronal convolutif. Nous considérons dans nos expériences la même variante du réseau U-Net utilisée dans MERLIN [2].

Dans le cas de la fusion amont (Fig.1b), l'image radar et optique recalées sont concaténées pour former une image ayant deux canaux. A la différence de MERLIN (Fig.1a), le réseau prendra donc en entrée non seulement la partie réelle (ou imaginaire) de l'image SAR complexe, mais aussi l'image optique. Par conséquent, seule la première couche est modifiée.

Dans la méthode de fusion intermédiaire (Fig.1c), deux encodeurs indépendants sont considérés. Chaque encodeur va indépendamment apprendre ses cartes de caractéristiques sur l'image qu'il a en entrée. La sortie de ces encodeurs est concaténée dans l'espace latent, pour ensuite utiliser un décodeur commun. Ce décodeur commun utilisera également les connexions résiduelles venant de deux encodeurs. Ceci permet de fusionner les représentations intermédiaires de deux images à des échelles de résolution différentes, respectant le principe du réseau U-Net.

Les réseaux MERLIN, FME et FMI ont tous été entraînés avec 8364 patches divisés en batches de 12 imagerie de taille  $256 \times 256$  pendant 30 époques avec un stride de 32, ce qui correspond à une image de  $3503 \times 2873$  pixels.

## 4.3 Résultats

La validation de la méthode proposée est réalisée en entraînant le réseau avec l'image TSX et l'image optique PNEO recalée sur l'image radar. Les résultats obtenus soulignent l'intérêt d'ajouter une image optique recalée pour l'entraînement. Tout d'abord, l'une des principales caractéristiques du modèle est sa capacité à préserver les informations du radar dans l'image reconstruite. C'est d'ailleurs primordial car cela permet de garantir que les données radar originales sont bien représentées dans l'image reconstruite. Toutes ces zones d'intérêt sont encadrées en rouge sur la Fig.2.

La Fig.2 compare les résultats de MERLIN à ceux de FME et FMI sur trois zones. Sur les deux premières, il est intéressant de voir que deux routes fines ont été beaucoup mieux reconstruites avec Fuse-MERLIN qu'avec MERLIN (encadrées en rouge) : le modèle est donc capable d'exploiter l'information auxiliaire pour mieux restaurer des détails très fins présents dans les images radar.

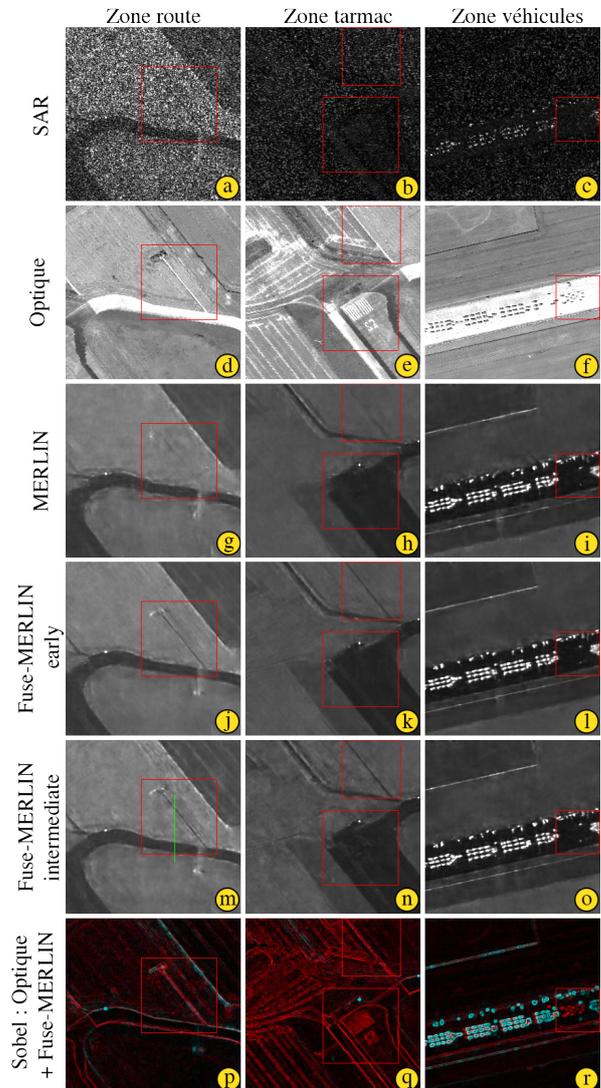


FIGURE 2 : Comparaison des résultats sur trois zones de Brétigny. Les différentes lignes présentent l'image SAR originale, l'image optique, les images débruitées par MERLIN, FME et FMI, la composition colorée de deux gradients.

Afin de vérifier la stabilité des reconstructions, nous avons effectué 9 entraînements des modèles MERLIN et FMI avec une initialisation des poids différente. Pour chaque variante, nous avons moyenné ces 9 images et calculé l'écart-type, puis comparé les résultats sur un profil croisant une petite et une grande route (Fig.3) et représenté en vert sur la Fig.2m. On voit bien le gain de FMI par rapport à MERLIN sur la petite route (entre les bandes verticales pointillées rouges) car elle se dégage plus nettement des zones environnantes. Grâce à l'écart-type pour les deux variantes, on constate que ce gain est stable. Cependant, sur la piste d'aéroport qui est plus large (pointillés noirs), les deux variantes ont des résultats équivalents. C'est donc sur les structures fines que FMI améliore les résultats de MERLIN.

On pourrait craindre que Fuse-MERLIN n'introduise sur l'image radar filtrée de fausses informations, spécifiques à l'image optique et absentes dans l'image radar. Les deux dernières zones contiennent de telles informations : sur la deuxième, des marquages blancs au sol sur l'image optique absents de l'image radar ; sur la troisième, certains véhicules présents lors de l'acquisition optique et qui étaient absents lors de l'acquisition radar et inversement. On constate sur les Fig.2(k,n,l,o), encadrés en rouge, que le réseau débruite l'image SAR sans faire apparaître les structures spécifiques à l'optique. Pour s'assurer de l'absence de biais dû à l'image optique, les gradients ont été calculés par le filtre de Sobel et combinés en une composition colorée RVB contenant dans le canal rouge le gradient de l'image optique et dans les canaux vert et bleu (soit du cyan) celui de l'image filtrée par FMI (dernière ligne de la Fig.2). On constate que FMI n'a pas rajouté d'informations spécifiques à l'optique car elles apparaissent en rouge sur la composition colorée. Les informations spécifiques à l'image radar (véhicules présents sur l'image radar et non sur l'optique) apparaissent bien en cyan.

Des tests ont été effectués sur d'autres acquisitions issues d'autres capteurs : une amélioration visuelle a également été constatée sur de fines structures sombres. Des tests préliminaires ont également montré la robustesse du réseau à un recalage optique radar précis à 2 pixels.

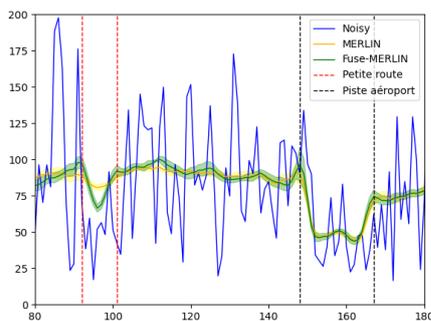


FIGURE 3 : Profils de l'image bruitée (bleu) et de la moyenne des images débruitées sur la ligne verte Fig.2m, obtenus par 9 simulations avec MERLIN (orange) et FMI (vert). L'écart-type est représenté par une bande colorée autour de la moyenne.

## 5 Discussion et perspectives

### 5.1 Discussion

L'intégration d'une information auxiliaire sous la forme d'une image optique au sein des modèles Fuse-MERLIN a mis en lumière une amélioration de la qualité de l'image restaurée par rapport à l'approche MERLIN. Au vu des résultats obtenus, on ne peut pour l'instant démontrer la supériorité de l'une ou l'autre des deux versions de Fuse-MERLIN. Cependant, le modèle FMI à double encodeur est plus intéressant que le modèle FME à unique encodeur pour plusieurs raisons. Tout d'abord il offre plus de flexibilité. En effet, avec un encodeur pour chaque modalité, il permet d'être plus adaptable pour l'ajout de modalités supplémentaires, ce qui pourra encore améliorer les résultats du débruitage. De plus, il sera également possible de fusionner les modalités de manière plus aisée, car chaque encodeur peut apprendre des caractéristiques

spécifiques de chaque modalité, indépendamment les unes des autres, et la fusion des caractéristiques se fera en aval de chaque encodeur. La capacité de généralisation du modèle à de nouvelles données est un point important pour continuer à améliorer le débruitage.

### 5.2 Perspectives

Les premiers tests présentés ici montrent que l'approche Fuse-MERLIN est prometteuse car elle permet d'améliorer sur certaines zones les résultats obtenus par MERLIN. Il convient cependant de la tester sur un plus grand nombre de données radar et optiques ou d'autres modalités, de différentes résolutions et acquises par différents capteurs. Il faudra également diversifier les sites d'étude.

Pour l'instant, aucune validation quantitative des performances de l'algorithme n'a été effectuée. Une perspective sera de calculer des métriques de validation permettant de comparer quantitativement les performances des différentes variantes de MERLIN ainsi que d'autres algorithmes de débruitage. En effet, un inconvénient des algorithmes non supervisés est que l'on ne dispose pas de vérité terrain à laquelle comparer la sortie de l'algorithme [5]. Ainsi, il faudrait disposer d'une image pour laquelle le speckle aurait été parfaitement filtré, en s'aidant notamment du filtrage multi-temporel d'une série d'acquisitions.

**Remerciements** Ce projet a été partiellement financé par l'ANR et la DGA dans le cadre du projet ASTRAL ANR-21-ASTR-0011.

## Références

- [1] E. DALSASSO, L. DENIS, M. MUZEAU et F. TUPIN : Self-supervised training strategies for SAR image despeckling with deep neural networks. *In EUSAR*, 2022.
- [2] E. DALSASSO, L. DENIS et F. TUPIN : As if by magic : self-supervised training of deep despeckling networks with MERLIN. *IEEE TGRS*, 60:1–13, 2021.
- [3] J. W. GOODMAN : *Speckle phenomena in optics : theory and applications*. Roberts and Company Publishers, 2007.
- [4] I. MERAOUMIA, E. DALSASSO, L. DENIS, R. ABERGEL et F. TUPIN : Multi-temporal speckle reduction with self-supervised deep neural networks. *IEEE TGRS*, 61:1–14, 2023.
- [5] A. B. MOLINI, D. VALSESIA, G. FRACASTORO et E. MAGLI : Speckle2void : Deep self-supervised SAR despeckling with blind-spot convolutional neural networks. *IEEE TGRS*, 2021.
- [6] B. PINEL-PUYSSÉGUR, L. MAGGIOLO, M. ROUX, N. GASNIER, D. SOLARNA, G. MOSER, S. B. SERPICO et F. TUPIN : Experimental comparison of registration methods for multisensor SAR-optical data. *In 2021 IEEE IGARSS*, pages 3022–3025, 2021.
- [7] S. VITALE, D. COZZOLINO, G. SCARPA, L. VERDOLIVA et G. POGGI : Guided Patchwise Nonlocal SAR Despeckling. *IEEE TGRS*, 57(9):6484–6498, 2019.