

VAE pour la détection d'anomalies en imagerie SAR

X. Huy NGUYEN^{1,2} Chengfang REN¹ Jérémy FIX³ Jean-Philippe OVARLEZ^{1,2} Joana FRONTERA-PONS²

¹SONDRA, CentraleSupélec, Université Paris-Saclay 91190 Gif-sur-Yvette, France

²DEMR, ONERA, Université Paris-Saclay, 91123 Palaiseau, France

³LORIA, CentraleSupélec, Université Paris-Saclay, 57000 Metz, France

Résumé – Dans ce papier, nous proposons une approche non supervisée pour la détection d'anomalies en imagerie SAR. Le modèle proposé combine une étape de *despeckling* en pré-traitement, un Auto-Encodeur Variationnel (VAE) de type β -annealing pour filtrer les anomalies de manière non supervisé et un détecteur d'anomalies basé sur le changement de la matrice de covariance en entrée et en sortie du réseau de β -annealing VAE. Des expériences ont été menées sur des images polarimétriques SAR de l'ONERA en bande X pour démontrer l'efficacité du Beta-Annealing VAE par rapport aux méthodes proposées dans la littérature.

Abstract – In this paper, we propose an unsupervised learning approach for anomaly detection in SAR imaging. The proposed model combines a preprocessing despeckling step, a β -annealing Variational Auto-Encoder (VAE) for unsupervised anomaly filtering, and an anomaly detector based on the change of the covariance matrix at the input and output of the β -annealing VAE network. Experiments have been carried out on X-band ONERA polarimetric SAR images to demonstrate the effectiveness of Beta-Annealing VAE compared with the methods proposed in the literature.

1 Introduction

La détection d'anomalies en télédétection est un défi majeur pour la surveillance de sites, la détection d'incendies de forêt, l'identification de cibles cachées et le suivi des changements environnementaux, etc. Peu de travaux en apprentissage profond ont été menés pour la détection d'anomalies en imagerie SAR [1, 2, 3, 4] bien qu'elle ait été étudiée dans de nombreux domaines comme l'imagerie médicale [5], l'imagerie hyperspectrale [6]. L'exploitation des données SAR s'est fortement accrue suite à la mise en orbite de nombreux satellites tels que TerraSAR-X et Sentinel. Cependant, le manque d'annotation d'anomalies en SAR demeure un problème crucial pour entraîner les algorithmes supervisés. De plus, les anomalies sont en général très rares, ce qui rend un entraînement supervisé inadapté. Dans ce contexte, l'emploi d'algorithmes non supervisés est généralement préféré tel que le détecteur de Reed-Xiaoli [7].

Nous abordons le problème de détection d'anomalies dans l'imagerie SAR en nous inspirant de la méthodologie proposée dans [3] qui consiste à quantifier une anomalie par l'écart des caractéristiques d'une zone à la distribution normale. L'objectif est de créer une cartographie des anomalies potentielles, pixel par pixel, à partir d'une image SAR polarimétrique. La première étape consiste à filtrer les fluctuations de *speckle* présentes dans l'image SAR afin de réduire les fausses détections. Ensuite, un *beta-annealing* VAE est mis en place pour apprendre la distribution de l'image SAR sans anomalie. Enfin, nous mesurons la différence entre l'image reconstruite et l'image *despecklée* afin de détecter les pixels hors-distributions.

2 Schéma de détection d'anomalies

Le *speckle* est naturellement présent dans les images SAR. Pour former une image, un système radar émet des impulsions électromagnétiques et collecte les signaux rétrodiffusés. Ces signaux dépendent de plusieurs facteurs, notamment la géométrie de la scène et les propriétés diélectriques des matériaux, qui peuvent absorber une partie de l'énergie du signal, le réfléchir ou le rétrodiffuser vers le capteur. Lors de la rétrodiffusion, les signaux interagissent entre eux, créant ainsi un phénomène de marche aléatoire. Ces interactions créent des variations d'amplitude et de phase qui génèrent de nombreuses fausses alertes, compliquant la détection d'anomalies. Il est donc crucial d'effectuer une étape de *despeckling*.

2.1 Despeckling

Contrairement à [3], nous adoptons la stratégie proposée par [8] pour filtrer les fluctuations de *speckle*. L'algorithme MERLIN consiste à minimiser la distance entre la partie réelle et la partie imaginaire des signaux à valeurs complexes. Le processus d'entraînement est auto-supervisé et ne requiert aucune donnée annotée, car la composante imaginaire sert de référence pour l'apprentissage de la composante réelle et vice versa. MERLIN présente de meilleures performances par rapport aux approches [9, 10, 11, 12] et comparable à SAR2SAR [13]. De plus, le processus d'entraînement ne requiert pas une série temporelle d'images SAR.

Pour traiter des données polarimétriques, MERLIN fonctionnant en mode mono-canal, nous l'appliquons successivement à chacun des canaux de l'image. Ces derniers peuvent être combinés ensemble pour former une image SAR polarimétrique *despecklée*, dénotée \mathcal{X} .

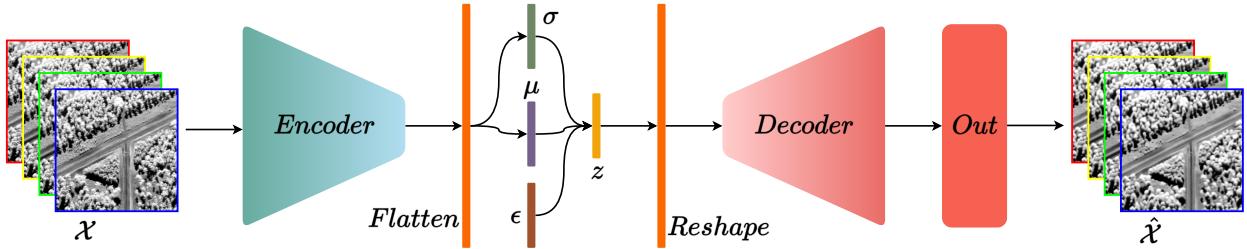


FIGURE 1 : Architecture d'un Auto-Encodeur Variationnel. \mathcal{X} et $\hat{\mathcal{X}}$ dénotent l'image SAR *despecklée* et reconstruite par le VAE, respectivement.

2.2 β -annealing VAE

Les Auto-Encodeurs Variationnels VAE font partie de la famille des réseaux génératifs basés sur l'apprentissage profond. Introduits dans [14], ils visent à approcher la distribution des données à l'aide d'une approche bayésienne variationnelle et peuvent se reformuler de manière équivalente comme un auto-encodeur régularisé. La fonction de perte à minimiser est alors composée d'une perte de reconstruction et d'un terme de régularisation. L'architecture globale du VAE est illustrée à la figure 1.

2.2.1 Architecture

Encodeur. L'encodeur comprime spatialement l'image SAR filtrée par MERLIN $\mathcal{F}_0 \triangleq \mathcal{X}$ avec une séquence de N opérations de sous-échantillonnage :

$$\mathcal{F}_i = \text{Pool}(E_i(\mathcal{F}_{i-1})), \forall i \in [|1; N|], \quad (1)$$

avec $\mathcal{F}_i \in \mathbb{R}^{h_i, w_i, c_i}$ la carte des caractéristiques issue de chaque bloc d'encodage, Pool l'opération MaxPooling avec un facteur 2 et E_i une opération formée de deux blocs successifs chacun composé d'une couche de convolution suivie d'une non-linéarité LeakyReLU. Des couches linéaires en sortie du réseau permettent de calculer les paramètres du modèle de la loi *a posteriori* $p(\mathbf{z}|\mathcal{X})$.

La reparamétrisation. Les VAEs modélisent la loi *a posteriori* $p(\mathbf{z}|\mathcal{X})$ par la distribution donnée en sortie de l'encodeur $q(\mathbf{z}|\mathcal{X})$ souvent supposée gaussienne, c-à-d $\mathbf{z}|\mathcal{X} \sim \mathcal{N}(\boldsymbol{\mu}(\mathcal{X}), \boldsymbol{\Sigma}(\mathcal{X}))$. Le processus d'apprentissage nécessiterait d'échantillonner cette loi mais cette échantillonnage induit une fonction de perte non différentiable en les paramètres de l'encodeur. Pour résoudre ce problème, [14] considère une reparamétrisation de la loi en séparant l'aléa des paramètres de cette loi, c'est-à-dire en échantillonnant $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$ et en effectuant un changement de variable $\mathbf{z} = \boldsymbol{\mu} + \boldsymbol{\sigma} \odot \epsilon$ conduisant à $\mathbf{z} \sim \mathcal{N}(\boldsymbol{\mu}, \text{diag}(\boldsymbol{\sigma}^2))$ où $\boldsymbol{\mu}$ et $\boldsymbol{\sigma}$ sont deux vecteurs estimés représentatifs de la moyenne et de l'écart-type des composantes de $\mathbf{z}|\mathcal{X}$, et où l'opérateur \odot^2 représente le carré de chaque composante du vecteur.

Décodeur. Le décodeur essaie de reconstruire l'image originale à partir de l'espace latent échantillonné, en passant séquentiellement $\mathcal{G}_0 \triangleq \mathbf{z}$ par une séquence de blocs de sur-échantillonnage :

$$G_i = D_i(\text{Upsample}(\mathcal{G}_{i-1})), \forall i \in [|1; N-1|], \quad (2)$$

Upsample caractérisant l'opération de sur-échantillonnage par un facteur 2 suivie par un bloc de décodage D_i ayant la même composition que E_i .

Opération de sortie. En sortie du décodeur, \mathcal{G}_{N-1} doit être surdimensionnée encore une fois afin de retrouver la taille de l'image originale. Cela est traduit par :

$$\mathcal{G}_N = \text{LeakyReLU}(\text{Conv}(\text{Upsample}(\mathcal{G}_{N-1}))). \quad (3)$$

Finalement, l'image reconstruite est obtenue par $\hat{\mathcal{X}} = \text{Sigmoid}(\text{Conv}(\text{Out}(\mathcal{G}_N)))$, Sigmoid étant la fonction d'activation finale et Conv l'opération de convolution.

2.2.2 Fonctions d'optimisation

Le formalisme du VAE [14, Eqn. 3] conduit à minimiser la fonction de perte ELBO (*Evidence Lower BOund*) :

$$\text{ELBO} = \mathcal{L}_{rec} - D_{KL}(q(\mathbf{z}|\mathcal{X})\|p(\mathbf{z})). \quad (4)$$

où $L_{rec} = \mathbb{E}_{q(\mathbf{z}|\mathcal{X})} [\log(p(\mathcal{X}|\mathbf{z}))]$ avec $p(\mathcal{X}|\mathbf{z})$ la distribution générative. Cette borne à minimiser comprend un terme de reconstruction et un terme de régularisation. Dans le cas d'un modèle gaussien, la divergence de Kullback-Leibler (KL) s'écrit sous la forme :

$$D_{KL}(q(\mathbf{z}|\mathcal{X})\|p(\mathbf{z})) = -\frac{1}{2} \sum_{j=1}^J (1 + \log(\sigma_j^2) - \mu_j^2 - \sigma_j^2), \quad (5)$$

où μ_j et σ_j sont les composantes des vecteurs moyenne et écart-type $\boldsymbol{\mu}$ et $\boldsymbol{\sigma}$ respectivement.

Nous considérons ici une extension du VAE, le β -annealing VAE, dans lequel l'importance du terme de régularisation est pondérée par le coefficient β qui évolue de manière cyclique durant l'apprentissage. La nouvelle fonction de perte \mathcal{L}_{VAE} à minimiser devient :

$$\mathcal{L}_{VAE} = \mathcal{L}_{rec} + \beta D_{KL}, \quad (6)$$

Dans le cas d'une loi génératrice normale, la perte de reconstruction s'exprime comme une perte L2. Dans ce papier, on explore également une perte de reconstruction originale (au sens du formalisme du VAE) en considérant une perte L1 (interprétant comme une loi génératrice de Laplace) qui s'avère plus robuste (moins sensible aux valeurs aberrantes) que la perte L2. La perte L1 permet de relâcher une contrainte de la perte L2 qui force le réseau à trop fidèlement reconstruire les anomalies (qui apparaissent comme des points très lumineux).

2.3 Détection de changements

Nous terminons le processus de détection d'anomalies en comparant les images reconstruites par le VAE avec les images *despecklées*. Les images SAR, ayant une forte dynamique spatiale

et spectrale, suscitent donc un calcul des statistiques de manière locale. Pour chaque pixel $k \in [0; h-1], l \in [0; w-1]$, nous définissons un *boxcar* $\mathcal{B}_{k,l}$, entourant et centrés sur le pixel k, l . La norme de Frobenius mesure ensuite entre la différence de variation autour de chaque pixel de l'image reconstruite et de l'image *despecklée* :

$$A_{k,l} = \left\| \hat{\Sigma}_{k,l}^{\mathcal{X}} - \hat{\Sigma}_{k,l}^{\hat{\mathcal{X}}} \right\|_F^2, \quad (7)$$

où $\hat{\Sigma}_{k,l}^{\mathcal{X}}$ est la matrice de covariance des vecteurs des canaux polarimétrique du bloc $\mathcal{B}_{k,l}$. Pour chaque bloc $\mathcal{B}_{k,l}$ de l'image déspecklée \mathcal{X} et identiquement pour l'image reconstruite $\hat{\mathcal{X}}$, elle est calculée comme suit :

$$\hat{\Sigma}_{k,l}^{\mathcal{X}} = \frac{1}{|\mathcal{B}_{k,l}|} \sum_{i,j \in \mathcal{B}_{k,l}} (\mathcal{X}_{i,j} - \hat{\mu}_{k,l}^{\mathcal{X}}) (\mathcal{X}_{i,j} - \hat{\mu}_{k,l}^{\mathcal{X}})^T, \quad (8)$$

et $\hat{\mu}_{k,l}^{\mathcal{X}} = \frac{1}{|\mathcal{B}_{k,l}|} \sum_{i,j \in \mathcal{B}_{k,l}} \mathcal{X}_{i,j}$ où $|\mathcal{B}_{k,l}|$ est le cardinal du *boxcar* $\mathcal{B}_{k,l}$ et où $\hat{\mu}_{k,l}^{\mathcal{X}}$ est sa moyenne empirique.

2.4 Détails d'implémentation

L'encodeur du VAE est constitué de $N = 5$ blocs d'encodage ayant chacun une taille de canaux égale respectivement à 16, 32, 64, 128, 128. Le nombre de canaux de chaque bloc du décodeur est l'inverse. Les couches de convolutions opèrent avec un noyau de taille 3×3 avec un pas de 1×1 . Les LeakyReLU ont tous une pente de 0.01. La taille de l'espace latent \mathbf{z} est fixée à 256.

Pour la mise en œuvre, nous regroupons les patches pour former un lot de taille 128 et utilisons l'optimiseur Adam avec un pas d'apprentissage de 0.0005. Le réseau est entraîné sur 500 époques. L'estimation de la matrice de covariance pour le calcul de la norme de Frobenius a été effectuée avec un *boxcar* de taille 11×11 .

Nous adoptons ainsi la stratégie de [15] pour définir β . Pendant les 5 premières époques de l'entraînement, β reste à 0. La variation cyclique commence à partir de la sixième époque, avec au total 4 cycles. Pour chaque cycle, β augmente linéairement entre 0 et 1, et reste à 1 uniquement sur 20% du temps à la fin du cycle.

3 Résultats expérimentaux

3.1 Données

Notre VAE est entraîné sur une image SAR polarimétrique (quad-pol) en bande X de dimensions spatial 4800×30000 , acquise par le capteur aéroporté SETHI de l'ONERA. Nous utilisons 80% de l'image pour l'entraînement, et réservons 20% restants pour les tests. Afin de réduire la dynamique de l'image, nous éliminons les 5 centiles des valeurs les plus élevées et les plus faibles. Durant la phase d'apprentissage, l'image SAR est découpée en plusieurs patches de taille 256×256 avec un pas d'avancement de 100 pixels. Chaque patch est soumis à une transformation logarithmique et est normalisé entre 0 et 1 après quantification. Les anomalies n'étant pas annotées au préalable, l'entraînement est donc non-supervisé.

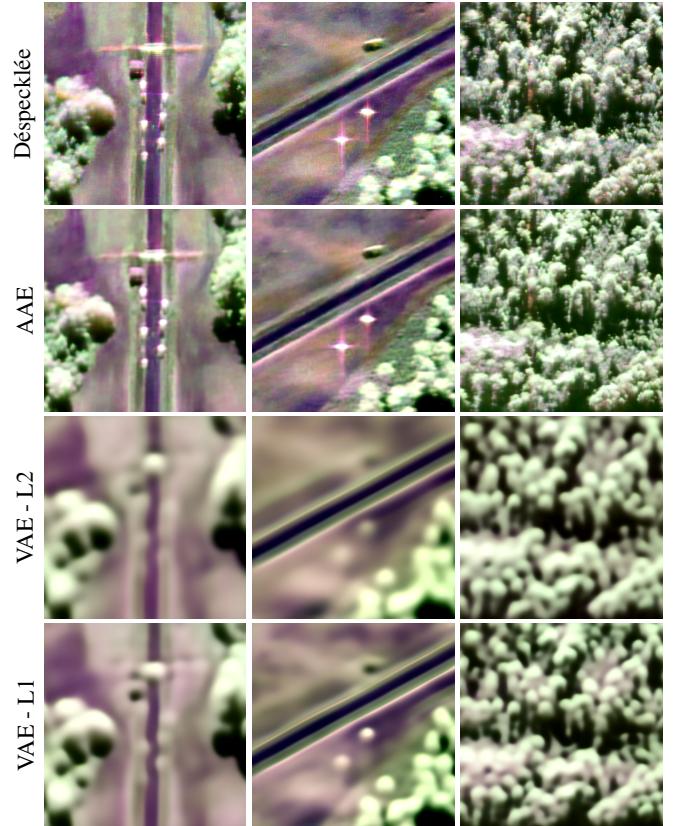


FIGURE 2 : Comparaison entre AAE, VAE entraîné avec la fonction *loss* L2 et L1

Métriques	AAE	VAE - L2	VAE - L1
PSNR	33.19	31.46	32.41
SSIM	0.866	0.886	0.892

TABLE 1 : Comparaison quantitative des trois réseaux.

3.2 Évaluations et performances

Nous comparons l'approche AAE proposée dans [3] et l'approche VAE avec deux fonctions d'optimisation de reconstruction différentes L2 et L1 sur trois zones différentes, extraites de l'image de test. On observe une route avec des véhicules et une barrière dans l'image de gauche. Sur l'image du milieu, on trouve trois véhicules stationnant en dehors de la route. L'image de droite montre une zone forestière.

On remarque que les points brillants ont été fortement atténus par les VAEs par rapport à l'AAE. Cependant, les images reconstruites par les Auto-Encodeur Variationnels sont plus flous. L'utilisation de la fonction d'optimisation L1 améliore certes la qualité des images reconstruites, mais elle reste inférieure à celle des images générées par l'Auto-Encodeur antagoniste. Les pixels les plus énergétiques sont moins affaiblis dans le cas du VAE avec la L1 contrairement à notre hypothèse initiale. Nous utilisons les métriques PSNR et SSIM [16] pour analyser quantitativement chaque approche. Les résultats sont présentés dans le tableau 1.

La figure 3 présente les cartes d'anomalies, dont les 2% des valeurs les plus élevées ont été coupées. Visuellement, les réseaux AAE et VAE génèrent moins de fausses alarmes que le détecteur Reed-Xiaoli. Cependant, ce dernier isole mieux les forts échos, car il est calculé directement sur une image SAR *Single Look Complex*. On observe une différence significa-

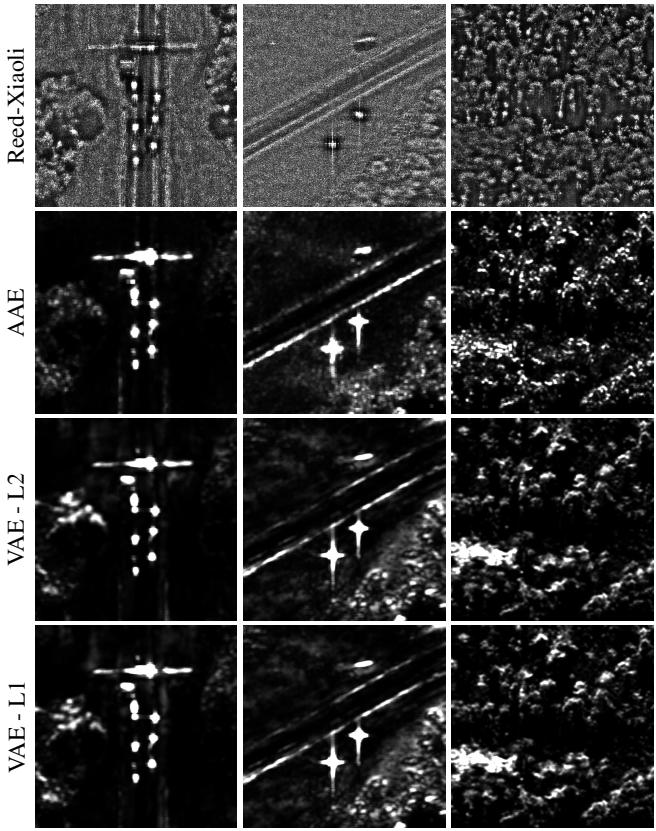


FIGURE 3 : Comparaison de la carte des anomalies obtenue par différentes méthodes de détection sur les patches décrits dans la figure 2 : le détecteur Reed-Xiaoli [7], la détecteur de changement entre l'image reconstruite et l'image déspecklée par notre VAE et par le réseau AAE proposé dans [3].

tive dans les zones arborées des cartes d'anomalies produites par les réseaux AAE et VAE, due à la meilleure qualité de reconstruction de l'AAE. Nous avons fixé la taille du *boxcar* à 11×11 pour permettre une comparaison équitable des cartes. Étant donnée la différence entre l'image reconstruite et l'image déspecklée, un *boxcar* de taille différente pourrait produire une carte avec des anomalies mieux définies.

4 Conclusion

Dans cet article, nous avons étudié l'application des Auto-Encodeurs Variationnels dans le schéma de détection d'anomalies proposé dans [3]. Visuellement, les cibles à fortes valeurs ont été fortement atténus. Les images générées par notre VAE ont certes une qualité inférieure, pourtant le calcul de la carte des anomalies avec la norme Frobenius n'a pas pas été fortement affecté. Une prochaine étape consiste à analyser l'utilisation des réseaux neuronaux à valeurs complexes afin d'exploiter l'information de la phase des images SAR polarimétriques.

Références

- [1] S. SINHA et al. “Variational autoencoder anomaly-detection of avalanche deposits in satellite SAR imagery”. In : *Proceedings of the 10th International Conference on Climate Informatics*. 2020, p. 113-119.
- [2] S. MABU, S. HIRATA et T. KUREMOTO. “Anomaly detection using convolutional adversarial autoencoder and one-class SVM for landslide area detection from synthetic aperture radar images”. In : *Journal of Robotics, Networking and Artificial Life* 8.2 (2021), p. 139-144.
- [3] M. MUZEAU et al. “Self-Supervised Learning Based Anomaly Detection in Synthetic Aperture Radar Imaging”. In : *IEEE Open Journal of Signal Processing* 3 (2022), p. 440-449.
- [4] M. MUZEAU et al. “Self-supervised SAR anomaly detection guided with RX detector”. In : *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*. IEEE. 2023, p. 1918-1921.
- [5] C. BAUR et al. “Autoencoders for unsupervised anomaly segmentation in brain MR images : A comparative study”. In : *Medical Image Analysis* 69 (2021), p. 101952. ISSN : 1361-8415.
- [6] T. JIANG et al. “Discriminative reconstruction constrained generative adversarial network for hyperspectral anomaly detection”. In : *IEEE Transactions on Geoscience and Remote Sensing* 58.7 (2020).
- [7] Xiaoli Y. et I. S. REED. “Adaptive detection of signals with linear feature mappings and representations”. In : *IEEE Transactions on Signal Processing* 43.12 (1995), p. 2953-2963.
- [8] E. DALSASSO, L. DENIS et F. TUPIN. “As if by magic : Self-supervised training of deep despeckling networks with MERLIN”. In : *IEEE Transactions on Geoscience and Remote Sensing* 60 (2021), p. 1-13.
- [9] C.-A. DELEDALLE et al. “MuLoG, or how to apply Gaussian denoisers to multi-channel SAR speckle reduction ?” In : *IEEE Transactions on Image Processing* 26.9 (2017), p. 4389-4403.
- [10] G. CHIERCHIA et al. “SAR image despeckling through convolutional neural networks”. In : *2017 IEEE international Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE. 2017, p. 5438-5441.
- [11] S. PARRILLI et al. “A nonlocal SAR image denoising algorithm based on LLMMSE wavelet shrinkage”. In : *IEEE Transactions on Geoscience and Remote Sensing* 50.2 (2011), p. 606-616.
- [12] C.-A. DELEDALLE et al. “NL-SAR : A unified nonlocal framework for resolution-preserving (Pol)(In) SAR denoising”. In : *IEEE Transactions on Geoscience and Remote Sensing* 53.4 (2014), p. 2021-2038.
- [13] E. DALSASSO, L. DENIS et F. TUPIN. “SAR2SAR : A semi-supervised despeckling algorithm for SAR images”. In : *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14 (2021), p. 4321-4329.
- [14] D. P. KINGMA et M. WELLING. “Auto-Encoding Variational Bayes”. In : *2nd International Conference on Learning Representations, ICLR 2014, Canada, Conference Track Proceedings*.
- [15] H. FU et al. “Cyclical annealing schedule : A simple approach to mitigating KL vanishing”. In : *arXiv :1903.10145* (2019).

- [16] Z. WANG et al. "Image quality assessment : from error visibility to structural similarity". In : *IEEE transactions on image processing* 13.4 (2004), p. 600-612.