

# Reconstruction 3D en tomographie radar : apprentissage profond basé sur un Matching Pursuit déroulé

Cristiano ULONDU MENDES<sup>1</sup> Loïc DENIS<sup>2,1</sup> Christophe KERVAZO<sup>1</sup> Florence TUPIN<sup>1</sup>

<sup>1</sup>LTCI, Télécom Paris, Institut Polytechnique de Paris, 19 place Marguerite Perey, BP 00000, 91120 Palaiseau Cedex, France

<sup>2</sup>Université Jean Monnet Saint-Etienne, CNRS, Institut d'Optique Graduate School, Laboratoire Hubert Curien UMR 5516, F-42023, SAINT-ETIENNE, France

**Résumé** – La tomographie radar en milieu urbain consiste à séparer des réflecteurs situés à des hauteurs différentes mais vus dans un même pixel car situés à une distance similaire du radar. Les méthodes d'apprentissage profond proposées récemment pour résoudre cette tâche sont basées sur le déroulement d'algorithmes de poursuites de base avec contrainte de parcimonie. Ils dépendent d'une discrétisation des hauteurs et ne permettent pas un contrôle simple du nombre de réflecteurs détectés. On présente dans cet article une approche alternative permettant d'estimer la position des cibles sur un intervalle continu. Notre approche s'inspire des itérations des algorithmes gloutons de reconstruction parcimonieuse tels que Matching Pursuit ou RELAX. Nous montrons des résultats de reconstruction satisfaisants sur des données simulées et sur une pile d'images satellitaires.

**Abstract** – Radar tomography in urban areas consists of separating targets of different heights projected into the same pixel due to their similar distance to the radar. Deep-learning methods recently proposed to solve this task are based on the unrolling of basis pursuit algorithms with a sparsity constraint. They depend on a discretization of the heights and don't allow for easy control of the number of targets to be detected. In this paper, we present an alternative approach that estimates the position of targets over a continuous interval. Our approach is drawn from the iterations of greedy algorithms for sparse reconstruction, such as Matching Pursuit or RELAX. We show satisfactory reconstruction results on simulated data and on a stack of satellite images.

## 1 Introduction

L'imagerie radar à synthèse d'ouverture (SAR) exploite le principe de l'écholocatisation afin d'obtenir un rendu bidimensionnel de très haute résolution d'une scène tridimensionnelle, par tout temps. La tomographie SAR (tomoSAR) permet de recouvrer des informations sur la dimension spatiale manquante en exploitant plusieurs acquisitions de la même scène vue sous différents angles. En milieu urbain, cela consiste à reconstruire un nuage de réflecteurs (ou "cibles radar") épars selon cette troisième dimension (estimation de leur position et de leur réflectivité). En raison de la parcimonie spatiale rencontrée dans ce contexte, des approches basées sur l'acquisition comprimée ont été proposées pour résoudre le problème de reconstruction [13, 1]. Ces approches nécessitent cependant des temps de calcul difficilement acceptables en pratique pour des images de grande dimension. Cette limitation a été levée récemment grâce à la technique schéma déroulés [4] consistant à réécrire quelques itérations d'un algorithme d'optimisation sous la forme d'une architecture de réseaux de neurones. Différents algorithmes résolvant le problème de l'acquisition comprimée, dont VAMP [2], ISTA [7, 12, 11] et ADMM [9] ont été considérés. L'entraînement de ces réseaux de neurones repose souvent sur une stratégie supervisée comparant, pixel par pixel, un vecteur parcimonieux de référence, dont la longueur dépend de l'échantillonnage de la dimension à restaurer, avec la sortie du réseau, qui est un vecteur de même longueur. Avec ces méthodes, les positions des cibles sont donc contraintes par une grille prédéfinie qui ne correspond pas nécessairement à la position réelle des cibles (positions hors grille). De plus, ces méthodes peuvent souffrir de la duplication des cibles lorsqu'une cible se situe entre deux valeurs de la grille prédéfinie

et d'un manque de contrôle du nombre total de cibles extraites.

Nous proposons une approche d'apprentissage profond s'inspirant des algorithmes gloutons de type "Matching Pursuit" comme RELAX [5] qui permet de considérer des cibles sur un intervalle continu et de contrôler le nombre maximal de cibles à restaurer. Ainsi, les sorties de notre réseau profond correspondent directement aux réflecteurs recherchés et ne nécessitent pas d'étape de post-traitement afin d'extraire les cibles significatives.

## 2 État de l'art : méthodes gloutonnes d'inversion tomographique

Dans la suite, on soulignera les variables à valeurs complexes pour les différencier de celles à valeurs réelles et on notera en gras les variables vectorielles (lettres minuscules) et matricielles (lettres majuscules).

La tomoSAR exploite une pile de  $D$  images SAR à valeurs complexes, acquises suivant des trajectoires parallèles. Cette configuration permet d'acquérir des images de la même scène avec des angles de vue très légèrement différents. Soit  $(O, \vec{x}, \vec{y}, \vec{z})$  le repère orthogonal de l'espace tel que  $(O, \vec{x}, \vec{y})$  est le plan horizontal et l'axe  $(O, \vec{x})$  coïncide avec la direction de déplacement du radar, comme représenté dans la figure 1. Pour une position donnée du capteur sur l'axe  $(O, \vec{x})$ , chaque pixel selon l'axe  $(O, \vec{y})$  de l'image correspond à la sommation suivant l'axe  $(O, \vec{z})$  des échos de toutes les cibles à équidistance par rapport au capteur. Après discrétisation des intégrales et recalage des  $D$  images, le vecteur  $\underline{v}$  de longueur  $D$  des amplitudes complexes mesurées dans chaque image, pour un pixel

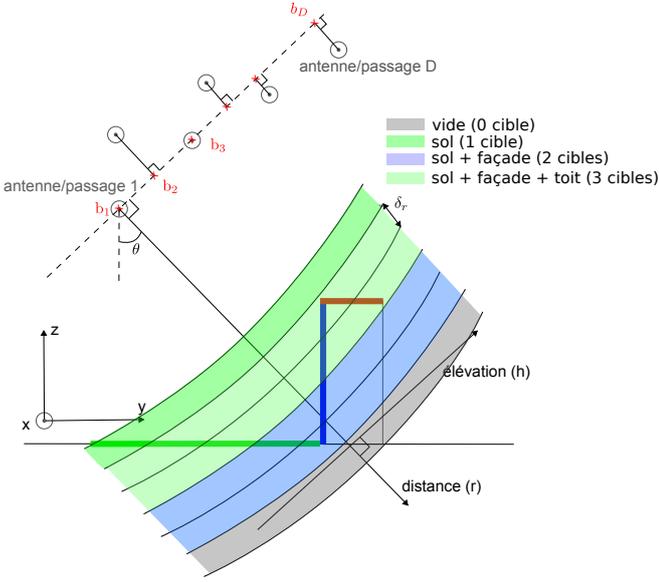


FIGURE 1 : Géométrie d'acquisition des données TomoSAR.

donné, s'écrit :

$$\underline{v} = \sum_{m=1}^K \underline{u}_m \underline{a}(z_m) + \underline{\epsilon} \quad (1)$$

où  $K$  est le nombre de cibles,  $\underline{a}(z_m)$  est le vecteur des amplitudes complexes vues sur les antennes pour une cible unitaire située à la hauteur  $z_m$  : l'amplitude sur la  $d$ -ième antenne vaut  $[\underline{a}(z_m)]_d = (\exp(-j\kappa_d z_m))$  avec  $\kappa_d = \frac{4\pi b_d}{\lambda r_0 \sin(\theta)}$  et enfin  $\underline{\epsilon}$  est un vecteur gaussien circulaire complexe de dimension  $D$  modélisant le bruit thermique. Concrètement, la TomoSAR consiste à estimer les couples  $(z_m, \underline{u}_m)$  des positions et réflectivités complexes de chacune des  $K$  cibles. Avec l'hypothèse faite sur la distribution du bruit thermique et sous réserve de connaître le nombre de cibles  $K$ , la méthode des moindres carrés non-linéaire correspond à l'estimateur du maximum de vraisemblance [10]. Elle consiste à résoudre le problème d'optimisation suivant :

$$\{\hat{z}, \hat{\underline{u}}\} = \arg \min_{\{z, \underline{u}\} \in \mathbb{R}^K \times \mathbb{C}^K} \left\| \underline{v} - \sum_{m=1}^K \underline{u}_m \underline{a}(z_m) \right\|^2. \quad (2)$$

En pratique  $K$  n'est pas connu et doit être estimé et la recherche conjointe des positions et échos des  $K$  cibles est computationnellement coûteuse. L'algorithme RELAX [5], rappelé dans l'algorithme 1, fournit une solution approchée au problème (2) en résolvant successivement des problèmes n'impliquant qu'une seule cible, la contribution des cibles précédemment estimée étant soustraite des mesures. Cette approche est moins coûteuse et offre plus de flexibilité quant au nombre de cibles restaurées sans nécessiter de post-traitement. Cette dernière caractéristique est fortement désirable dans les approches de reconstruction en tomoSAR.

### 3 Méthode proposée inspirée du déroulement de l'algorithme RELAX

Dans cet article, on propose de réaliser les itérations de RELAX par une chaîne de  $K_{\max}$  réseaux de neurones, chacun

#### Algorithme 1 : Algorithme RELAX

---

**Entrées :**  $\underline{v}$ ,  $\text{tol} \in \mathbb{R}^+$ ,  $K_{\max} \in \mathbb{N}^*$   
**Sorties :**  $\hat{z} \in \mathbb{R}^K$ ,  $\hat{\underline{u}} \in \mathbb{C}^K$ , ( $K \leq K_{\max}$ )

- 1  $k \leftarrow 1$  ;
- 2 **tant que**  $\left\| \underline{v} - \sum_{m \in \llbracket 1, k \rrbracket} \hat{\underline{u}}_m \underline{a}(\hat{z}_m) \right\|^2 > \text{tol}$  &
- $k < K_{\max}$  **faire**
- 3 **tant que Pas convergé faire**
- 4 **pour**  $m=k, \dots, I$  **faire**
- 5  $\underline{v}^{(m)} \leftarrow \underline{v} - \sum_{n \in \llbracket 1, k \rrbracket \setminus \{m\}} \hat{\underline{u}}_n \underline{a}(\hat{z}_n)$  ;
- 6  $\hat{z}_m \leftarrow \arg \max_z |\underline{a}(z)^H \underline{v}^{(m)}|^2$  ;
- 7  $\hat{\underline{u}}_m \leftarrow \frac{\underline{a}(\hat{z}_m)^H \underline{v}^{(m)}}{\underline{a}(\hat{z}_m)^H \underline{a}(\hat{z}_m)}$  ;
- 8 **fin**
- 9 **fin**
- 10  $k \leftarrow k + 1$
- 11 **fin**

---

étant chargé de prédire la position d'une cible, l'incertitude de cette position ainsi que la probabilité d'existence de la cible (permettant d'indiquer l'absence de cible). Le choix du nombre de cibles est donc automatique et n'est pas à fixer par l'utilisateur. Chaque réseau reçoit la concaténation des caractéristiques associées au signal initial et aux signaux résiduels après le retrait de la contribution des cibles détectées par les réseaux précédents (voir la figure 2 pour plus de détails). Cela permet de pallier la baisse progressive de la qualité de l'information dans les signaux résiduels. On peut noter sur cette figure que le point de départ du calcul des caractéristiques est le produit par  $\underline{A}(z)^H$ , où la matrice  $\underline{A}(z)$  correspond à une transformée de Fourier discrète ( $\underline{A}(z)$  est obtenue en concaténant les vecteurs  $\underline{a}(\hat{z}_n)$  pour des valeurs de hauteurs  $z_n$  régulièrement espacées). De cette façon, la méthode, une fois entraînée avec un nombre d'acquisitions  $D$  fixé, peut être potentiellement appliquée à des signaux  $\underline{v}$  de longueur variable. En phase d'inférence, les probabilités d'existence prédites par les réseaux serviront de condition d'arrêt et les incertitudes peuvent être utilisées pour exclure les cibles dont la position est trop incertaine.

### 4 Entraînement de la chaîne de réseaux de neurones

Comme dans la plupart des approches en TomoSAR basées sur le déroulement, on traite les pixels indépendamment. Les différents réseaux de la chaîne sont composés d'une succession de couches de convolution complexes (CConv) avec un noyau  $1 \times 1$  suivi d'une fonction non-linéaire de type ReLU adaptée aux nombres complexes [8]. Nous avons fixé le nombre maximal de cibles dans un pixel à  $K_{\max} = 3$ . La simulation des données d'entraînement couvre les cas les plus fréquents en milieu urbain illustrés en figure 1. Elle utilise les paramètres d'acquisition du capteur TerraSAR-X lors de la mission ayant servi à acquérir les données réelles que nous considérons en phase de validation de la méthode. La chaîne est entraînée sur ces données synthétiques en optimisant la fonction de perte (3) définie ci-dessous. Il s'agit d'une somme pondérée de deux termes où les poids  $\tau_1$  et  $\tau_2$  sont appris. Le premier

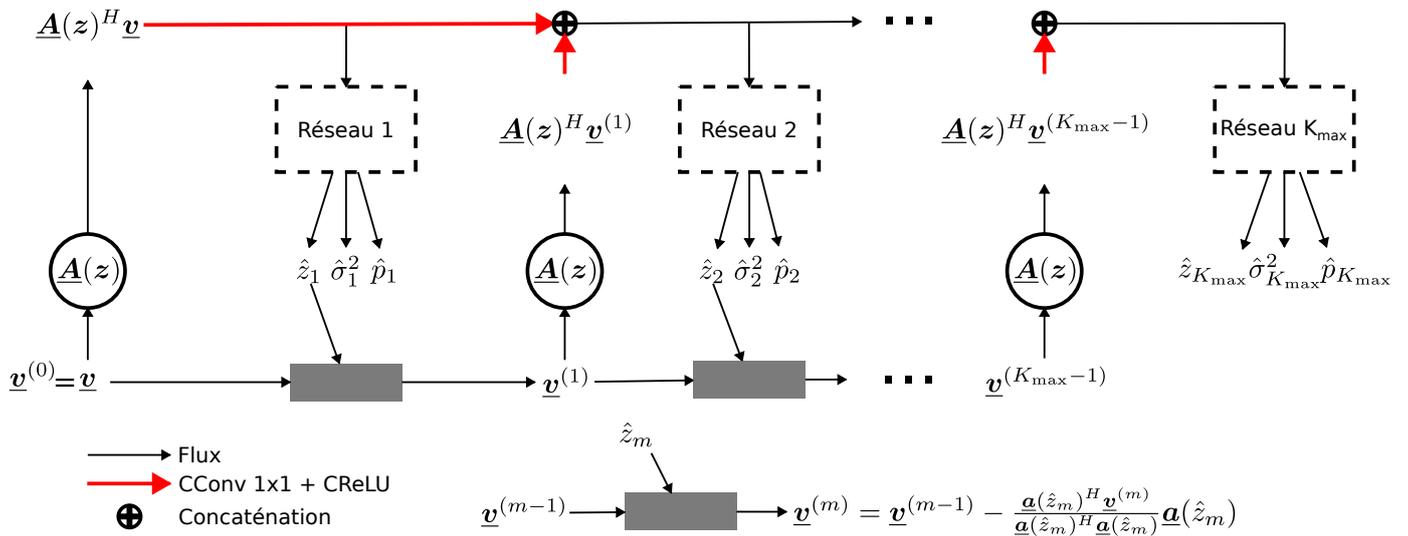


FIGURE 2 : Méthode proposée : chaîne de réseaux de neurones imitant les itérations de l’algorithme RELAX.

TABLE 1 : Paramètres de simulation pour l’entraînement

Paramètres	
longueur d’onde $\lambda$	0.031 m
distance $r_0$	$6.15e^5$ m
angle de vue $\theta$	0.6 rad
nombre d’acquisitions $D$	40
ouverture $\Delta b =  b_1 - b_D $	778 m
résolution de Rayleigh $\delta_z = \frac{\lambda r_0}{2\Delta b} \sin \theta$	8.3 m
amplitudes des échos	$\Gamma(5, 1)$
phase des échos	$\mathcal{U}(0, 2\pi)$
écart entre cibles ( $\delta_z$ )	$\{0.1, 0.2, \dots, \frac{z_{max} - z_{min}}{\delta_z}\}$
snr du signal	$[0, 10]$ dB

terme compare, via l’entropie-croisée binaire (ECB), la prédiction de présence d’une cible  $p_j$  donnée par le  $j$ -ième réseau à l’existence de cette cible (1 si  $j \leq K$ , 0 sinon). L’objectif est d’encourager les réseaux à ne pas déléguer la prédiction de la position d’une cible aux réseaux en aval. Le deuxième terme évalue la prédiction de la position des cibles  $z_i$  via un mélange gaussien dont les moyennes et écart-types correspondent aux positions prédites  $\hat{z}_j$  et aux incertitudes associées  $\hat{\sigma}_j$ . La fonction de perte comprend également un terme de régularisation (deuxième terme de l’équation (3)), proposé dans [6], pour éviter le délaissement d’une des deux tâches.

$$\mathcal{L} = \frac{1}{2\tau_1^2} \sum_{j=1}^{K_{max}} \text{ECB}(p_j, \mathbb{1}_{K-j>0}) + \log((1 + \tau_1^2)(1 + \tau_2^2)) - \frac{1}{2\tau_2^2} \sum_{i=1}^K \log \left[ \sum_{j=1}^{K_{max}} \frac{p_j}{\sum_{k=1}^{K_{max}} p_k} \frac{1}{\sqrt{2\pi\hat{\sigma}_j^2}} \exp \left( -\frac{(z_i - \hat{z}_j)^2}{2\hat{\sigma}_j^2} \right) \right]. \quad (3)$$

## 5 Validation de la méthode proposée

Dans cette section, on valide la méthode proposée sur une scène 3D simulée avec les paramètres d’acquisition du capteur

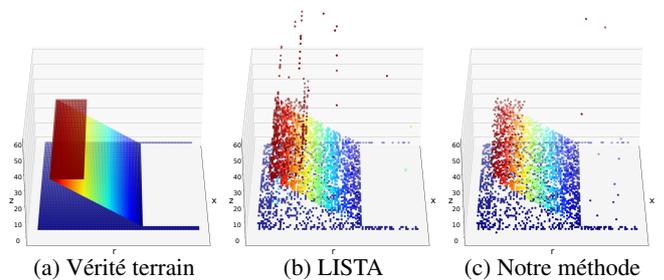


FIGURE 3 : Comparaison de LISTA et de notre méthode sur une scène 3D simulée.

TerraSAR-X et sur une vraie pile d’images. Les résultats sont analysés en comparaison à ceux de la méthode LISTA [3] entraînée sur les données décrites dans la section 4.

### 5.1 Résultats sur des données synthétiques

La scène simulée correspond à un bâtiment de 30 mètres de haut sur un sol plat. Elle est représentée dans la figure 3a en géométrie radar, c’est-à-dire dans l’espace  $(O, \vec{x}, \vec{r}, \vec{z})$ . On peut alors remarquer qu’on a de 0 à 3 cibles se mélangeant dans un même pixel  $(x, r)$ . La figure 3 montre les résultats de la méthode LISTA et de la méthode proposée. Dans le cas de LISTA (figure 3b), seules les cibles d’amplitude supérieure à un certain seuil sont affichées. De même, les cibles de probabilité d’existence inférieure à 0.5 ou dont l’incertitude sur la position est supérieure à 1 mètre sont masquées dans notre approche (figure 3c). Les reconstructions par ces deux méthodes sont globalement proches de la vérité terrain. On peut cependant souligner qu’il y a moins de cibles isolées avec notre approche.

### 5.2 Résultats sur des données réelles

On considère ici une zone contenant la tour-Mirabeau située dans le 15<sup>e</sup> arrondissement de Paris et mesurant environ 70 mètres de haut. Les nuages de cibles produits avec les deux

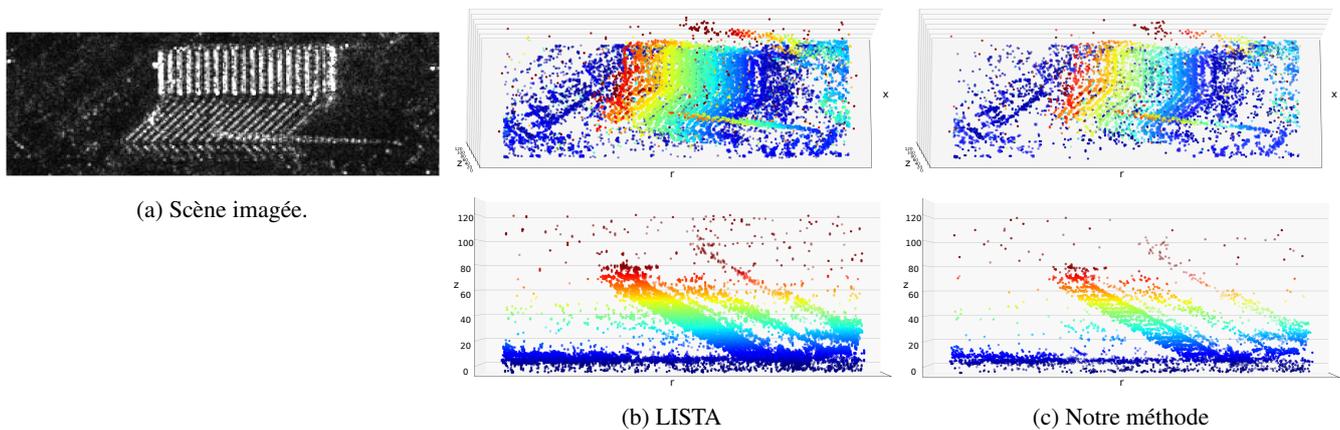


FIGURE 4 : Comparaison de LISTA et de notre méthode sur des données du capteur TerraSAR-X.

méthodes (figures 4b et 4c) semblent fidèles à ce qu'on peut voir sur l'image d'amplitude 4a. En effet, on reconnaît dans les deux cas la structure du bâtiment et les deux colonnes à proximité qui sont également visibles dans l'image 4a. De plus, la hauteur du bâtiment est plutôt bien estimée (voir la deuxième ligne de la figure 4). Comme avec les données simulées, on constate que la reconstruction avec LISTA contient plus de cibles éloignées des structures principales et donc qui ont moins de chances de correspondre à de réelles cibles. On peut également observer, sur la deuxième ligne de la figure 4, une plus nette séparation entre les différents étages du bâtiment avec la méthode proposée comme c'est le cas dans l'image d'amplitude.

## 6 Conclusion

Nous avons proposé une approche basée sur l'apprentissage profond imitant les itérations de l'algorithme de type "Matching Pursuit" RELAX qui approche la solution du maximum de vraisemblance des paramètres d'intérêt. À l'inverse des approches de l'état de l'art, notre méthode s'affranchit de l'utilisation d'une grille et permet d'estimer la position des cibles sur un intervalle continu. Les expériences sur les données simulées et réelles montrent que la méthode proposée permet de déterminer la position des cibles avec une bonne précision et produit peu de fausses alarmes (cibles isolées). De futurs travaux exploreront des formes de régularisation permettant d'exploiter la géométrie du voisinage local des cibles détectées.

**Remerciements :** Ce travail est financé partiellement par l'Agence de l'Innovation de Défense - AID - via le Centre Interdisciplinaire d'Études pour la Défense et la Sécurité - CIEDS (projet 2023 ALIA).

## Références

- [1] A. BUDILLON, A. EVANGELISTA et G. SCHIRINZI : Three-Dimensional SAR Focusing From Multipass Signals Using Compressive Sampling. *IEEE Transactions on Geoscience and Remote Sensing*, 2011.
- [2] J. GAO, Y. YE, S. LI, Y. QIN, X. GAO et X. LI : Fast Super-resolution 3D SAR Imaging Using an Unfolded Deep Network. 2019.
- [3] K. GREGOR et Y. LECUN : Learning fast approximations of sparse coding. *In Proceedings of the 27th International Conference on International Conference on Machine Learning*, 2010.
- [4] J. R. HERSHEY, J. Le ROUX et F. WENINGER : Deep Unfolding : Model-Based Inspiration of Novel Deep Architectures. *ArXiv*, 2014.
- [5] J. LI et P. STOICA : Efficient mixed-spectrum estimation with applications to target feature extraction. *IEEE Transactions on Signal Processing*, 1996.
- [6] L. LIEBEL et M. KÖRNER : Auxiliary tasks in multi-task learning, 2018.
- [7] K. QIAN, Y. WANG, Y. SHI et X. X. ZHU :  $\gamma$ -Net : Superresolving SAR Tomographic Inversion via Deep Learning. *IEEE Transactions on Geoscience and Remote Sensing*, 2022.
- [8] Y. REN, W. JIANG et Y. LIU : A New Architecture of a Complex-Valued Convolutional Neural Network for PolSAR Image Classification. *Remote Sensing*, 2023.
- [9] R. SHEN, S. WEI, Y. WEN, J. SHI et X. ZHANG : A 3-D Imaging Method Of Building With Tomosar Based On DUADMM-Net. *In IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium*, 2023.
- [10] P. STOICA et R.L. MOSES : *Spectral Analysis of Signals*. 2005.
- [11] M. WANG, Z. ZHANG, X. QIU, S. GAO et Y. WANG : ATASI-Net : An Efficient Sparse Reconstruction Network for Tomographic SAR Imaging With Adaptive Threshold. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [12] Y. WANG, C. LIU, R. ZHU, M. LIU, Z. DING et T. ZENG : MAda-Net : Model-Adaptive Deep Learning Imaging for SAR Tomography. *IEEE Transactions on Geoscience and Remote Sensing*, 2023.
- [13] X. X. ZHU et R. BAMLER : Tomographic SAR Inversion by  $L_1$  -Norm Regularization—The Compressive Sensing Approach. *IEEE Transactions on Geoscience and Remote Sensing*, 2010.