

Reconstruction tomographique par représentation neuronale implicite de volumes régularisée pour la microscopie électronique

Gauthier MULTARI¹ Julie DIGNE² Nicolas BONNEEL² Lucian ROIBAN³ Voichița MAXIM¹ Sébastien VALETTE¹

¹INSA-Lyon, CREATIS, UMR 5220 CNRS, Villeurbanne, France

²LIRIS, CNRS, Université Lyon 1, Villeurbanne, France

³INSA Lyon, Université Claude Bernard Lyon 1, CNRS, MATEIS, UMR5510, Villeurbanne, France

Résumé – La tomographie électronique permet l’analyse des matériaux en 3D à l’échelle nanométrique. Mais des contraintes mécaniques limitent la plage d’angles d’acquisition à environ 120 degrés. Ceci conduit à des artefacts d’élongation lors de l’utilisation de méthodes de reconstruction tomographiques classiques. Ce travail propose d’utiliser une représentation implicite du volume encodée dans un réseaux de neurones. Nous montrons que cette méthode est moins sujette aux artefacts d’angles manquants et proposons une régularisation de type Variation Totale pour améliorer la reconstruction.

Abstract – Electron tomography enables 3D analysis of materials at the nanometric scale. But mechanical constraints limit the range of acquisition angles to around 120 degrees. This leads to elongation artifacts when using conventional tomographic reconstruction methods. This work proposes the use of an implicit volume representation encoded in a neural network. We show that this method is less prone to missing angle artifacts and propose a Total Variation regularization to better improve reconstruction.

1 Introduction et contexte

La Microscopie Électronique en Transmission (MET) constitue l’état de l’art en termes de visualisation d’échantillons à l’échelle nanométrique. On peut citer des cas d’utilisation en science des matériaux, biologie structurale ou encore la nanotechnologie, où la visualisation précise de la morphologie, de la porosité ou encore de la composition chimique est critique. Cependant, cette technique présente des défis majeurs qui se répercutent lors de la reconstruction tomographique en 3D des échantillons. Les échantillons sont inclinés et imagés à différents angles. Les limitations mécaniques du microscope restreignent l’inclinaison à environ $\pm 60^\circ$ conduisant à des artefacts directionnels le long des faisceaux manquants, lors de la reconstruction avec des méthodes classiques. Un deuxième problème est lié à la sensibilité des échantillons et au fait qu’ils se dégradent sous l’effet des électrons. Ceci impose l’utilisation de très faibles doses d’électrons, ce qui donne des projections bruitées difficilement exploitables. Nous considérons le modèle direct de formation d’image suivant :

$$\mathbf{p} = \mathbf{A}\mathbf{f} + \varepsilon, \quad (1)$$

où $\mathbf{p} \in \mathbb{R}^M$ représente le vecteur de données de projection mesurées, $\mathbf{f} \in \mathbb{R}^N$ est une approximation en dimension finie de l’image à reconstruire, $\mathbf{A} \in \mathbb{R}^{M \times N}$ est la matrice du système, et ε le bruit. Ce problème discret correspond à l’observation de données échantillonnées de la fonction $\mathcal{A}\mu$, où \mathcal{A} est un opérateur linéaire et μ est le coefficient d’atténuation linéaire du milieu. En MET, pour chaque coupe 2D orthogonale à l’axe de rotation, \mathcal{A} est la transformée de Radon à angle manquant. Les méthodes classiques de reconstruction cherchent à résoudre le problème :

$$\mathbf{f}^* = \arg \min_{\mathbf{f}} d(\mathbf{A}\mathbf{f}, \mathbf{p}) + \lambda R(\mathbf{f}), \quad (2)$$

avec le terme d’attache aux données $d(\mathbf{A}\mathbf{f}, \mathbf{p})$ dépendant du type de bruit dans les images. La résolution numérique peut se faire par l’algorithme EM [3], dont les variantes sans régularisations donnent l’algorithme SIRT (*Simultaneous Iterative Reconstruction Technique*) [4] pour un bruit gaussien et MLEM (*Maximum Likelihood Expectation Maximization*) [12] pour un bruit poissonien. SIRT comme MLEM peuvent être vus comme des méthodes de descente de gradient, ce qui rend possible l’utilisation de l’algorithme FISTA (*Fast Iterative Shrinkage-Thresholding Algorithm*) [1]. Les versions FISTA de SIRT et MLEM ont été évaluées dans un contexte de MET par [7].

Nous présentons dans cet article une approche de reconstruction tomographique s’appuyant sur des réseaux neuronaux, que nous appelons T-NeF, enrichie par une régularisation de variation totale (T-NeF-TV). Contrairement aux méthodes classiques qui résolvent le problème discret, l’approche choisie dans ce travail cherche un estimateur $\hat{\mu}$ défini en tout point du volume par un réseau de neurones Γ_θ .

Nous comparons les résultats de la méthode proposée à des reconstructions obtenues par reconstruction itérative sur volume voxelisé. Nous considérons deux algorithmes : SIRT-TV et MLEM-TV. Le premier est un choix classique en MET. Le deuxième l’est moins, car il suppose un bruit poissonien. Il a été montré cependant dans [7] qu’il conduisait à des résultats supérieurs pour des acquisitions à angle manquant et que sa convergence était plus rapide.

2 Champs neuronaux

En 2022, Mildenhall et al. [8] ont proposé une représentation implicite de scène 3D encodée par un réseau de neurones, appelée NeRF (Champs de Radiance Neuronaux). Plusieurs adaptations et évaluations de cette approche ont été proposées en tomographie classique [15], ainsi qu’en tomographie élec-

tronique, que ce soit sans TV [5, 2, 14], ou avec TV [11, 16]. Notre approche se démarque du fait que la variation totale n'est ni calculée dans l'espace de l'encodage, ni dans l'espace image discret mais dans l'espace du champ neuronal après encodage des caractéristiques.

À condition de disposer d'un nombre suffisant de paramètres, les réseaux de neurones sont capables d'encoder des signaux continus dans des dimensions arbitraires. Ainsi, le volume à reconstruire est encodé dans une fonction continue, à la différence d'une représentation en voxels ou en maillage. Par ailleurs, leur nature adaptative leur permet de moduler dynamiquement l'information associée à une fréquence donnée. Une telle représentation présente plusieurs avantages directs, notamment la possibilité de modifier les dimensions d'entrée et de sortie sans nécessiter d'adaptation structurelle du modèle et de nouvelles possibilités de régularisation, comme celle proposée dans la suite de cet article.

2.1 Encodage positionnel

Les réseaux de neurones privilégient l'apprentissage des basses fréquences, ce qui conduit à une incapacité à apprendre les détails haute fréquence d'un volume donné avec une simple entrée 3D. L'une des approches possibles pour pallier cette caractéristique est d'encoder les coordonnées spatiales dans un espace de plus haute dimension. On définit $\mathcal{H} : \mathbb{R}^3 \mapsto \mathbb{R}^d$ comme étant un encodage positionnel transformant des coordonnées spatiales $\mathbf{x} = (x, y, z)$ en un vecteur de caractéristiques de dimension $d \gg 3$ et servant d'entrée à un MLP $\Gamma_\theta : \mathbb{R}^d \mapsto \mathbb{R}$, paramétré par θ , qui prédit la valeur d'atténuation $\hat{\mu}(\mathbf{x})$ associée. Soit $f_\theta = \Gamma_\theta \circ \mathcal{H}$:

$$\hat{\mu}(\mathbf{x}) = f_\theta(\mathbf{x}), \quad (3)$$

2.2 Fonction de perte

Le modèle de projection utilisé est similaire à celui des méthodes de reconstruction classiques et simule la loi de Beer-Lambert le long de rayons virtuels $\mathbf{r}(s)$ tirés pour chaque pixel des projections. Chaque rayon est discrétisé en L points $\mathbf{s}_i \in \mathbb{R}^3$ échantillonnés spatialement pour lesquels le MLP Γ_θ guidé par l'encodage positionnel \mathcal{H} prédit la densité locale.

L'intégrale du rayon est approximée par quadrature numérique :

$$\mathbf{p}_\theta = \int_{\mathbf{s}} f_\theta\{\mathbf{r}(s)\} ds \approx \sum_{i=1}^L f_\theta\{\mathbf{r}(\mathbf{s}_i)\} \cdot \Delta \mathbf{s}_i. \quad (4)$$

Le modèle ainsi défini est différentiable et physiquement réaliste. L'attache aux données est assurée par une fonction de perte $\mathcal{L}_{\text{data}}$ quadratique qui minimise l'écart entre les projections simulées et celles acquises pour chaque rayon. En pratique, des lots (*batches*) \mathcal{R} de sous-ensembles \mathbf{j} de rayons sont sélectionnés à chaque itération pour des angles de projection distincts. Cela permet un meilleur apprentissage en plus de réduire la charge mémoire nécessaire à chaque descente de gradient :

$$\mathcal{L}_{\text{data}} = \frac{1}{|\mathcal{R}|} \sum_{\mathbf{j} \in \mathcal{R}} \left\| \mathbf{p}_\theta^{(\mathbf{j})} - \mathbf{p}^{(\mathbf{j})} \right\|_2^2. \quad (5)$$

Régularisation de Variation Totale Nous proposons de régulariser l'optimisation en ajoutant une contrainte de variation totale à la fonction de perte $\mathcal{L}_{\text{data}}$: de nouveaux rayons r_{TV} traversant le volume à reconstruire sont tirés aléatoirement. Similairement à l'approche d'optimisation, chaque rayon est discrétisé en L points. Il en résulte un vecteur de valeurs échantillonnées $f_\theta\{r_{TV}(s_i)\}$, et la variation totale de l'ensemble des rayons r_{TV} contribue à une valeur de perte additionnelle :

$$\mathcal{L}_{TV} = \sum_{r_{TV}} \sum_{i=1}^{L-1} |f_\theta\{r_{TV}(s_{i+1})\} - f_\theta\{r_{TV}(s_i)\}|. \quad (6)$$

La fonction de perte complète proposée dans ce travail est la suivante :

$$\mathcal{L}_{\text{Total}} = \mathcal{L}_{\text{Data}} + \lambda \mathcal{L}_{TV} \quad (7)$$

Avec $\lambda \in \mathbb{R}_+$ un scalaire contrôlant le poids de la régularisation par rapport à l'attache aux données, fixé à 0.1 dans nos expériences. Il a été constaté expérimentalement que de meilleurs résultats étaient obtenus lorsque les rayons $r(s)$ et $r_{TV}(s)$ étaient différents entre eux à chaque étape d'optimisation.

2.3 Optimisation et entraînement

L'architecture du réseau de neurones utilisé est la suivante : Perceptron Multi-Couches 6×128 avec des fonctions d'activation ReLU entre ses couches. La taille de la couche d'entrée du réseau est définie par les paramètres de l'encodage positionnel. Nous adoptons dans ce travail un encodage par table de hachage multirésolution (*Multiresolution Hash Encoding*) proposé par [9] avec 32 niveaux, chaque niveau générant 2 caractéristiques, combinées pour former la représentation finale. La grille commence par une résolution de 1 au niveau le plus grossier (soit l'ensemble du volume normalisé), chaque niveau suivant ayant une résolution multipliée par un facteur 4. L'encodage positionnel possède environ 33,5 millions de paramètres et le réseau de neurones environ 91 mille. L'échantillonnage des rayons est effectué avec 256 échantillons, et la taille de lot est fixée à 2048 rayons. Nous employons un taux d'apprentissage de 1×10^{-4} avec l'optimiseur Adam (paramètres par défaut : $\beta_1 = 0,9$ et $\beta_2 = 0,99$), ainsi qu'un planificateur de taux d'apprentissage multi-étapes avec $\gamma = 0,7$. Le schéma d'entraînement comprend 3 *epochs*, elles-mêmes constituées de $\left\lceil \frac{M}{|\mathcal{R}|} \right\rceil$ itérations, avec M correspondant au nombre total de rayons dans le jeu de données.

3 Résultats

Nous présentons les résultats des différentes méthodes décrites sur deux types de données. Pour produire des résultats quantitatifs, un fantôme de type Shepp-Logan de résolution 256^3 a été simulé. Deux plages de projections ont été générées, une "idéale" couvrant 180 degrés ($\pm 90^\circ$) et une plus proche du cas MET avec une plage de 120 degrés. Les deux plages sont constituées d'incrément d'angles d'un degré. Des données MET d'une bipyramide d'or et d'argent ont été mises à disposition par le laboratoire MATEIS.

La table 1 compare la qualité (PSNR) des reconstructions sur le fantôme Shepp-Logan, selon plusieurs méthodes : SIRT,

TABLE 1 : Comparaison quantitative entre les différentes approches de reconstruction sur le fantôme Shepp-Logan.

angles α	bruit σ	Méthode	PSNR
180	0	SIRT	39.81
		MLEM	48.96
		T-NeF	34.40
		T-NeF-TV	36.00
120	0	SIRT-TV	25.23
		MLEM-TV	27.37
		T-NeF	27.30
		T-NeF-TV	30.55
	0.01	SIRT-TV	27.19
		MLEM-TV	21.46
		T-NeF	26.56
		T-NeF-TV	29.92
	0.1	SIRT-TV	22.41
		MLEM-TV	21.42
		T-NeF	21.90
		T-NeF-TV	23.41

MLEM, notre approche T-Nef et notre approche contrainte par Variation T-Nef-TV, selon plusieurs conditions. α indique la plage d'angles disponible (180° signifiant qu'il n'y a pas d'angle manquant). σ indique le niveau de bruit gaussien relatif. En dehors du cas $\alpha = 180$, notre approche engendre des résultats de meilleure qualité que les approches classiques. Les figures 1 et 2 comparent les résultats de reconstruction de notre méthode avec la méthode classique MLEM-TV, respectivement sur des données sans et avec bruit.

La figure 3 montre les résultats de reconstruction de nos approches sur des données expérimentales en microscopie électronique. La première ligne montre des coupes du volume reconstruit avec T-NeF, la deuxième avec T-NeF-TV, et la troisième montre des reconstructions 3D surfaciques de la bipyramide.

4 Conclusion et perspectives

Nous avons présenté une approche de reconstruction tomographique prometteuse, exploitant à la fois les représentations neuronales implicites et la régularisation par mesure de variation totale. Ce travail s'inscrit dans la lignée des travaux expérimentaux investiguant le rôle de la représentation de l'espace dans les problèmes inverses, en particulier les représentations par noyaux et par réseaux de neurones. Il a été montré dans la littérature que les deux étaient reliés, et qu'un réseau dont la dimension des couches tend vers l'infini et pour un taux d'apprentissage qui tend vers zéro pouvait être représenté linéairement avec des noyaux neuronaux tangents [6]. Les propriétés de ces noyaux peuvent être modifiées en utilisant un encodage spatial, afin d'enrichir le contenu fréquentiel des images [13]. Il est intéressant de noter que ces représentations sont étudiées au-delà de l'apprentissage profond et elles donnent de meilleurs résultats tout en permettant une réduction du nombre de "paramètres", au sens du remplacement des voxels par les paramètres caractérisant les noyaux [10].

Remerciements

Ce travail a bénéficié du soutien financier de la Fédération Informatique de Lyon (FIL)

Références

- [1] A. BECK *et al.* : A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.
- [2] V. DEBARNOT, V. KISHORE, R. RIGHETTO et I. DOKMANIC : Ice-tide : Implicit cryo-et imaging and deformation estimation. *IEEE Transactions on Computational Imaging*, PP:1–14, 01 2024.
- [3] A. P. DEMPSTER *et al.* : Maximum likelihood from incomplete data via the em algorithm. *Journal of the royal statistical society*, 39(1):1–22, 1977.
- [4] P. GILBERT : Iterative methods for the three-dimensional reconstruction of an object from projections. *Journal of Theoretical Biology*, 36(1):105–117, 1972.
- [5] H. KNIESEL *et al.* : Clean implicit 3d structure from noisy 2d stem images. *In CVPR, IEEE*, 2022.
- [6] J. LEE *et al.* : Wide neural networks of any depth evolve as linear models under gradient descent. *NEURIPS*, 32, 2019.
- [7] T. LEULIET *et al.* : Efficiency of TV-regularized algorithms in computed tomography with Poisson-Gaussian noise. *In EUSIPCO*, pages 1294–1298. IEEE, 2021.
- [8] B. MILDENHALL *et al.* : Nerf : Representing scenes as neural radiance fields for view synthesis. *In ECCV*, 2020.
- [9] T. MÜLLER *et al.* : Instant neural graphics primitives with a multiresolution hash encoding. *ACM Trans. Graph.*, 41(4):102 :1–102 :15, juillet 2022.
- [10] E. OZSAR *et al.* : Parametric level-sets enhanced to improve reconstruction (PaLEnTIR). *Inverse Problems*, 41(2):025004, 2025.
- [11] D. RUCKERT *et al.* : Neat : neural adaptive tomography. *ACM Trans. Graph.*, 41(4), juillet 2022.
- [12] L. A. SHEPP *et al.* : Maximum likelihood reconstruction for emission tomography. *IEEE Transactions on Medical Imaging*, 1(2):113–122, 1982.
- [13] M. TANCIK *et al.* : Fourier features let networks learn high frequency functions in low dimensional domains. *NEURIPS*, 33:7537–7547, 2020.
- [14] V. VEEN *et al.* : Missing wedge completion via unsupervised learning with coordinate networks. *International Journal of Molecular Sciences*, 25(10):5473, 2024.
- [15] X. WANG *et al.* : Neural radiance fields in medical imaging : A survey, 2025.
- [16] Y. WANG *et al.* : Learning adaptive tensorial density fields for clean cryo-et reconstruction. *In NEURIPS*, volume 36, pages 72270–72283. Curran Associates, Inc., 2023.

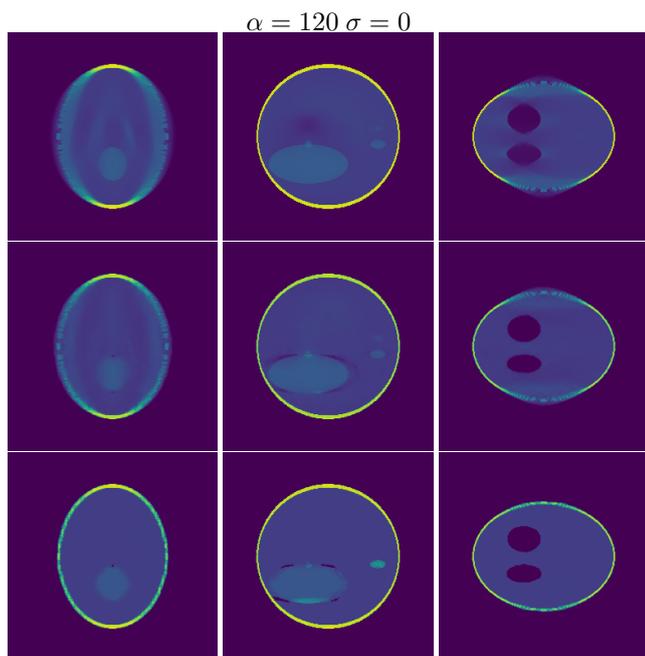
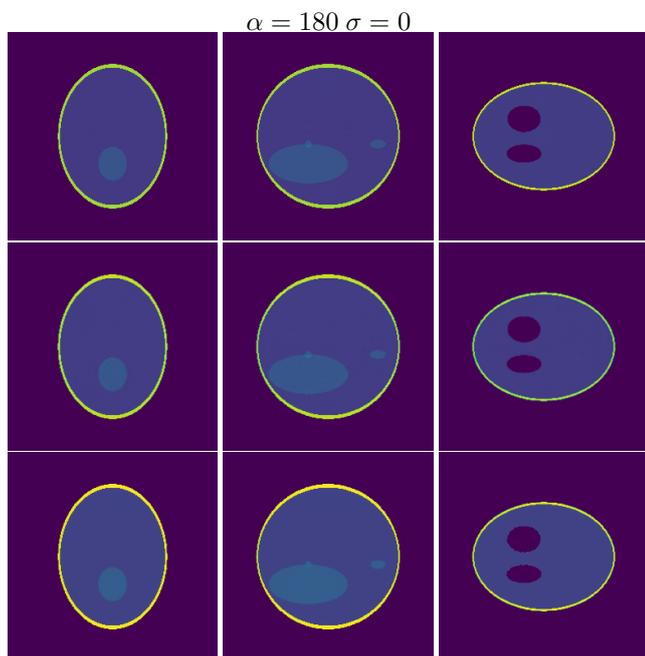


FIGURE 1 : Comparaison des reconstructions avec NeRF et les méthodes traditionnelles, données sans bruit. Pour chacun des deux sous-blocs, chaque ligne représente respectivement les reconstructions par MLEM, T-NeF et T-NeF-TV

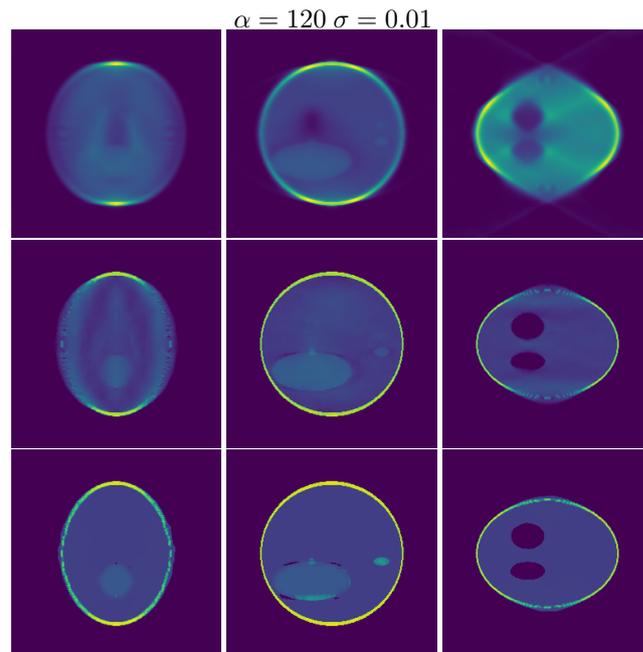


FIGURE 2 : Comparaison des reconstructions avec NeRF et les méthodes traditionnelles. Pour chacun des trois sous-blocs, chaque ligne représente respectivement les reconstructions par MLEM-TV, T-NeF et T-NeF-TV

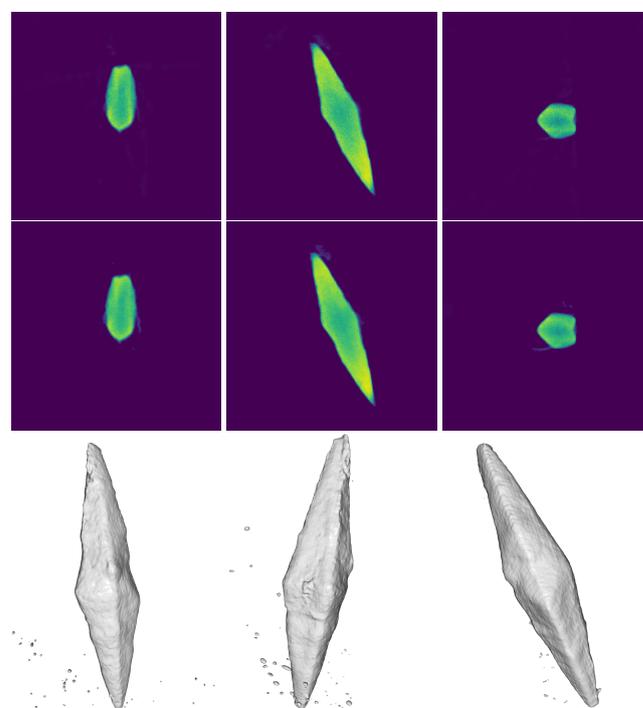


FIGURE 3 : Reconstruction d'une nano-particule (bipyramide d'or et d'argent) obtenue par tomographie électronique