



ÉCOLE CENTRALE DE NANTES

PROJET D'OPTION

REPORT

Représentation neuronale implicite en tomographie par émission de positrons

AUTEURS :

MAURICIO SALIM GÓMEZ CHICRE
ÁUREO HENRIQUE E SILVA MARQUES

TUTEURS :

DIANA MATEUS
YOUNÈS MOUSSAOUI

13 mai 2024

Table des matières

1	Introduction	2
2	Travail Réalisé	3
2.1	Étude bibliographique	3
2.2	Modèle utilisé - SIREN	4
2.3	Mise en ouvre de l'expérience	4
2.4	Résultats	5
3	Prochaines étapes	5

1 Introduction

La tomographie par émission de positons (TEP) est une modalité d'imagerie nucléaire qui mesure l'activité métabolique des cellules des tissus corporels. La TEP est en fait une combinaison de médecine nucléaire et d'analyse biochimique. Utilisée principalement pour des patients souffrant d'affections cérébrales ou cardiaques et de cancer, en leur permettant de quantifier un de leurs processus physiologiques [4]. [5]

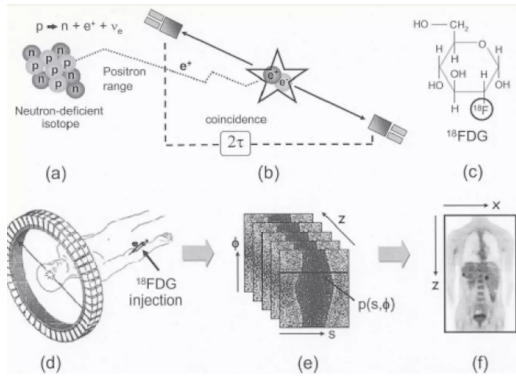


FIGURE 1 – v [2].

La TEP fonctionne comme un dispositif de balayage qui détecte les photons émis par un radionucléide (Figure 1 (a), (b) et (c)) dans l'organe ou le tissu examiné. Les radionucléides sont fabriqués en attachant un atome radioactif à des substances chimiques qui sont utilisées naturellement par l'organe ou le tissu en question au cours de son processus métabolique. Un des traceurs les plus utilisés en clinique est le fluorodéoxyglucose (FDG), qui est fabriqué en introduisant un atome radioactif dans la molécule de glucose. Etant le cerveau utilise le glucose pour son métabolisme, le FDG s'avère être une substance très utile dans certaines applications médicales[4].

Dans ce contexte, la transformée de Radon [10] constitue un outil mathématique fondamental pour la reconstruction d'images à partir des données de projection acquises lors du processus d'imagerie. La transformée de Radon, notée $R[f](s, \theta)$, prend une fonction $f(x, y)$ représentant la distribution des radiopharmaceutiques émetteurs de positrons (ou la distribution spatiale des coefficients d'atténuation des rayons X pour la tomodynamométrie) à l'intérieur du corps et la transforme en une fonction de deux variables, s et θ , correspondant à la distance depuis l'origine le long d'une ligne particulière et l'angle de cette ligne par rapport à un axe de référence, respectivement. Mathématiquement, la transformée de Radon est définie comme suit :

$$R[f](s, \theta) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x, y) \delta(x \cos(\theta) + y \sin(\theta) - s) dx dy$$

où δ représente la fonction delta de Dirac. Cette intégrale calcule l'intégrale de ligne de $f(x, y)$ le long de chaque ligne spécifiée par les paramètres s et θ . En collectant ces intégrales de ligne sur toutes les directions et distances possibles, un sinogramme est généré, qui encapsule des informations essentielles sur la distribution de notre fonction. L'application ultérieure de techniques de transformée de Radon inverse permet la reconstruction d'une image détaillée représentant la distribution spatiale des sources émettrices de positrons, facilitant l'interprétation diagnostique et la prise de décision médicale en imagerie TEP. La quantité de radionucléide recueillie dans le tissu affecte la luminosité du tissu sur l'image et indique le niveau de fonctionnement de l'organe ou du tissu [10].

Cependant, les scanners TEP présentent deux problèmes principaux : premièrement, toutes les coïncidences des deux photons émis ne sont pas correctes (il existe également des coïncidences aléatoires, multiples et diffuses qui contribuent au bruit [1]) ; deuxièmement, les radionucléides sont nocifs pour l'organisme, ce qui oblige à réduire la quantité appliquée, et donc réduit la qualité des images. Il est également intéressant de réduire le temps de balayage afin de ne pas analyser tous les angles, puisque la durée des scanners PET peut varier de 10 à 40 minutes, en fonction de la section du corps à examiner et de la taille de la section (sans compter le temps nécessaire pour que le radionucléide soit absorbé par le tissu objectif) [3, 4].

Étant donné que le scanner TEP fournit un sinogramme et non une image de l'organe en question, une reconstruction est nécessaire. Cette reconstruction repose sur des modèles mathématiques et des algorithmes itératifs, impliquant la minimisation d'un critère d'erreur entre le sinogramme d'entrée et le sinogramme estimé par le modèle, ou sur la maximisation d'une vraisemblance de ces deux derniers.

Le développement des outils basés sur l'intelligence artificielle (IA) permet l'utilisation de techniques d'IA et d'apprentissage profond (AP) pour reconstruire l'image TEP, ou simplement pour réaliser un débruitage [7]. C'est ici que l'on s'intéresse par des Représentations Neuronales Implicites (RNI), car les approches conventionnelles encodent le signal d'entrée dans des représentations discrétisées ou partitionnées de l'espace d'entrée (nuages de points, grilles de voxels et maillages). Cependant, l'RNI associe les coordonnées d'entrée à leur valeur correspondante dans l'espace d'entrée, ce qui permet une représentation continue.

Ici, un perceptron multicouche (MLP) est entraîné à paramétrer le signal d'intérêt, tel qu'une image ou une forme, en utilisant des coordonnées comme entrée. L'objectif est de prédire les valeurs de données correspondantes à ces coordonnées. Ainsi, le MLP sert de fonction de représentation neuronale implicite qui encode la représentation du signal dans ses poids.

Les modèles RNI se sont révélés efficaces en termes de mémoire et capables de relever les défis posés par le coût élevé des données étiquetées et les ressources limitées en mémoire (qui ont fait l'objet d'une grande

attention, comme en témoigne la croissance exponentielle du nombre d'articles de recherche). Ils offrent une solution pour alléger le fardeau de la collecte de données étiquetées dans les domaines médicaux en éliminant le besoin de données d'apprentissage et d'étiquettes explicites. Au lieu de cela, ils exploitent les mesures ou les signaux disponibles sans nécessiter de données étiquetées (ce qui est laborieux, coûteux, prend du temps et demande un effort important dans le contexte de l'imagerie médicale) pour chaque cas, ce qui permet la reconstruction de structures anatomiques détaillées en 3D ou la génération de scans en 2D [7].

En outre, l'obtention de scans IRM, CT et PET de haute qualité pose un défi en raison des temps d'analyse plus longs (et de l'exposition accrue aux radionucléides pour les scans PET). Les approches conventionnelles ne permettent pas de traiter efficacement ces problèmes par la super-résolution ou la reconstruction d'images, car elles ne sont pas agnostiques en termes de résolution et nécessitent des quantités considérables de données. Cependant, les RNI peuvent traiter la super-résolution de manière plus efficace en prenant en compte les entrées du domaine des coordonnées continues et en étant indépendantes de la résolution. En utilisant les RNI, il devient possible de reconstruire des CT ou des IRM directement à partir du domaine des capteurs [7].

Plus précisément, la tâche de reconstruction d'images est confrontée à de nombreux défis, notamment la nécessité de disposer d'ensembles de données d'entraînement à grande échelle, l'instabilité en présence de modifications structurelles et les difficultés de généralisation à diverses modalités d'images ou localisations anatomiques. Il est essentiel de surmonter ces obstacles pour améliorer la fiabilité et l'applicabilité de la reconstruction d'images dans le domaine médical. Dans le domaine de l'imagerie médicale, les méthodes RNI comme NeRP [8] ont déjà montré leur efficacité dans la reconstruction de la progression structurelle des tumeurs et leur applicabilité polyvalente à diverses modalités d'imagerie, qui ont été démontrées par des expériences menées sur des données 2D et 3D provenant de CT et IRM [7].

Ainsi, dans ce travail nous étudions l'application des Représentations Neuronales Implicites (RNI) en imagerie TEP. Bien que les RNIs ont principalement été développées pour représenter des scènes naturelles complexes en 2D et 3D (très efficaces grâce à qu'elles utilisent des représentations de signaux implicitement définies, continues et différentielles [6]), ils ont déjà démontré sa capacité dans le domaine médical [7]. Les RNI se sont révélées être un paradigme puissant, offrant de nombreux avantages par rapport aux représentations conventionnelles dans les tâches de reconstruction, de débruitage et de super-résolution. Donc, l'objectif de ce projet est d'adapter ces représentations au contexte de l'imagerie TEP et CT (tomographie par ordinateur) pour les tâches de débruitage et de reconstruction d'images.

2 Travail Réalisé

2.1 Étude bibliographique

Les RNI présentent plusieurs avantages tels que leur nature agnostique en termes de résolution, leur efficacité en termes de mémoire, leur capacité à éviter les biais de localité et leur différentiabilité, ce qui leur permet de s'adapter à différentes tâches. Grâce à l'efficacité des RNI en termes de mémoire et de données, ils sont largement utilisés dans de nombreuses tâches d'imagerie médicale. L'un des défis les plus importants de l'imagerie médicale automatisée est la collecte de données annotées de vérité terrain à partir de sources fiables, telles que les cliniciens et les professionnels de la santé. Ce processus est laborieux, coûteux, prend du temps et nécessite des efforts considérables. [7]

En outre, les RNI sont particulièrement utiles pour des tâches telles que la super-résolution. Pendant les examens d'imagerie médicale (CT, PET, IRM ou ultrasons), les mouvements du patient peuvent provoquer des artefacts de mouvement. L'obtention de scans de haute qualité pose un défi en raison des temps de balayage plus longs. Les RNI peuvent résoudre ce problème par le biais de la super-résolution de manière plus efficace en prenant en compte les entrées du domaine des coordonnées continues et en étant agnostiques en termes de résolution. [7]

Les RNIs sont également utilisés dans les applications biomédicales, en particulier pour résoudre les problèmes d'imagerie inverse, où ils prennent en entrée des mesures bruitées ou sous-échantillonnées d'une image médicale et visent à générer en sortie une image complète reconstruite. Le caractère incomplet de l'entrée peut être dû à des contraintes de temps, à une exposition réduite aux rayonnements ou aux mouvements du patient. Le cadre NeRP [8], qui propose d'intégrer des réseaux neuronaux implicites pour reconstruire des images médicales (MRI et CT spécifiquement) faiblement échantillonnées en trois étapes sans nécessiter de données d'apprentissage, est l'un des exemples les plus marquants de mise en œuvre dans ce domaine.

D'autre part, les RNI peuvent ne pas encoder les détails des hautes fréquences par rapport aux représentations discrètes, ce qui conduit à une qualité de représentation supprimée appelée biais spectral. Pour résoudre ce problème, plusieurs approches ont été explorées : Premièrement, en modifiant l'entrée, on peut envisager de changer le type d'entrée en le mapant dans un espace à plus haute dimension pour permettre au réseau de capturer des détails plus complexes dans le signal, comme le fait très bien NERF [6] en codant

son entrée dans un espace à plus haute dimension de sinus et de cosinus (et aussi en utilisant un échantillonnage hiérarchique des volumes) ; Deuxièmement, remplacer la fonction d'activation ReLU par une nouvelle qui facilite davantage l'apprentissage des composantes à haute fréquence, comme l'ont fait les acteurs de SIREN [9] en utilisant un sinus comme fonction d'activation périodique ; et enfin, modifier également la sortie du MLP dans un espace de dimension supérieure, où chaque nœud est responsable de la reconstruction d'une partie spécifique du signal, ce que nous n'aborderons pas pour l'instant.

2.2 Modèle utilisé - SIREN

En nous basant sur les modèles étudiés, nous avons décidé de débiter la partie pratique de notre étude en utilisant une architecture simple appelée SIREN (Sinusoidal Representation Networks). Le modèle SIREN propose une approche novatrice pour les représentations neuronales implicites. Cette architecture se distingue par sa simplicité intrinsèque, utilisant la fonction sinus comme fonction d'activation périodique. Symboliquement, chaque couche du réseau, représentée par ϕ_i , effectue une transformation affine sur l'entrée x_i . Pour la première couche, dans notre étude, cette entrée x_i est représentée par les coordonnées 2D de l'image qui sera reconstruite. Cette transformation implique une multiplication matricielle avec la matrice de poids W_i et l'addition des biais b_i . Ensuite, une non-linéarité sinusoidale est appliquée à chaque élément du vecteur résultant, encapsulant l'essence de l'architecture SIREN [9]. La fonction d'activation appliquée aux neurones de chaque couche peut être visualisée dans l'équation ci-dessous. La valeur de $\phi_i(x_i)$, la sortie résultante de la couche i , est une valeur comprise entre -1 et 1, étant la réponse de la fonction sinus. Lorsque ces valeurs sont transmises à la couche suivante, cela implique que ces couches ont reçu en entrée une distribution d'arcsinus. Enfin, la dernière couche renverra, dans notre étude, les intensités de l'image reconstruite.

$$\Phi(x) = W_n(\phi_{n-1} \circ \phi_{n-2} \circ \dots \circ \phi_0)(x) + b_n, \quad \phi_i(x_i) = \begin{cases} \sin(W_i x_i + b_i), & \text{si } i \neq 0 \\ \sin(\omega_0 W_i x_i + b_i), & \text{si } i = 0 \end{cases}$$

La propriété distinctive de SIREN est sa capacité à avoir ses dérivées également conformes à la structure SIREN, grâce à la propriété trigonométrique selon laquelle la dérivée du sinus est le cosinus—un sinus déphasé. Cette propriété intrinsèque facilite la supervision des dérivées complexes [9].

Une application notable de l'architecture SIREN est sa capacité d'initialisation unique, permettant le contrôle sur les distributions d'activation et facilitant la construction d'architectures profondes, ce qui est possible grâce à un bon choix du hyperparamètre ω_0 . Le choix de ω_0 sert de schéma d'initialisation pour préserver la distribution des activations à travers le réseau, de sorte que la sortie finale à l'initialisation ne dépende pas du nombre de couches. Comme cela affecte la distribution des poids (ils doivent être uniformément distribués), ne pas le choisir avec soin peut entraîner de mauvaises performances à la fois en termes de précision et de vitesse de convergence. L'objectif est alors que l'initialisation de la première couche du réseau sinusoidal avec la fonction sinus $\sin(\omega_0 Wx + b)$ couvre plusieurs périodes sur $[-1, 1]$. Une explication plus détaillée et une preuve sont présentées dans le matériel supplémentaire de SIREN. De plus, la convergence accélérée des SIRENs, comparée aux architectures de référence, a été validée empiriquement [9].

Dans des scénarios expérimentaux, les modèles SIREN démontrent une précision remarquable dans la représentation implicites d'images, réalisant des ajustements précis en quelques centaines d'itérations et en quelques secondes sur des GPU contemporains. L'efficacité de la convergence, associée à la capacité de gérer des différentes fréquences d'image et de maintenir un comportement stable, a joué un rôle central dans le choix du modèle SIREN pour la tâche spécifique de reconstruction et de débruitage d'images de tomographie.

2.3 Mise en oeuvre de l'expérience

Dans un premier volet de notre projet, nous étudions l'application de l'architecture SIREN à l'imagerie CT, afin de démontrer la faisabilité de cette architecture pour la reconstruction d'une modalité de l'imagerie médicale, sans faire appel à des méthodes inverses. Pour cela, on considère un fantôme synthétique "Shepp-Logan", avec une projection Radon réalisée en utilisant la bibliothèque NumPy de Python, en faisant varier le nombre d'angles projetés. Un bruit gaussien d'écart-type égal à 2 est ajouté au sinogramme généré. L'objectif est d'analyser si le modèle est capable de reproduire l'image reconstruite sans considérer le bruit du sinogramme. Étant donné que le bruit constitue un détail de haute fréquence, on s'attend à ce que le modèle apprenne ces caractéristiques en dernier. Par conséquent, nous pouvons arrêter l'exécution du modèle avant que cela ne se produise ou enregistrer le point où il a présenté la meilleure performance.

L'expérience consiste à fournir les coordonnées de l'image CT en tant qu'entrée (*Model input*) au modèle (*SIREN*) pour obtenir en sortie l'image CT reconstruite (*Model output*). La partie importante ici est la conception de la fonction de coût (*Loss*). Pour ce calcul, la transformation de Radon de *Model output* a été réalisée et l'erreur MSE (*Mean Square Error*) entre le sinogramme résultant et le sinogramme original a été calculée. Un

schéma du modèle est présenté dans la Figure 2. Cette expérience ressemble à un cas pratique, étant donné qu'en tomographie, nous avons accès uniquement au sinogramme et non à l'image finale.

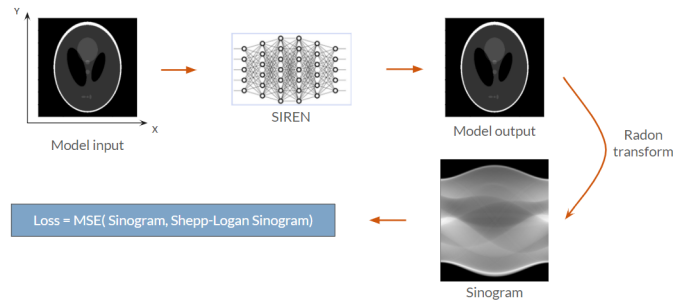


FIGURE 2 – Schéma de l'expérience proposée

Nous avons exécuté le modèle avec l'optimiseur Adam, un taux d'apprentissage de 10^{-4} , 6000 étapes, en faisant varier le paramètre hyperparamètre ω_0 , caractéristique du modèle Siren. Dans le but de reproduire un scénario réaliste, le sinogramme de test a été construit avec différents nombres de réalisations (40, 60, 80 et 100).

2.4 Résultats

Après l'analyse des tests pour ω_0 égal à 40 (Tableau 1, indiqué principalement par les mesures de PSNR), nous avons constaté que à mesure que le nombre de réalisations diminue, la performance diminue également. Par conséquent, les nouveaux tests ont considéré uniquement un nombre de réalisations égal à 40 en faisant varier ω_0 de 10 à 50. Les nouveaux résultats sont présentés ci-dessous (Tableau 2). En observant les résultats, on remarque que le meilleur résultat est obtenu avec un ω_0 de 20. On peut également observer les résultats qualitatifs dans la Figure 3. Dans l'image, on peut remarquer qu'à l'intérieur de la structure, les détails sont assez flous au point que certains d'entre eux ne peuvent pas être remarqués. À partir du graphique PSNR, il est perceptible qu'il existe un point à partir duquel une diminution débute. Cela correspond au moment où le modèle commence à assimiler les détails de haute fréquence, en l'occurrence le bruit gaussien. Dans notre tâche, l'intérêt est d'interrompre prématurément l'entraînement du modèle avant qu'il n'acquière ces fréquences élevées.

TABLE 1 – Tests avec $\omega_0 = 40$

N° de Réalisations	MSE	PSNR
100	0,0017	27,2866
80	0,0018	27,4443
60	0,0021	26,7677
40	0,0033	24,8172

TABLE 2 – Tests avec N de réalisations = 40

ω_0	MSE	PSNR
10	0,0022	26,546
20	0,0018	27,0504
30	0,0021	26,8212
40	0,0033	24,8172
50	0,0063	22,0083

3 Prochaines étapes

La prochaine étape de notre travail consiste à appliquer l'architecture SIREN dans le contexte de l'imagerie TEP, pour une tâche de débruitage et de reconstruction. Nous cherchons à reconstruire une image TEP avec une loss définie comme la MSE entre le sinogramme d'entrée et celui de l'image de sortie.

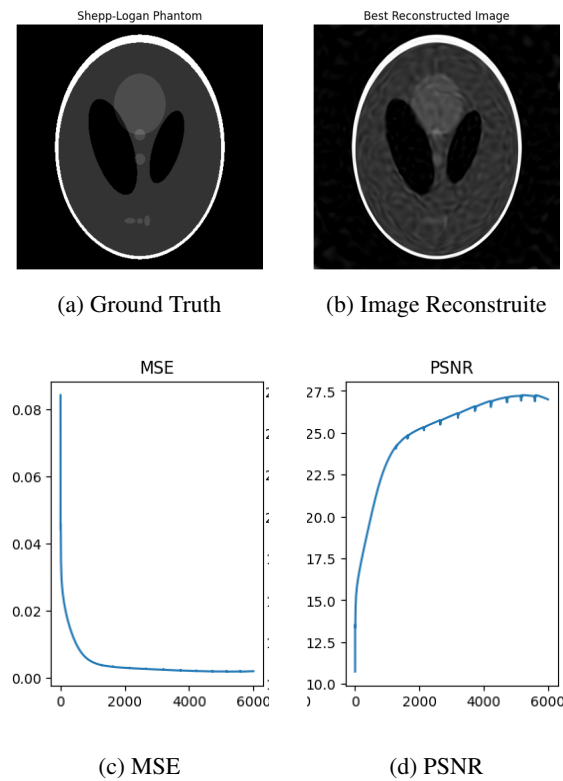


FIGURE 3 – Résultats pour l'expérience avec la meilleure performance de PSNR ($\omega_0 = 20$)

Références

- [1] T. CARLIER. *Tomographie à Émission de Positons : quelques aspects techniques*. 2023.
- [2] Vibha CHASWAL. *Causes of Noise in PET imaging*. URL : <https://es.slideshare.net/vibhachaswal/causes-of-noise-in-pet-imaging>.
- [3] BMJ Publishing GROUP. *Positron Emission Tomography (PET)*. URL : <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1126321/>.
- [4] Johns HOPKINS. *Positron Emission Tomography (PET)*. URL : [https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/positron-emission-tomography-pet#:~:text=Positron%20emission%20tomography%20\(PET\)%20is,nuclear%20medicine%20and%20biochemical%20analysis..](https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/positron-emission-tomography-pet#:~:text=Positron%20emission%20tomography%20(PET)%20is,nuclear%20medicine%20and%20biochemical%20analysis..)
- [5] Daniel LEJEUNE et al. *TITAN : Bringing The Deep Image Prior to Implicit Representations*. 2022. arXiv : 2211.00219 [eess.IV].
- [6] Ben MILDENHALL et al. « Nerf : Representing scenes as neural radiance fields for view synthesis ». In : (2021).
- [7] Amirali MOLAEI et al. « Implicit Neural Representation in Medical Imaging : A Comparative Survey ». In : (2023).
- [8] Liyue SHEN, John PAULY et Lei XING. « Nerf : implicit neural representation learning with prior embedding for sparsely sampled image reconstruction ». In : *IEEE Transactions on Neural Networks and Learning Systems* (2022).
- [9] Vincent SITZMANN et al. « Implicit neural representations with periodic activation functions ». In : (2020).
- [10] WIKIPEDIA CONTRIBUTORS. *Radon Transform*. Wikipedia, The Free Encyclopedia. 2024. URL : https://en.wikipedia.org/wiki/Radon_transform.