



HAL
open science

Réseaux de neurones convolutifs en médecine nucléaire : applications à la segmentation automatique des tumeurs gliales et à la correction d'atténuation en TEP/IRM

Paul Blanc-Durand

► **To cite this version:**

Paul Blanc-Durand. Réseaux de neurones convolutifs en médecine nucléaire : applications à la segmentation automatique des tumeurs gliales et à la correction d'atténuation en TEP/IRM. Médecine humaine et pathologie. 2018. dumas-02116080

HAL Id: dumas-02116080

<https://dumas.ccsd.cnrs.fr/dumas-02116080>

Submitted on 30 Apr 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial - NoDerivatives 4.0 International License

AVERTISSEMENT

Cette thèse d'exercice est le fruit d'un travail approuvé par le jury de soutenance et réalisé dans le but d'obtenir le diplôme d'Etat de docteur en médecine. Ce document est mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt toute poursuite pénale.

Code de la Propriété Intellectuelle. Articles L 122.4

Code de la Propriété Intellectuelle. Articles L 335.2-L 335.10

UNIVERSITÉ PARIS DESCARTES
Faculté de Médecine PARIS DESCARTES

Année 2018

N° 199

THÈSE
POUR LE DIPLÔME D'ÉTAT
DE
DOCTEUR EN MÉDECINE

Réseaux de neurones convolutifs en médecine nucléaire :
applications à la segmentation automatique des tumeurs
gliales et à la correction d'atténuation en TEP/IRM

Présentée et soutenue publiquement
le 4 octobre 2018

Par

Paul BLANC-DURAND

Né le 3 mai 1989 à Paris (75)

Dirigée par M. Le Professeur Emmanuel Itti, PU-PH

Jury :

Mme La Professeure Aurélie Kas, PU-PH.....Présidente

M. Le Professeur Emmanuel Durand, PU-PH

M. Le Professeur Jérôme Hodel, PU-PH

M. Le Professeur Vincent Lebon, PU-PH

Table des matières

1	Introduction	6
1.1	Contexte scientifique	6
1.1.1	Radioactivité	6
1.1.2	Principales techniques d'imageries en médecine nucléaire	7
1.1.2.1	La tomographie par émission de positons . .	7
1.1.2.2	Les γ caméras	7
1.1.2.2.1	Le collimateur :	7
1.1.2.2.2	Le détecteur :	8
1.1.3	Principales techniques d'imageries anatomiques utilisées en médecine nucléaire	9
1.1.3.1	La TDM	9
1.1.3.2	L'IRM	9
1.1.4	Images et médecine nucléaire	10
1.1.4.1	Généralités	10
1.1.4.2	Traitement numérique des images	11
1.2	Neurobiologie et réseaux de neurones	12
1.2.1	Neurophysiologie et modélisation élémentaire	12
1.2.1.1	Le neurone biologique	12
1.2.1.2	Modélisation élémentaire du neurone	12
1.2.2	Réseaux de neurones	14
1.2.2.1	Architectures	14
1.2.2.1.1	Perceptron et Perceptron multi-couches :	15
1.2.2.1.2	Auto-encodeurs :	15
1.2.2.1.3	Réseaux Générateurs Adversatifs :	15
1.2.2.1.4	Réseaux de neurones récurrents :	15
1.2.2.1.5	Réseaux de neurones profonds avec connections résiduelles :	16
1.2.2.2	Types de problèmes résolubles par réseau de neurones	16

1.2.3	Apprentissage et régularisation	17
1.2.3.1	Apprentissage	17
1.2.3.2	Itérations	18
1.2.3.3	Sur et sous-apprentissage, Régularisation . .	19
1.2.4	Résumé des étapes	21
1.3	Vision et réseaux de neurones convolutifs	22
1.3.1	La vision	22
1.3.2	Principe général des réseaux de neurones convolutifs .	23
1.3.2.1	Les grands types de couches	24
1.3.2.1.1	Les couches de convolution : <i>Convolutional Layers</i>	24
1.3.2.1.2	Les couches de regroupements : <i>Pooling Layers</i>	25
1.3.2.1.3	Les couches entièrement connectées : <i>Fully connected layers</i>	25
1.4	Objectif de la recherche	26

2 Détection et segmentation automatique des tumeurs gliales en TEP à la ^{18}F -FET par réseau de neurones convolutifs de type U-net **27**

2.1	Contexte scientifique	27
2.2	Matériels et méthodes	29
2.2.1	Patients	29
2.2.2	Acquisition TEP à la ^{18}F -FET	30
2.2.3	Distribution normale de la ^{18}F -FET	30
2.2.4	Vérité terrain des masques de segmentation de ^{18}F -FET	31
2.2.5	Pre-processing	31
2.2.6	Implémentation	32
2.2.7	Architecture du réseau de neurone convolutif U-net .	32
2.2.8	Métriques	33
2.3	Résultats	35
2.3.1	Caractéristiques des patients	35
2.3.2	Détection	35
2.3.3	Segmentation	36
2.4	Discussion	38
2.5	Conclusion	41

3	Correction d'atténuation en TEP-IRM : évaluation de deux types de réseaux de neurones convolutifs	42
3.1	Introduction	42
3.1.1	Rappels physiques de l'atténuation des rayonnements électromagnétiques	43
3.1.2	Correction d'atténuation en TEP/TDM	43
3.1.3	Correction d'atténuation en TEP/IRM	43
3.1.3.1	À partir des données TEP	44
3.1.3.2	À partir des données IRM	44
3.1.4	Objectifs	46
3.2	Materiels et méthodes	46
3.2.1	Population	46
3.2.2	Acquisition TEP/IRM	47
3.2.3	Génération des AC map et reconstruction des images TEP	47
3.2.3.1	Architecture U-net et génération des pCT _{U-net}	48
3.2.3.2	Architecture Resnet et génération des pCT _{Resnet}	48
3.2.3.3	Génération des cartes de μ , post traitement et reconstruction TEP	50
3.2.4	Évaluation de la performance	51
3.2.4.1	Exactitude des pseudoCT	51
3.2.4.2	Impact des MRAC sur la quantification TEP	51
3.2.4.3	Quantification régionale des erreurs	52
3.2.4.4	Processing	53
3.3	Résultats	53
3.3.1	Exactitude des pseudoCT	53
3.3.2	Quantification régionale des erreurs	56
3.4	Discussion	57
3.5	Conclusion	63
	Conclusion	66
	Table des figures	70
	Liste des tableaux	72
	Bibliographie	73

Liste des Abréviations

- AC : Correction d'atténuation
- AAL : *Anatomic Atlas Labelling*
- BTV : *Biological Tumoral Volume* : Volume tumoral biologique
- CT : *Computed Tomography*
- CTAC : *Computed Tomography Attenuation Correction*
- CNN : *Convolutional Neural Networks*, Réseaux de neurones convolutifs
- DSC : *Dice Similarity Coefficient*, Coefficient de similarité de Dice
- FCL : *Fully Connected Layer*
- FDG : Fluoro-Déoxy-Glucose
- FET : Fluoro-Ethyl-Thyrosine
- FCL : *Fully Connected Layer*
- GAN : *Generative Adversarial Network*
- IQ : Inter-Quartile
- IRM : Imagerie par résonance magnétique
- keV : kiloelectron-volt
- MBq : MégaBecquerel
- MRAC : *Magnetic Resonance Attenuation Correction*
- ms : millisecondes
- ml : millilitres
- mV : milliVolts
- pCT : pseudoCT
- R^2 : Coefficient de détermination
- ReLU : *Rectified Linear Unit*
- ROI : *Region Of Interest* : Région d'intérêt
- RMSE : *Root Mean Square Error* : Erreur quadratique moyenne
- SUV : Standardized Uptake Value
- SPM : *Statistical Parametric Mapping*
- TDM : Tomodensitomètre
- TEP : Tomographie par Émission de Positron
- UH : Unités Hounsfields
- UTE : Ultra Short Time Echo
- VOI : Volume d'intérêt
- ZTE : Zero Time Echo

Chapitre 1

Introduction

1.1 Contexte scientifique

1.1.1 Radioactivité

La médecine nucléaire utilise des médicaments radiopharmaceutiques afin de tracer des molécules vectrices. Les isotopes radioactifs possèdent plusieurs propriétés fondamentales qui confèrent à nos images leurs particularités. En effet, les isotopes se désintègrent de manière aléatoire selon une constante λ propre à chaque isotope et avec une demi-vie $T_{1/2} = \ln 2/\lambda$. Le dénombrement de ces derniers sur un laps de temps fixé se comporte donc comme une variable aléatoire suivant une loi de Poisson. Les applications de médecine nucléaire en imagerie s'attèlent donc à estimer le nombre de désintégrations et à en préciser la localisation. La localisation peut être déduite par la trajectoire rectiligne des rayonnements électromagnétiques en l'absence d'interaction avec la matière (que ce soit grâce aux collimateurs des γ -caméras ou à l'estimation de la ligne de réponse en tomographie par émission de positons (TEP)). Par ailleurs, l'estimation du nombre de désintégrations est un comptage simple des coups, corrigé de la décroissance radioactive et de l'atténuation avec le milieu (essentiellement par effet photoélectrique et effet Compton). Ces estimations peuvent être calculées lors de la reconstruction tomographique grâce à l'opérateur de projection.

1.1.2 Principales techniques d'imageries en médecine nucléaire

1.1.2.1 La tomographie par émission de positons

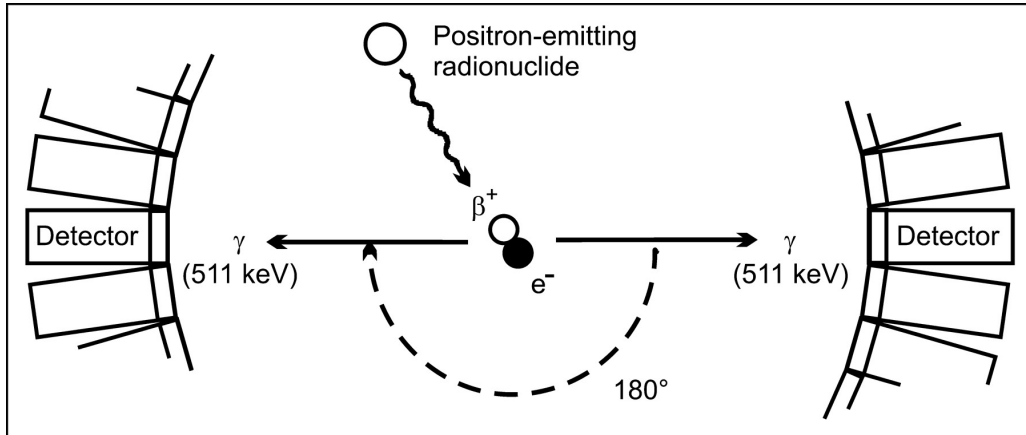
La TEP est une technique d'imagerie moléculaire capable de détecter et de compter les annihilations issues d'isotopes émetteurs β^+ . Ainsi une molécule vectrice marquée à un isotope β^+ , émettra lors de sa désintégration un positon. Une fois son énergie cinétique déposée dans le milieu environnant, elle entrera en collision avec un électron et émettra de façon simultanée lors d'une réaction dite d'annihilation, 2 photons anti-colinéaires de 511 kiloelectron-volts (keV) (les photons d'annihilations) à 180° l'un de l'autre. Ces photons sont détectés par une couronne de détecteurs disposés en anneau autour du patient, permettant d'estimer la localisation du lieu de désintégration. Les processus d'émission, d'annihilation, et de coïncidence sont représentés schématiquement dans la Figure 1.1. Les isotopes principalement utilisés pour l'imagerie TEP sont le fluor 18 (^{18}F), le gallium 68 (^{68}Ga).

1.1.2.2 Les γ caméras

De la même manière, après administration d'un médicament radiopharmaceutique émetteur γ , les γ -caméras peuvent estimer la localisation du photon incident et produire l'image scintigraphique. Les γ -caméras sont composées successivement d'un collimateur, de détecteurs et d'une électronique (Fig 1.2).

1.1.2.2.1 Le collimateur : Le collimateur est une galette de plomb ou d'un autre métal atténuant ayant un rôle de filtre restreignant les photons à

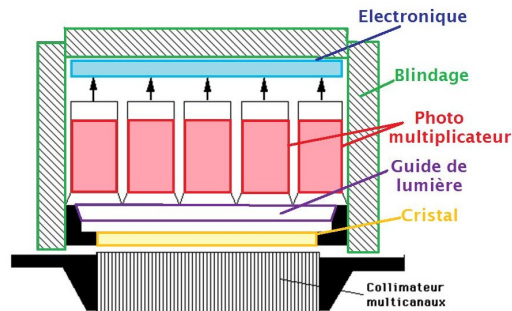
FIGURE 1.1 – Émission de positon, annihilation et détection de coïncidence par TEP (d'Iris Verel, The Promise of Immuno PET, JNM 2005)



ceux venant d'une direction donnée. Les collimateurs sont le plus souvent *parrallèles* mais peuvent être aussi *pinhole*.

1.1.2.2.2 Le détecteur : Le détecteur est lui même composé successivement d'un scintillateur souvent de NaI dopé avec des impuretés de Thallium, de guides de lumière et de photomultiplicateurs. Le scintillateur, après excitation par un rayonnement électromagnétique, émet des photons (phénomène de scintillation). Les électrons déplacés dans le cristal à la suite des interactions élémentaires perdent progressivement leur énergie en excitant les atomes de thallium qui émettent des photons lumineux lorsqu'ils se déséxcitent. Des guides de lumière puis des photomultiplicateurs amplifient le signal. En résulte une impulsion électrique négative pour chaque photon γ incident et dont l'amplitude est proportionnelle à l'énergie déposée dans le cristal par le rayonnement gamma initial.

FIGURE 1.2 – Les γ -caméras sont composées successivement d'un collimateur, de détecteurs et d'une électronique



1.1.3 Principales techniques d'imageries anatomiques utilisées en médecine nucléaire

1.1.3.1 La TDM

La TDM se compose d'un tube à rayons X rotatif. Les photons émis issus du rayonnement de freinage d'électrons (*Bremsstrahlung*), sont atténués lorsqu'ils traversent le patient puis sont récupérés par un détecteur opposé au tube. Les projections ainsi obtenues peuvent ensuite être reconstruites par différentes techniques comme par exemple la rétroprojection filtrée. La TDM mesure donc l'atténuation et estime la densité électronique des tissus. En médecine nucléaire elle possède un rôle à la fois anatomique pour mieux préciser la localisation et la morphologie des anomalies observées mais également afin d'estimer la cartographie d'atténuation pour la correction de cette dernière.

1.1.3.2 L'IRM

L'IRM est une technique d'imagerie dont le signal est produit par la résonance des noyaux d'hydrogènes, principalement de l'eau et des lipides. En fonction des séquences d'impulsion radiofréquence et des gradients, on peut pondérer sur des composantes biophysiques des tissus que sont le T1,

le T2 et la densité de protons. L'IRM offre un fort contraste tissulaire, et de nombreuses informations physiologiques (perfusion, diffusion). Combiner la TEP et l'IRM en un seul examen a été un défi technologique car les tubes photomultiplicateurs utilisés pour l'imagerie TEP sont sensibles au champs magnétique, de même l'électronique TEP peut induire des hétérogénéités du champs magnétique local perturbant le signal RMN [1]. Ce n'est qu'à partir de 2015 que des solutions ont pu être trouvées autorisant une imagerie TEP et IRM simultanée grâce à des photodiodes à avalanches (solution choisie par Siemens dans sa biograph mMR) ou des Photomultiplicateurs Siliconés (General Electric SIGNA PET/MR).

1.1.4 Images et médecine nucléaire

1.1.4.1 Généralités

On désigne par "image" la représentation d'une scène qui est le plus souvent planaire. Une image est donc un signal particulier dont le support est spatial et multidimensionnel (généralement à 2 dimensions). Pour parler de volume toutes les dimensions doivent être homogènes, à savoir qu'elles doivent être de la même nature. La grandeur physique dépend du détecteur utilisé pour obtenir l'image. Ainsi l'on sera confronté en médecine nucléaire à des décomptes de désintégrations (Bécquerel, s^{-1} , Bq), des concentrations (un nombre de coups par voxel, par exemple des MBq/ml), des temps (en IRM par exemple l'on mesure le T1, temps de relaxation longitudinal des tissus le long de l'axe B_0) ou des densités en scanner (Unité Hounsfields UH). De plus, les images en médecine nucléaire peuvent être bi-dimensionnelles comme pour la scintigraphie conventionnelle, tri-dimensionnelle (tomo-scintigraphie ou images TEP) voire quadri-dimensionnelles (dynamiques TEP). Enfin on distingue classiquement les images continues des images discrètes. Dans ces dernières il existe un système de coordonnées dépendant de l'échantillonnage et découpant l'espace spatial en pixels dans le cadre d'images 2D (*picture*

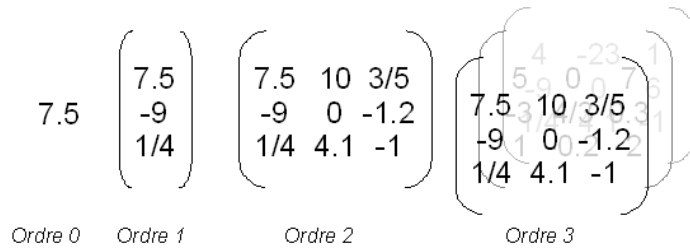


FIGURE 1.3 – Les tenseurs sont des objets mathématiques issus de l’algèbre linéaire permettant de généraliser les scalaires et les vecteurs, Wikipédia

element) ou en voxels pour (*volume element*) l’imagerie 3D. Cet échantillonnage est défini lors de la reconstruction, à priori et dépend notamment de la résolution du détecteur afin d’optimiser le rapport signal sur bruit.

1.1.4.2 Traitement numérique des images

Par abus de langage, on utilise souvent le terme de matrice pour désigner la représentation échantillonnée de nos images. Néanmoins, les matrices en mathématiques étant limitées à 2 dimensions, le terme de tenseur est préféré. Un tenseur désigne une généralisation mathématique du concept de vecteur. Ainsi un vecteur est un tenseur d’ordre 1, une matrice un tenseur d’ordre 2. Des exemples de tenseurs sont représentés en Figure 1.3.

Nos images sont donc l’association d’un tenseur et de métadonnées. Les métadonnées sont l’ensemble des informations permettant l’identitovigilance, la traçabilité, les paramètres d’acquisitions et de reconstructions. Enfin, le traitement numérique des images représente un ensemble de techniques permettant d’améliorer, de segmenter, de reconnaître, de représenter ou de compresser nos images.

1.2 Neurobiologie et réseaux de neurones

1.2.1 Neurophysiologie et modélisation élémentaire

1.2.1.1 Le neurone biologique

Le neurone est une cellule fonctionnelle spécialisée qui capte et transmet des signaux électriques à des muscles, des glandes ou à d'autres neurones et dont le corps (stroma) est doté de deux types de prolongements, soit des ramifications courtes et rayonnantes (les dendrites), soit une longue fibre qui se termine par une synapse (l'axone). Ces prolongements permettent d'établir de nombreuses connections. En effet, chaque neurone communique, en moyenne, avec 10^4 autres neurones et le cerveau humain comportant de l'ordre de 10^{11} neurones, il existe environ 10^{15} connexions. Par ailleurs, le neurone, du fait de sa membrane dialysante et des différences de concentrations intra et extracellulaires possède des membranes polarisées avec un potentiel de repos d'environ -70mV . Les dendrites jouent un rôle de réception de l'information dont le signal dépend de neuromédiateurs synaptiques. Il existe ensuite une intégration spatiale et temporelle des différents potentiels dendritiques représentés en Figure 1.4 qui si elle dépasse la valeur critique de -40mV , entraîne une inversion brutale du potentiel membranaire du stroma neuronal puis un rapide retour à l'état de repos en 3.5ms qui est ensuite transmise à l'axone. Cette cascade événementielle est appelée potentiel d'action et met en jeu des échanges ioniques du potassium et du sodium.

1.2.1.2 Modélisation élémentaire du neurone

Un modèle est une représentation simplifiée de la réalité. Le neurone est ainsi souvent représenté par une unité élémentaire qui calcule une somme pondérée de ses entrées et s'active selon une fonction [2]. L'unité modèle la plus simple est une unité qui produit la combinaison linéaire (z) de ses entrées (x) (Figure 1.5) par leurs poids (w) respectifs, auquel un biais (b) est ajouté (équation 1.1). La sortie de l'unité est la valeur $g(z)$ où g est

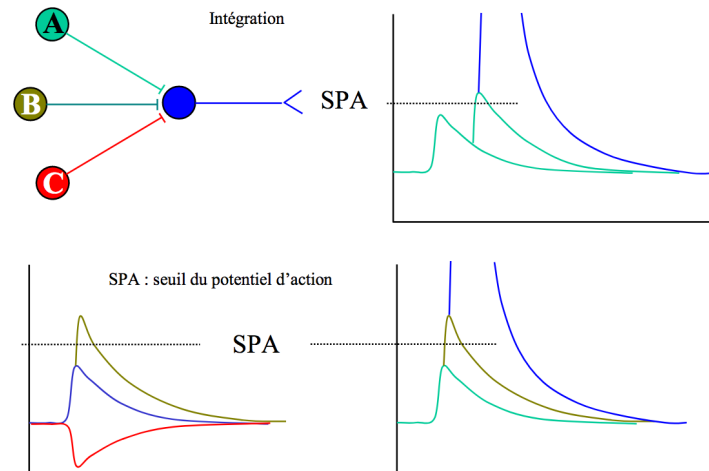


FIGURE 1.4 – Intégration synaptique des potentiels dendritiques

En haut à gauche : trois axones de neurones A,B,C interagissent avec un quatrième neurone.

En haut à droite : la sommation temporelle deux potentiels post-synaptique excitateurs dépassant le seuil d'activation est à l'origine d'un potentiel d'action.

En bas à gauche : la sommation spatiale de deux potentiels post-synaptique excitateurs et d'un potentiel post-synaptique inhibiteur ne dépasse pas le seuil d'activation.

En bas à droite : la sommation spatiale de deux potentiels post-synaptique excitateurs dépassant le seuil d'activation est à l'origine d'un potentiel d'action.

Issu de JM Mienville, unice.fr/JMM/Cours/Neuro

la fonction d'activation retenue (équation 1.2). La fonction d'activation sert à définir une partition, un seuil pour séparer la population. Elle est classiquement non linéaire car la combinaison linéaire de fonctions linéaires est une fonction linéaire, ce qui empêchera le futur réseau d'apprendre des fonctions plus complexes. Ainsi en utilisant des fonctions d'activation non linéaires on peut modéliser quasiment n'importe quelle fonction, sous réserve que la taille du réseau de neurones soit suffisamment grand [3].

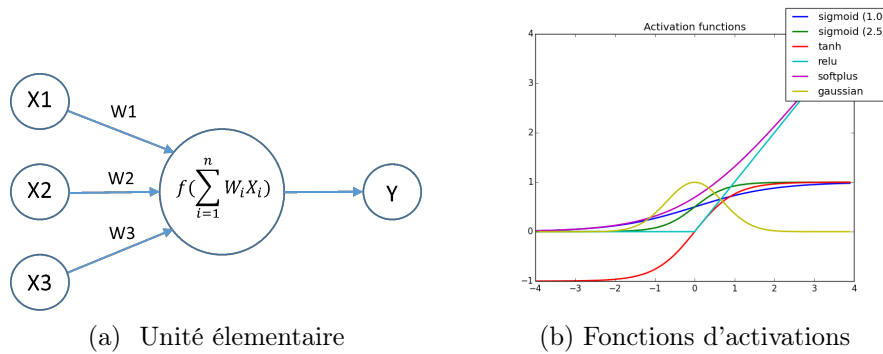


FIGURE 1.5 – Une unité neuronale typique calcule la combinaison linéaire z de ses entrées x_i en utilisant les poids w_i . La sortie y de l'unité est une fonction de z , appelée fonction d'activation. Les fonctions d'activation populaires incluent les fonctions sigmoïdes (tangente hyperbolique représentée ici en rouge) ou les fonctions linéaires rectifiées (ReLU, turquoise).

$$z = \sum_{i=1}^n w_i \times x_i + b \tag{1.1}$$

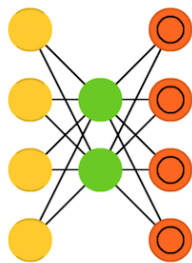
$$y = g(z) \tag{1.2}$$

1.2.2 Réseaux de neurones

1.2.2.1 Architectures

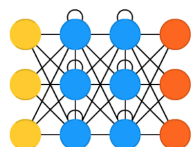
Une combinaison de ces neurones et des liens entre ces unités définit le réseau de neurones.

1.2.2.1.1 Perceptron et Perceptron multi-couches : Le perceptron [4] est le réseau le plus simple et possède au moins deux neurones d'entrée et un neurone de sortie. Il peut être utilisé pour modéliser des portes logiques ou pour entraîner des séparateurs linéaires. Le perceptron multi-couches est composé de multiples couches "cachées", généralement entièrement connectées aux couches adjacentes. Il permet d'apprendre des fonctions plus complexes.



1.2.2.1.2 Auto-encodeurs : L'idée de base derrière les auto-encodeurs est d'encoder des informations et d'apprendre une représentation (encodage) d'un ensemble de données. L'ensemble du réseau ressemble à un sablier, avec des couches cachées plus petites que les couches d'entrées et de sorties. La plus petite couche est presque toujours au milieu et représente l'endroit où l'information est la plus compressée (le *chokepoint* du réseau). La première moitié est dénommée encodage, la seconde moitié décodage et le milieu le code.

1.2.2.1.3 Réseaux Générateurs Adversatifs : Les GAN (*generative adversarial networks*) [5] sont constitués de deux réseaux travaillant en parallèle, l'un ayant pour tâche de générer du contenu (le générateur) et l'autre d'en juger la qualité (le discriminateur). L'objectif des réseaux génératif est d'augmenter le taux d'erreur du réseau discriminant (c'est-à-dire de "tromper" le discriminateur en produisant de nouveaux cas qui semblent provenir de la vraie distribution de la population d'entraînement).

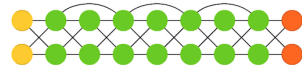


1.2.2.1.4 Réseaux de neurones récurrents : Les réseaux de neurones récurrents sont des types de réseaux de neurones présentant des connections récurrentes repré-

sentées par des boucles. Elles correspondent à des connexions au travers du temps dans l'espace des poids. Les données d'entrées qui sont fournies aux neurones dépendent à la fois des poids des couches précédentes mais également des valeurs des poids antérieurs. Ils sont donc particulièrement utilisés pour l'analyse des structures dynamiques temporelles.

1.2.2.1.5 Réseaux de neurones profonds avec connections résiduelles :

Les réseaux de neurones profonds avec connections résiduelles[6] sont des réseaux de neurones avec des connections résiduelles (*skip-connection*) permettant de sauter des couches profondes. De cette manière le réseau est obligé d'apprendre à partir de données d'entrées et de sorties. Ce type de réseau permet d'entraîner des réseaux beaucoup plus profonds et ainsi d'élargir le champs de réception.



1.2.2.2 Types de problèmes résolubles par réseau de neurones

Les réseaux de neurones permettent de résoudre deux grands types de problèmes :

- Les problèmes de classifications, où l'on cherche à prédire une classe, une catégorie. L'on distingue classiquement les classifications binaires (deux classes à prédire), des classifications "multi-class" où plusieurs classes sont à prédire.
- Les problèmes de régressions, où l'on cherche à prédire une fonction continue.

Cette distinction est importante car pour chaque type de problème, l'évaluation des performances de ce réseau se fera de manière différente. Cette mesure de performance, également appelée fonction de coût "*cost-function*",

reflète de combien se trompe le réseau de neurone pour la tâche qui lui est affectée. Par exemple dans le cadre de problèmes de classifications la *log-loss* 1.3 ou l'entropie-croisée est classiquement utilisée.

$$J = - \sum_{c=1}^M y_{o,c} \log p_{o,c} \quad (1.3)$$

J est la fonction de coût, M représente le nombre de classes à prédire, $y_{o,c}$, un indicateur binaire valant 0 ou 1 en fonction de si la classe c est correctement prédite pour l'observation o et $p_{o,c}$ la probabilité prédite pour chacune des classes de l'observation o . Ce genre de fonction de coût pénalise donc particulièrement les résultats qui se trompent et sont sûrs d'eux.

1.2.3 Apprentissage et régularisation

1.2.3.1 Apprentissage

L'apprentissage est le mécanisme par lequel le système nerveux s'adapte au monde extérieur. Les mécanismes précis de l'apprentissage par le cerveau restent très largement incompris et impliquent à la fois des modifications génomiques en partie sous le contrôle d'OTX2 [7] (méthylations des citosines, rétrotransposons) , des modifications neuronales (augmentation de la surface d'échange des synapses, du nombre de récepteurs en post synaptique ou du nombre de neuromédiateurs en pré-synaptiques) mais aussi du micro-environnement où les pieds astrocytaires jouent également un rôle primordial [8].

L'apprentissage en machine learning est très probablement fondamentalement différent de celui du cerveau. En effet, aujourd'hui les principales techniques d'apprentissage sont dites supervisées, à savoir que l'on fournit à l'algorithme des vérités terrains. Notre cerveau lui semble apprendre de manière largement non supervisé. Une des définitions historiques de Tom

Michels de l'apprentissage machine disait qu'un système était intelligent si pour une tâche T mesurée par une performance P, P s'améliore à la tâche T avec l'expérience E. Les récentes avancées en calcul parallèle ont permis dans de nombreuses applications visuelles d'égaliser voire de dépasser les capacités humaines pour la tâche qui leur est affectée. Néanmoins l'enjeu actuel est la transférabilité de ces connaissances et de passer de réseaux dits discriminatifs à des réseaux génératifs en les intégrant dans des représentations plus larges que de simples tâches.

L'apprentissage machine se fait essentiellement par itérations successives en cherchant à minimiser l'erreur globale d'un modèle mesurée par sa fonction de coût. Ce type de problème se réfère à l'ensemble des techniques d'optimisation et plus particulièrement de minimisation. Une des techniques de minimisation les plus utilisées en apprentissage machine est la descente de gradient où les erreurs sont rétropropagées aux réseaux de neurones au cours d'itérations successives. L'idée de la rétropropagation est de déterminer pour chaque poids, combien une petite variation de ce dernier impactera l'erreur globale. Le poids est ensuite ajusté en proportion de son impact sur l'erreur. On parle de gradient (de combien l'erreur varie) et est estimé comme la dérivée de l'erreur par rapport au poids. La descente de gradient regarde seulement au voisinage dans l'espace des poids et ne garantit pas de trouver ni un minimum global ni la solution optimale. Elle peut cependant trouver des solutions qui marchent suffisamment bien en pratique.

1.2.3.2 Itérations

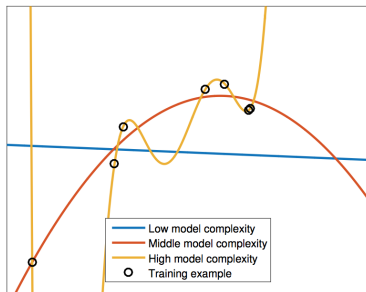
Quand les données sont trop volumineuses pour être traitées en même temps, on divise la population d'entraînement en lots (*batch*). Seuls les données du *batch* sont présentées au réseau de neurones et les poids sont adaptés en fonction de ces derniers. L'on parle d'une époque (*epoch*) lorsque toutes les données du *batch* ont été présentées. Le nombre d'itérations par époque

est donc le nombre nécessaire de *batch* pour compléter l'époque (nombre de patients dans l'entraînement divisé par le nombre de patients dans le *batch*).

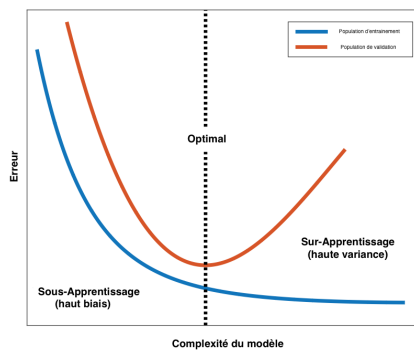
1.2.3.3 Sur et sous-apprentissage, Régularisation

Une des principales craintes lors de l'entraînement d'un réseau de neurones est le risque de sur-apprentissage ou de sous-apprentissage. En effet, un réseau de neurones, peut approximer à peu près n'importe quelle fonction sous réserve qu'il contienne suffisamment de paramètres. La complexité du modèle dépend à la fois du nombre de paramètres à entraîner mais aussi du nombre d'itérations réalisées. Néanmoins, le réseau de neurones n'apprenant pas sur la population globale mais sur un échantillon de cette dernière, pourrait ne pas être en mesure de généraliser les résultats. Ainsi, il existe un équilibre délicat entre le risque de sur-apprentissage et l'optimisation des performances du modèle. Cet équilibre fait souvent référence à l'expression du "compromis biais-variance". En effet, si le modèle ne possède pas assez de paramètres il ne sera pas en mesure d'apprendre de l'échantillon, et par voie de conséquence de prédire correctement dans une population de validation. On parle dans ce cas de *haut biais*, car l'erreur de prédiction dans cette population de validation est grande. En revanche si ce dernier contient trop de paramètres, à partir d'un certain moment, le modèle ne s'ajustera qu'en fonction du bruit qui par définition n'est absolument pas corrélé avec le signal mesuré. On observe alors une diminution de l'erreur dans la population d'entraînement (par sur-apprentissage) et une augmentation dans la population de validation. On parle alors de *haute variance*.

Un des moyens utilisés en apprentissage machine pour contrôler ce nombre de paramètres est la régularisation qui est une pénalisation soustraite à la fonction de coût. Les deux grands types de régularisation sont respectivement la régularisation $\mathcal{L}1$ et $\mathcal{L}2$ chacune ayant des spécificités propres dans le processus d'optimisation.



(a) Trois modèles ajustent leurs fonctions pour minimiser leurs erreurs : En vert, le modèle bien que passant par tous les points est trop complexe, et ne sera probablement pas en mesure de généraliser les résultats dans une population de validation. En bleu, le modèle semble trop simple et n'arrive pas à s'ajuster correctement dans la population d'entraînement.



(b) Concept du compromis biais-variance : Évolution de l'erreur en fonction de la complexité d'un modèle dans une population d'entraînement et dans une population de validation

FIGURE 1.6 – Le compromis biais-variance.

$$\mathcal{L1} = \sum_{i=1}^n |w_i|$$
$$\mathcal{L2} = \sum_{i=1}^n w_i^2$$

La régularisation $\mathcal{L1}$ aura tendance à rendre la matrice des poids W sparse pendant l'optimisation ce qui signifie qu'elle introduira des zéros dans le modèle qui dépendra ainsi de moins de paramètres. La régularisation $\mathcal{L2}$ quant à elle aura tendance à privilégier des poids plus petits. Le modèle final aura plus probablement une plus faible dépendance vis-à-vis de nombreux paramètres au lieu de dépendre fortement de quelques uns.

1.2.4 Résumé des étapes

Pour résumer et pour bien garder en tête les différentes étapes nécessaires à l'entraînement d'un réseau de neurones :

1. Choisir son architecture de réseau de neurone, son type de fonction de coût et de régularisation.
2. Initialiser les poids (au hasard, égaux à 0,..)
3. entraîner son réseau par itérations successives
 - (a) Implémenter la propagation "par l'avant" "*forward-propagation*" : calculer la fonction d'activation des produits entre les entrées et les poids au fur et à mesure des couches
 - (b) Calculer la fonction coût, en estimer sa dérivée par rapport au poids
 - (c) Implémenter la propagation arrière "*back-propagation*" : rétro-propagation des erreurs proportionnellement aux poids

1.3 Vision et réseaux de neurones convolutifs

1.3.1 La vision

L'oeil humain est extrêmement performant pour reconnaître des objets ou des scènes ce qui lui permet d'interagir facilement avec son environnement. Le cortex visuel est organisé de manière hiérarchisée, comme une succession de champs visuels corticaux ayant des organisations rétinotopiques, et de nombreuses connections "entrantes" ou "sortantes" entre les différents champs [9]. Ces connections suggèrent une organisation de bas vers hauts niveaux d'abstractions. Le cortex visuel (primaire et associatif) occupe environ 20% du cortex cérébral. Il est localisé dans le lobe occipital et s'étend aux parties postérieures des lobes temporaux et occipitaux. Pour qu'une région corticale soit considérée comme une carte de champs visuel, elle ne doit pas contenir plus d'une seule représentation pour chaque point du champ visuel ; c'est-à-dire que deux parties du cortex qui répondent préférentiellement à un même emplacement du champ visuel doivent se trouver dans des cartes de champs visuels distinctes [10]. A partir de ces propriétés une quinzaine de champs visuels corticaux ont pu être indentifié.

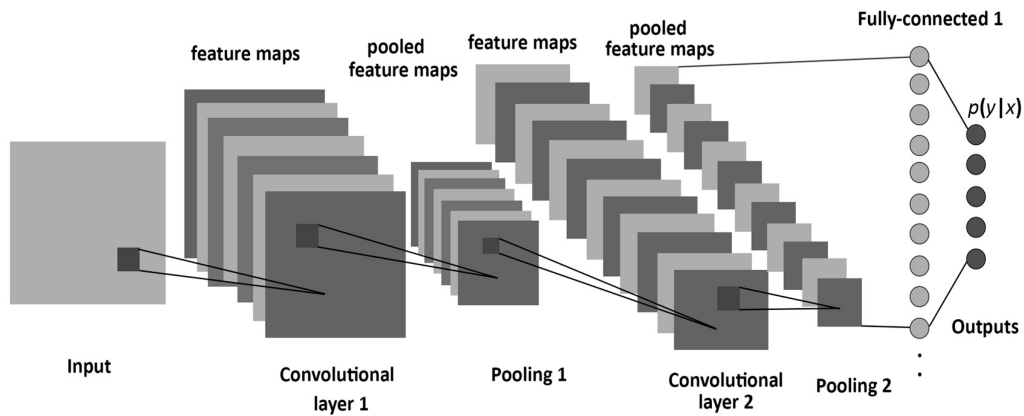
La rétine convertit les photons lumineux en un signal électrique qui est ensuite transmis au cortex visuel primaire par le nerf optique puis les radiations optiques. La première étape de traitement se situe dans le champs visuel V1 avec une correspondance bi-univoque entre un point sur la rétine et les centres de traitement pour ce point dans les cartes du champ visuel, de sorte que les points voisins de la rétine sont également traités à proximité dans les cartes du champ visuel. L'aire visuelle primaire (V1) est localisée dans et autour du sillon calcarin. L'image projetée sur la rétine est inversée et le centre du champs visuel (provenant de la fovéa) est considérablement élargi (grosissement cortical). L'image est par ailleurs conservée topographiquement intacte dans les étapes suivantes de traitement cortical [11]. L'hémi-champ visuel est projeté dans l'hémisphère controlatéral avec une topographie rela-

tivement systématique proposée initialement par Schwartz [12]. Les travaux pionniers de Hubel [13] sont les premiers à établir chez le chat une réponse préférentielle des neurones de V1 à différents angles et différentes positions dans le champs visuel. Les voies de la vision suivent ensuite deux voies celle du "*what*" inféro temporale et une voie du "*where*" pariétale [14]. La voie ventrale inféro-temporale, traite les couleurs et les formes de manière hiérarchique afin de les intégrer dans des représentations intermédiaires [15]. La voie dorsale quant à elle se dirige vers les aires temporales médiales et les aires temporales médiales supérieures et extrait des informations de mouvement. Ces voies se distinguent même plus en amont où dès la rétine des neurones spécialisés : les cellules magnocellulaires fournissent l'entrée au traitement du mouvement dans la voie dorsale. Ces cellules sont sensibles à de faibles luminances mais aussi à des fréquences spatiales et temporelles plus basses. Elles sont en revanche insensibles aux couleurs.

1.3.2 Principe général des réseaux de neurones convolutifs

Les réseaux de neurones convolutifs sont très similaires aux réseaux de neurones. Ils exploitent néanmoins une des caractéristiques importantes des images à savoir la distribution spatiale de l'échantillonnage. Un exemple d'architecture type de réseau de neurone convolutif est présenté en figure 1.7. Ils sont constitués successivement de couches de convolutions, de couches de regroupements, et de couches connectées. Le terme d'apprentissage profond *deep learning* se réfère aux nombreuses couches qui doivent être apprises au fur de l'entraînement. Les réseaux de neurones convolutifs ne sont pas directement inspirés de la biologie et reposent sur des algorithmes d'apprentissage qui peuvent fondamentalement différer de ceux des cerveaux biologiques. Cependant, ils apprennent des représentations internes qui ressemblent fortement aux idées que l'on imagine des représentations du cortex visuel.

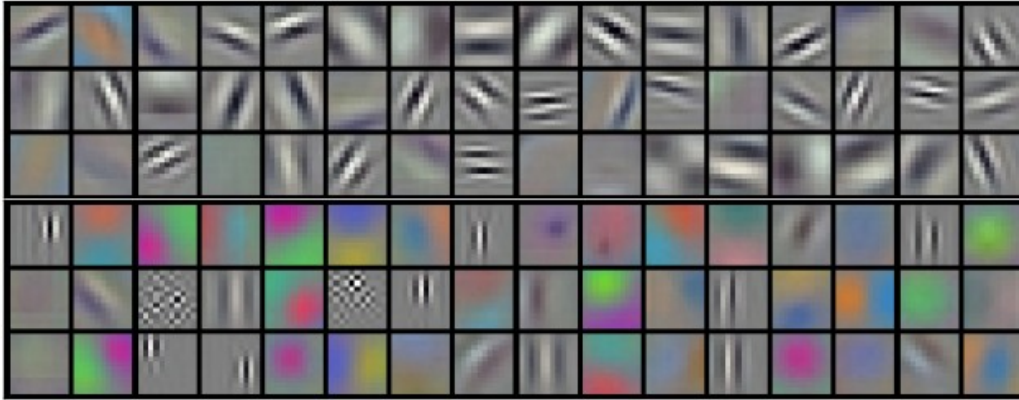
FIGURE 1.7 – Architecture classique d’un réseau de neurones convolutif. Une image est fournie en donnée d’entrée (input) et est convoluée avec des filtres (première couche de convolution) et dont les cartes d’activation sont regroupées et concaténées.



1.3.2.1 Les grands types de couches

1.3.2.1.1 Les couches de convolution : *Convolutional Layers* Les couches de convolution sont un ensemble de filtres qui sont appris au cours de l’entraînement. La taille et le nombre de ces filtres sont définis à priori. Par exemple : 96 filtres de $11 \times 11 \times 3$ sont appris pour la première couche convolutive de Krizhevsky [16] et sont représentés Fig 1.8. Au cours de la propagation avant, chaque filtre glisse (plus précisément est convolué) et les produits scalaires entre les entrées (par exemple les valeurs des pixel d’une image) et ces filtres sont calculés. Au fur et à mesure que le filtre se déplace sur l’image une carte d’activation est produite. Elle représente la réponse de ce filtre à chaque position spatiale. Ces cartes d’activations sont concaténées le long de la dimension $n+1$ où n représente le nombre de dimensions initiales de l’image pour former un nouveau tenseur. Comme précédemment ces cartes d’activations sont modulées par une fonction d’activation, classiquement une ReLU. On peut par ailleurs moduler le déplacement sur l’image de ces filtres (le *stride*). Ainsi la carte d’activation aura des dimensions dépendant à la fois des dimensions des filtres mais également du *stride*. On peut aussi ap-

FIGURE 1.8 – 96 filtres de convolution de taille 11x11x3 appris par la première couche de Krizhevsky et al[16].



plier à l'image un *zero-padding* consistant à agrandir l'image d'entrée par des zéros afin que la carte d'activation conserve les mêmes dimensions de l'images natives.

1.3.2.1.2 Les couches de regroupements : *Pooling Layers* Il est commun d'insérer périodiquement après les couches de convolutions des couches de *Pooling*. Ces fonctions sont prédéfinies et réduisent pour les couches ultérieures le nombre de paramètres à apprendre tout en élargissant le champs de réception (*receptive field*). Elles opèrent indépendamment aux différentes profondeurs du réseau et ne nécessitent pas de poids à entraîner. L'une des opérations classiques opérée est la fonction maximum, où au voisinage de N pixels seul le maximum est retenu dans la couche de regroupement.

1.3.2.1.3 Les couches entièrement connectées : *Fully connected layers* Les neurones dans ces couches sont tous connectés à l'ensemble des neurones des cartes d'activations précédentes. Ainsi, si l'issue d'une couche

de convolution ou de regroupement produit une carte d'activation de 512 filtres de 7x7 et que la couche de FCL contient 4096 neurones, il y'aura 512x7x7x4096 poids à entraîner. La FCL est la dernière couche et permet de faire une transition afin de transformer les cartes d'activations en probabilités. Ainsi si l'on cherche à prédire 10 catégories, la FCL contiendra 10 neurones. Ceci est permis grâce aux propriétés et aux liens entre les fonctions d'activations sigmoïdes classiquement utilisées sur ces couches et les lois de probabilités (notamment logistiques).

1.4 Objectif de la recherche

Notre objectif au travers de deux exemples est de tenter de mieux appréhender les potentialités de l'apprentissage profond en médecine nucléaire. Nous étudierons donc dans un premier temps une procédure de segmentation qui peut être assimilée à un problème de classification où l'on cherche à prédire pour chaque voxel une catégorie. Dans un second temps nous nous intéresserons à estimer des TDM à partir d'une séquence IRM. Ce type de problème peut être assimilé à une régression où l'on cherche à prédire une fonction continue à partir de nos images.

Chapitre 2

Détection et segmentation automatique des tumeurs gliales en TEP à la ^{18}F -FET par réseau de neurones convolutifs de type U-net

2.1 Contexte scientifique

Les gliomes font partie du groupe des tumeurs cérébrales les plus fréquentes [17] et comme la TEP au ^{18}F -Fluoro-Deoxy-Glucose (FDG), retrouve une activité physiologique intense du cortex sain, les acides-aminés radiomarqués (la ^{11}C -Methionine (^{11}C -MET), la ^{18}F -Fluoro-L-Dopamine (^{18}F -FDOPA) ou la ^{18}F -Fluoro-Ethyl-Tyrosine (^{18}F -FET) [18]) sont plus appropriés en raison du faible métabolisme cérébral physiologique de ces derniers et peuvent donc caractériser les tumeurs cérébrales avec un meilleur contraste. Ces radiotraceurs sont de plus en plus utilisés dans la prise en charge des tumeurs gliales que ce soit pour la recherche des diagnostics différentiels, dans l'évaluation de l'extension ou dans la planification du traitement [19]. Même si la TEP est une modalité de choix pour sa haute sensibilité, elle possède toutefois plusieurs limitations comme une faible résolution spatiale avec des effets de volume partiel et un relativement faible rapport signal sur bruit. Pour ces deux raisons, et du fait des larges variations de taille et de métabolisme des tumeurs gliales, la segmentation des tumeurs gliales en TEP est complexe.

La segmentation est une tâche courante en imagerie médicale qui peut être réalisée manuellement, semi-automatiquement ou automatiquement. Elle a plusieurs objectifs, comme par exemple mieux définir le volume tumoral macroscopique (*Gross tumoral volume*, GTV) ou le volume tumoral biologique (*Biological Tumor Volume*, BTV) pour la planification de la radiothérapie et le monitoring de la réponse thérapeutique. De plus, l'interprétation d'un examen à la ^{18}F -FET peut nécessiter la définition d'un volume d'intérêt (VOI). En effet, l'analyse des courbes temps-activité (*Time Activity Curves*, TAC) nécessite la définition d'un VOI, dont l'intérêt est d'aider à la gradation des tumeurs gliales [20, 21], à la caractérisation des phénotypes moléculaires [22] ou pour différencier la progression tumorale des radionécroses [23].

Les VOI des TEP à la ^{18}F -FET sont généralement réalisés en contournant manuellement une sphère centrée sur l'activité maximale. Un élagage (*pruning*) peut éventuellement être ajouté par seuillage fixe ou adaptatif. Les seuillages basés sur l'intensité sont couramment utilisés dans les segmentations semi-automatiques des ^{18}F -FET. Deux approches peuvent être utilisées soit en utilisant un seuil fixe (seuil fixe par exemple $\text{SUV} > 2.5$ ou défini par rapport à un bruit de fond) ou en utilisant un seuil adaptatif (par exemple 41% de SUV_{max}). Ces approches ne tiennent néanmoins pas compte des corrélations spatiales des voxels dans une image ce qui rend de telles méthodes sensibles au bruit, aux hétérogénéités locales de métabolisme et à l'effet de volume partiel. En outre, ces types de segmentations sont sensibles au seuil retenu ce qui peut conduire à de grandes différences de volume [24]. Enfin, le contourage manuel est soumis à une variabilité inter et intra-observateur en plus d'être consommateur de temps. Pour éviter ces problèmes, des méthodes de segmentation automatique des TEP ont été suggérées et récemment revues par Hatt et al. [25]. Certaines d'entre elles sont des méthodes 2D qui considèrent chaque coupe transaxiale TEP indi-

viduellement. D'autres sont des méthodes 3D qui semblent plus appropriées car elles considèrent les informations contextuelles du volume en entier. Les principales techniques 3D incluent la segmentation basée sur les gradients, les croissances de région (comme le fuzzy c-mean), les algorithmes statistiques, l'apprentissage machine et des techniques de segmentation basées sur la texture. Récemment, les progrès dans la vision par ordinateur et l'apprentissage automatique ont été appliqués à l'imagerie médicale. Les réseaux de neurones convolutifs (CNN) peuvent rassembler des informations locales autour d'un voxel et produire une carte de vraisemblance [26]. Nous évaluons ici la faisabilité d'un CNN de type U-net pour la segmentation des tumeurs gliales en TEP à la ^{18}F -FET.

2.2 Matériels et méthodes

2.2.1 Patients

Trente-sept patients ont été inclus rétrospectivement. Les caractéristiques des patients sont résumées en table 2.1. Tous les patients ont réalisé une TEP/TDM à la ^{18}F -FET au moment du diagnostic avant toute biopsie ou traitement (résection tumorale, chimiothérapie, radiothérapie). Les patients ayant nécessité une chirurgie rapide (<2 semaines) en raison d'un effet de masse ou d'une hémorragie intracérébrale, ainsi que les patients ayant des antécédents de biopsie cérébrale ou d'opération neurochirurgicale ont été exclus. Tous les patients ont réalisé ces examens d'imagerie en tant que soins courants et ont donné un consentement éclairé écrit avant l'inclusion. L'analyse rétrospective a été approuvée par le Comité local d'éthique de l'Etat de Vaud (2017-00758).

2.2.2 Acquisition TEP à la ^{18}F -FET

Les patients ont réalisé une TEP/TDM à la ^{18}F -FET dynamique sur une Discovery D690 temps de vol (27 patients) ou sur une Discovery LS (10 patients) (GE Healthcare, Waukesha, WI, USA) au sein du même service d'imagerie (CHUV, Lausanne, Suisse). Un jeûne d'au moins 4 heures avant l'injection de la ^{18}F -FET était recommandé selon les directives internationales. Après l'injection intraveineuse de 214 ± 25 MBq (de 145 à 295 MBq) de ^{18}F -FET, les images TEP ont été acquises en utilisant un protocole dynamique sur 50 minutes (10 images de 5 min). L'étalonnage pour les deux machines était le même. Les acquisitions ^{18}F -FET ont été reconstruites par méthode itérative de maximisation de l'espérance par sous ensembles ordonnés (OSEM, 3 itérations et 16 sous-ensembles) et incluaient un post-filtre gaussien (FWHM = 5 mm).

2.2.3 Distribution normale de la ^{18}F -FET

La distribution cérébrale normale de la ^{18}F -FET retrouve une captation relativement faible du bruit de fond cortical et de la substance blanche. Il existe un drainage et parfois une stase dans les compartiments veineux des sinus latéral, sagittal ou le sinus caverneux ainsi qu'un métabolisme physiologique de la graisse sous-cutanée ou péri-orbitaire. Des artefacts de bordure ont parfois été vus dans les limites de champs. Aucun faux positif n'était attendu de cet ensemble de données car les patients étaient inclus dans un essai clinique et toute hyperfixation dans la substance blanche ou grise du cerveau était considérée comme pathologique. Les volumes TDM n'étaient disponibles pour aucun patient et, par conséquent, la segmentation manuelle n'était basée que sur les données TEP. Tenant compte du fait que la captation maximale de la ^{18}F -FET ne correspondait pas nécessairement à la tumeur, afin de réaliser la segmentation des tumeurs cérébrales, nous avons décidé de suivre une procédure en deux étapes. Tout d'abord, une détection

de la tumeur gliale en utilisant un réseau convolutif de type U-Net, suivie d'une seconde partie de segmentation à proprement parler.

2.2.4 Vérité terrain des masques de segmentation de ^{18}F -FET

Les acquisitions dynamiques TEP ^{18}F -FET ont été réalignées temporellement sur la première acquisition dynamique puis normalisées spatialement dans le référentiel de l'Institut neurologique de Montréal (MNI - Université McGill, Montréal, Canada). Les dimensions des voxels résultants étaient de $2 \times 2 \times 2$ mm³. Une image de sommation de la 20^{ème} à la 40^{ème} minute après l'injection du radiopharmaceutique a servi de référence pour la segmentation manuelle ou l'entraînement. La normalisation spatiale a été réalisée à l'aide du logiciel SPM (SMP12) implémentée dans Matlab version R2015b (Mathworks Inc. , Sherborn, MA). Pour effectuer les vérités terrains, un masque de contour a été dessiné manuellement autour de la tumeur à l'aide du logiciel LifeX (<https://www.lifexsoft.com>, CEA, Saclay, France) [27, 28] et un seuillage semi-adaptatif a été fixé à 1,3 x le bruit de fond comme suggéré par Bette et al. [29, 30] et où le bruit de fond était défini comme la valeur moyenne d'une VOI sphérique de 10 cm³ dans l'hémisphère controlatéral.

2.2.5 Pre-processing

À des fins de calcul, toutes les images (masques et images de sommation) ont été redimensionnées avec une interpolation linéaire en un volume de $64 \times 64 \times 40$. En raison des limitations de résolution de la TEP et des effets de flou liées à la fonction d'étalement ponctuel (PSF : point spread fonction), nous avons également effectué un calcul d'amplitude de gradient de l'image native qui a été ajouté à l'image originale pour renforcer les contours. Chaque volume a été normalisé en intensité à partir de la moyenne et de la déviation

standard. Pour éviter le sur-apprentissage, nous avons également utilisé une stratégie d’augmentation des données afin d’élargir l’ensemble de données d’apprentissage. Ces opérations d’augmentation comprenaient des rotations (-30° à $+30^\circ$), des translations $(-0.1, 0.1)$, des cisaillements (-10° à $+10^\circ$) et des zooms $(0.85, 1.15)$. Puis pour évaluer la procédure de segmentation dans des conditions cliniques, l’ensemble de données a été séparé au hasard entre un groupe d’entraînement et un groupe de validation (représentant respectivement 70 et 30% de la population totale).

2.2.6 Implémentation

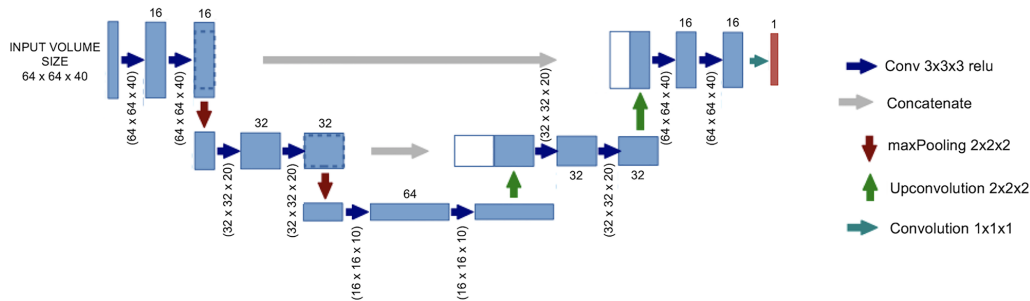
Notre réseau de neurones convolutif a été implémenté à partir d’une bibliothèque Keras supportant les opérations 3D. L’entraînement a été réalisé sur un poste de travail équipé d’une carte graphique NVIDIA TITAN X 1080 Ti. Le réseau de neurones a été entraîné par descente de gradient stochastique en utilisant un optimisateur de type ADAM. Le Dice (*Dice Similarity Coefficient*, DSC) a été utilisé comme mesure de similarité car il reflète l’efficacité de la procédure de segmentation. C’est également le DSC qui a été utilisé comme fonction de coût et était rétropropagée pour adapter les poids. La taille du batch était de 2 résultant en 13 itérations par époque. Le taux d’apprentissage (*learning rate*) a été défini initialement à 10^{-4} et le modèle a été entraîné jusqu’à 150 époques.

2.2.7 Architecture du réseau de neurone convolutif U-net

L’architecture du réseau suivait celle décrite par Ronneberger et al. [31] et est illustrée à la figure 2.1. Il se composait d’une voie d’encodage (côté gauche) et d’une voie de décodage (côté droit de la figure 2.1). Le chemin d’encodage suivait l’architecture typique d’un CNN avec l’application répé-

tée de deux convolutions successives de $3 \times 3 \times 3$ (avec un padding de 2), chacune suivie d'une activation par unité linéaire rectifiée (ReLU) et d'une opération de maxPooling de $2 \times 2 \times 2$. À chaque étape de sous-échantillonnage (d'encodage), le nombre de filtres a été doublé en commençant par 16 filtres à la première convolution. Chaque étape du chemin d'expansion consistait en un sur-échantillonnage de la carte des caractéristiques suivie d'une convolution de $3 \times 3 \times 3$. Au niveau de la couche finale, une convolution de $1 \times 1 \times 1$ a été utilisée pour mapper chaque carte d'activation au nombre souhaité de classes (ici 2 : tumorale et non tumorale). Au total, le réseau comptait 11 couches de convolution et 387 889 paramètres devaient être estimés. Le masque prédit par l'U-net a été suivi d'une dilatation morphologique avec une connectivité carrée de $3 \times 3 \times 3$. Nous avons également effectué un seuillage fixe, où le seuil a été fixé à 1,3 fois la valeur moyenne d'un *swap* hémisphérique du masque U-net prédit pour correspondre à la procédure qui avait été effectuée pour la vérité terrain.

FIGURE 2.1 – Architecture d'un auto-encodeur convolutif de type U-net comme décrit par Ronneberger et al [31]



2.2.8 Métriques

Afin d'évaluer la qualité de cette procédure de segmentation, nous avons calculé la mesure de similarité de Dice (DSC) qui a été utilisée à la fois

comme mesure de performance mais également comme fonction de coût. Le DSC a été calculé de la manière suivante où P correspondait à la prédiction de l'U-net et T la vérité terrain :

$$\text{Dice Similarity Coefficient} = \frac{2 \times P \cap T}{P + T}, \quad (2.1)$$

Les valeurs de sensibilité (Se), de spécificité (Sp), de valeurs prédictives positives (VPP), de valeurs prédictives négatives (VPN) ont été calculées à l'échelle du voxel dans la population de validation de 11 patients. Se, Sp, VPP et VPN ont été définis comme suit :

$$Se = \frac{VP}{VP + FN} \quad (2.2)$$

$$Sp = \frac{VN}{VN + FP} \quad (2.3)$$

$$VPP = \frac{VP}{VP + FP} \quad (2.4)$$

$$VPN = \frac{VN}{VN + FN} \quad (2.5)$$

VP (Vrais Positifs) : nombre de voxels correctement classifiés tels quels,
 VN (Vrais Négatifs) : nombre de voxels non tumoraux correctement classifiés
 FP (Faux Positifs) : nombre de voxels incorrectement classifiés comme tumoral
 FN (Faux Négatifs) : nombre de voxels incorrectement classifiés comme non tumoral.

2.3 Résultats

2.3.1 Caractéristiques des patients

Trente-sept patients ont été inclus dont 26 (71%) dans la population d'entraînement et 11 (29%) dans la population de validation. Les caractéristiques des patients inclus sont résumées en Table 2.1.

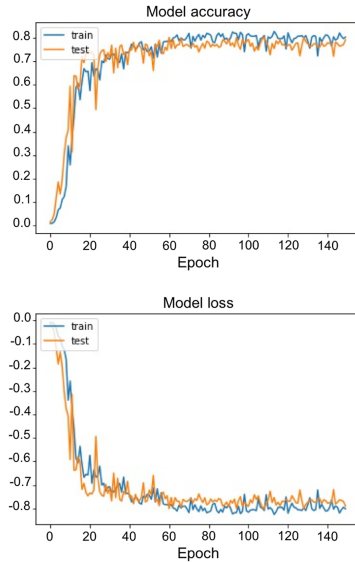
TABLE 2.1 – Caractéristiques de la population

Caractéristiques	Patients (n=37)
Âge (années)	44.7 (36 ;53)
Sexe Féminin	14 (37.8)
Délais entre la ^{18}F -FET et le diagnostic histologique (mois)	1.1 (0.5 ;1.7)
Données histopathologiques	
Astrocytomes Diffus	10 (27)
Astrocytomes Anaplasiques	9 (24.3)
Oligodendrogliomes	8 (21.7)
Oligodendrogliomes Anaplasiques	1 (2.7)
Gangliogliome	1 (2.7)
Glioblastomes	6 (16.2)
Gliomes Non Anaplasiques (pas de nature histologique précise)	2 (5.4)
Traitements	
Chirurgie	6 (16.2)
Radiochimiothérapie	11 (29.7)
Chimiothérapie	9 (24.3)
Chirurgie + Radiochimiothérapie	8 (21.6)
Aucun	3 (8.2)

2.3.2 Détection

Chez les 11 patients inclus dans la population de validation, toutes les lésions ont été détectées sans faux positif résultant en une sensibilité et une spécificité de 100% pour la détection tumorale.

FIGURE 2.2 – Courbes d’apprentissage avec évolution du DSC (en haut) et de la fonction de coût (en bas) au fil des époques .



2.3.3 Segmentation

Après 150 époques, le DSC a atteint 0. 7924 dans la population d’entraînement et 0. 7911 dans la population de validation. Les courbes d’apprentissage et l’évolution du DSC au fil des époques sont présentées en Figure 2.2.

La dilatation morphologique et le seuillage adaptatif du masque prédit par l’U-net ont substantiellement augmenté le DSC dans la population de validation à 0.8231 (+ 4.1%). À l’échelle du voxel, la matrice de confusion dans la population de validation est montrée dans la Table 2.2. Il en résultait une sensibilité de 0.88 [IC95%, 87.1 - 88.2%], une spécificité de 0.99 [99.9 - 99.9%], une VPP de 0.78 [76.9 - 78.3%] et une VPN de 0.99 [99.9 to 99.9%]. Ces résultats prennent en compte le grand nombre de voxels non tumoraux. Des exemples de segmentations sont montrés en figure 2.3.

FIGURE 2.3 – Exemples de segmentations avec à chaque fois : les images natives 18F-FET (en haut), la vérité terrain (milieu haut), le masque prédit (milieu bas) et la différence entre le masque prédit et la vérité terrain (en bas, avec les faux négatifs en rouge et les faux positifs en bleu)

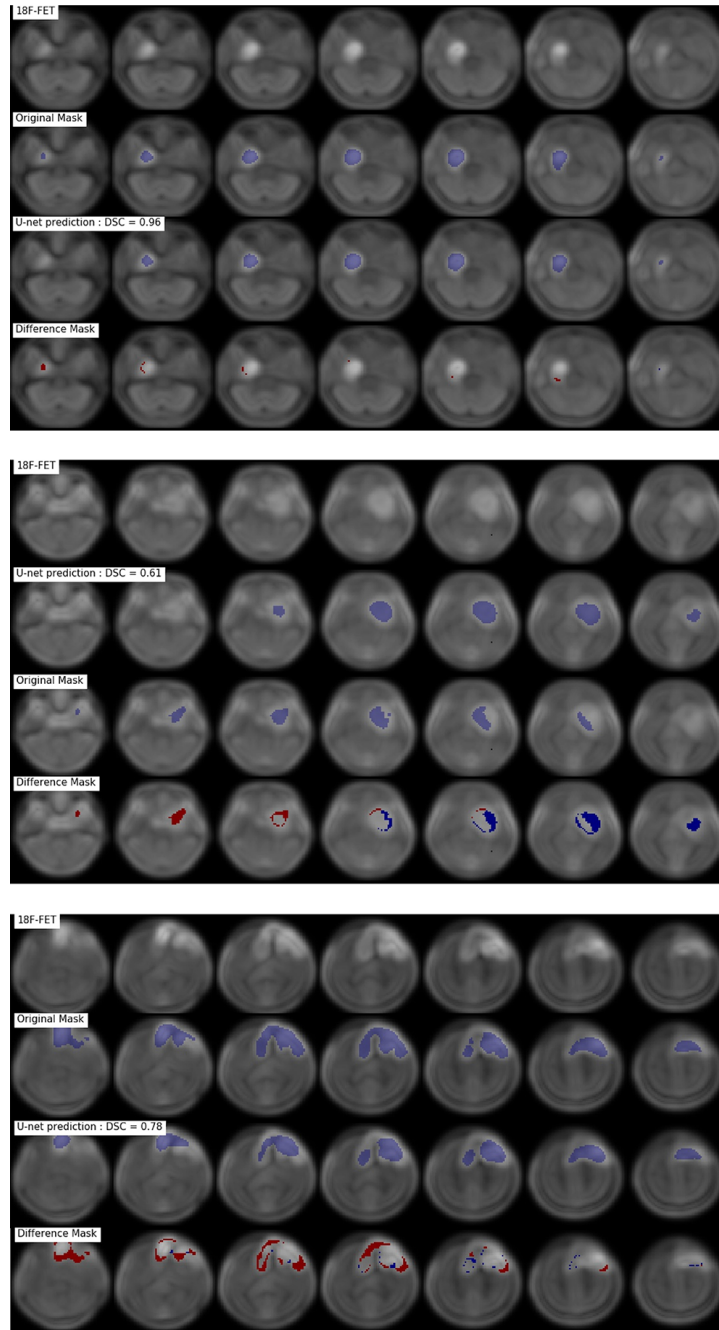


TABLE 2.2 – Table de contingence résumant les performances pour la segmentation des voxels tumoraux ou non la population de validation.

Voxel Prédit / Voxel vérité terrain	Tumoral	Non tumoral
Tumoral	10478	1426
Non tumoral	3025	1787261

2.4 Discussion

L’objectif de cette partie était d’implémenter et d’évaluer un CNN reprenant une architecture de type U-Net pour la détection et la segmentation automatique des tumeurs cérébrales en TEP à la ^{18}F -FET. Au terme de 150 époques nous avons réussi à détecter 100% des lésions tumorales et atteindre un DSC de 0.82.

Néanmoins comme le montre le nombre relativement élevé des faux négatifs dans la population de validation résultant en une sensibilité de 88%, nous avons cependant essayé des ressources additionnelles pour récupérer les voxels voisins qui ont été manqués par l’U-Net. Par conséquent, nous avons également essayé une approche de croissance de région et une segmentation par partage des eaux en utilisant des voxels aléatoires du masque U-net prédit. Les deux méthodes n’ont pas amélioré significativement le DSC, et même si ont diminué le nombre de faux négatifs ont eu un impact négatif important sur le nombre de faux positifs (données non montrées).

Deuxièmement, il est important de discuter de la génération de la vérité terrain. À cette fin, nous avons effectué une segmentation manuelle des ^{18}F -FET suivie d’un seuillage fixe, tel que pratiqué en routine pour l’estimation du BTV. Il n’y a pas de recommandations claires sur la manière dont le seuillage devrait être appliqué. Filss et al ont montré que le choix du seuil appliqué avait une grande importance sur le volume tumoral [35]. Il est à noter qu’en raison de la faible résolution et de la segmentation manuelle effectuée, l’absence de seuillage (fixe ou adaptatif) aurait entraîné un débordement de contour et pourrait également avoir diminué les performances.

Le choix d'un seuillage défini par rapport au bruit de fond controlatéral reposait à la fois sur des considérations cliniques mais aussi sur le fait que deux générations différentes de TEP ont été utilisées dans cette étude, ce qui aurait complexifié le choix d'un seuil fixe de SUV. Il est probablement vrai que des modifications de ces ROI de vérité terrain conduiraient à des différences substantielles. En effet, en raison des opérations de seuillage qui ne tiennent pas compte de la corrélation spatiale des voxels, certains voxels tumoraux peuvent avoir manqué la vérité comme le montre la figure 3B. Certains voxels négatifs étaient clairement à l'intérieur de la tumeur, mais n'étaient pas considérés comme tels, car le masque de référence était seuillé. En ce qui concerne le nombre de faux négatifs, deux raisons principales peuvent être avancées. Soit ils ont été manqués par le CNN U-net ou ont été omis après seuillage. Comme le masque original manuel et le masque U-net étaient suivis d'un seuillage, les différents niveaux de seuil entre la vérité terrain et le masque prédit (1.3 fois la valeur moyenne de 10 cm³ dans l'hémisphère controlatéral dans le cas du masque manuel, et 1.3 fois la valeur moyenne d'un échange hémisphérique du masque prédit par l'U-net) peut expliquer cette différence. Néanmoins, nous avons décidé d'appliquer ce seuil, car il diminuait le nombre de faux positifs de 3 049 à 1 476 sans impacter significativement le nombre de vrais positifs (de 10 051 à 10 478).

Une autre des limitations de ce travail comprend le nombre relativement faible de patients. Certains auteurs soutiennent en effet qu'une petite population peut être problématique dans l'apprentissage machine car une plus grande taille d'échantillon conduit généralement à une amélioration des performances [32]. Cette limitation peut néanmoins être partiellement corrigée par une stratégie d'augmentation des données [16]. Par conséquent, la stratégie principale d'augmentation des données utilisant des translations ou des cisaillements, comme elle risque d'augmenter le risque de sur-apprentissage, doit être envisagée en complément de stratégies de régularisations ou de *drop-out* (suppression des liens faibles). En outre, ces derniers semblent être

les plus efficaces pour améliorer les performances du modèle [33]. D'autres stratégies d'augmentation de données auraient pu également être employées à l'instar des réseaux adversatifs générateurs (où deux réseaux de neurones se combattent l'un contre l'autre, le premier (le générateur) crée un volume d'image fictif et le second (le discriminateur) ce volume peut être "statistiquement" tiré de la distribution de l'ensemble des données [34]. Si le discriminateur est trompé et que le volume est considéré comme un volume réel, ces images peuvent être utilisées pour l'entraînement. Nous avons également procédé à un ré-échantillonnage des volumes TEP à des fins de calcul, ce qui néanmoins entraîne généralement une perte d'information et peut dégrader les performances du réseau de neurones. Cette dernière étape était cependant obligatoire en raison des limitations de notre processeur graphique (GPU).

À l'avenir, nous pensons que ces résultats peuvent encore être améliorés à plusieurs niveaux. Premièrement, certaines opérations de pré-traitement afin de standardiser les données pourraient être utilisées telles que l'utilisation de l'égalisation d'histogramme (par exemple par des techniques telles que l'égalisation de l'histogramme adaptatif avec limitation de contraste : CLAHE) afin d'améliorer le contraste de ces images et de peut être mieux permettre la différenciation de l'arrière-plan du tissu tumoral. En effet, les pré-traitements de type CLAHE sont déjà utilisés dans les mammographies numériques pour améliorer le contraste pour l'oeil du radiologue, cela reste néanmoins une hypothèse que ces derniers amélioreront également les performances pour la vision par ordinateur. L'architecture de ce CNN pourrait également être adaptée si plus de patients étaient inclus en y intégrant des couches plus profondes. Enfin, à l'ère de l'imagerie hybride, l'on pourrait combiner les informations simultanées TDM / IRM en autorisant des données d'entrées multi-canaux.

2.5 Conclusion

Cette étude est la première à proposer une approche entièrement automatique pour segmenter les tumeurs gliales en TEP à la ^{18}F -FET et en utilisant un réseau de neurones convolutif de type U-net. Nous avons atteint de relativement bonnes performances compte tenu des limites de la machine et de la technique.

Chapitre 3

Correction d'atténuation en TEP-IRM : évaluation de deux types de réseaux de neurones convolutifs

NB : cette partie a été réalisée sous la direction du Pr Aurélie Kas lors de mon semestre à la Pitié Salpêtrière.

3.1 Introduction

Le 18-F-Fluoro-déoxy-glucose (FDG) est de plus en plus utilisé dans le diagnostic et le suivi des patients atteints de pathologies neurodégénératives car cet examen peut quantifier la perte neuronale classiquement retrouvée dans ces pathologies. En effet, il existe des anomalies régionales du métabolisme glucidique cortical permettant de diagnostiquer avec une bonne sensibilité la maladie Alzheimer [36]. Le FDG pourrait également être en mesure de distinguer les maladies d'Alzheimer d'autres démences corticales [37, 38]. L'intégration récente de la TEP et de l'IRM en une seule modalité offre l'opportunité aux médecins de conforter leurs diagnostics et de faciliter la prise en charge des patients en réalisant ces deux examens sur une même session [39]. Un des principaux défi technologique des TEP/IRM combinées est d'obtenir une estimation précise de la cartographie d'atténuation (CA) de la TEP et donc de trouver un substitut fiable à la CA standard basée sur

la tomodensitométrie (TDM).

3.1.1 Rappels physiques de l'atténuation des rayonnements électromagnétiques

Les rayonnements électromagnétiques subissent une atténuation avec le milieu qu'ils traversent et qui dépend essentiellement de la densité électronique Z de ce dernier mais également de l'énergie des photons. Les deux principaux modes d'interactions aux domaines d'énergies en imagerie médicale et en médecine nucléaire sont l'effet photoélectrique et la diffusion Compton. Le coefficient global d'atténuation linéaire μ représente la probabilité globale d'interaction par unité d'épaisseur de matériau traversé et s'exprime en cm^{-1} .

3.1.2 Correction d'atténuation en TEP/TDM

Historiquement les corrections d'atténuation étaient réalisées par un "scanner" de transmission à l'aide d'une source de $^{68}\text{Germanium}/^{68}\text{Gallium}$. L'arrivée de TDM intégrées à la TEP a permis d'estimer à partir des données TDM la carte des μ (*Computed Tomography Attenuation Correction : CTAC*). En effet, les images TDM sont une carte d'atténuation à un niveau d'énergie différent de ceux des photons d'annihilation car liés au spectre énergétique émis par le tube à Rayons X. Il a été démontré par Carney et al. qu'une simple transformation bilinéaire des voxels du TDM était suffisamment fiable pour estimer la carte d'atténuation à un autre niveau d'énergie [40].

3.1.3 Correction d'atténuation en TEP/IRM

En TEP/IRM la cartographie d'atténuation ne peut être précisément mesurée mais peut être estimée de plusieurs manières. En effet, en IRM le signal mesuré n'est pas linéairement proportionnel à la densité proto-

nique. De plus, du fait de la structure cristalline riche en hydroxyapatite ($\text{Ca}_5(\text{PO}_4)_3(\text{OH})$) de la corticale osseuse, ces derniers contiennent peu de noyaux d'hydrogènes à l'origine d'un faible signal. Or, la densité osseuse et liée à cette concentration d'hydroxyapatite et donc les structures les plus atténuantes sont également celles avec le moins de signal en IRM conventionnelle.

Deux grandes classes de techniques peuvent être utilisées pour estimer la cartographie d'atténuation : celles estimant la carte des μ à partir de données TEP ou celles l'estimant à partir de séquences IRM. De plus, de par la manière particulière de recueillir le signal en IRM en dedans d'antennes réceptrices, une partie de l'atténuation ne peut pas être visualisée car l'IRM est aveugle en dehors de ces dernières. Il est donc indispensable d'ajouter au cours de la reconstruction le matériel externe (antennes de radiofréquence et de réception, cages de gradients, tables) [41] .

3.1.3.1 À partir des données TEP

L'on distingue classiquement les corrections d'atténuation par transmission où une source externe d'activité connue permet l'estimation de cette dernière et les corrections d'atténuation par émission où ce sont les données TEP qui sont utilisées. La correction d'atténuation par transmission utilise des sources précédemment citées et participe à une irradiation additionnelle du patient. La correction d'atténuation par émission peut intervenir lors de plusieurs étapes à savoir lors de la mesure du temps de vol [42] ou lors de la reconstruction [43, 44].

3.1.3.2 À partir des données IRM

La correction d'atténuation à partir de l'IRM (MRAC) s'attèle à créer une TDM de substitution à l'aide de séquences IRM. À ce jour la correction d'atténuation en TEP/IRM repose essentiellement sur les techniques par

atlas ou par segmentations. Dans l’atlas, il existe une coregistration initiale d’un modèle de TDM à l’IRM du patient. Ces méthodes ont montré certaines limites lorsqu’elles sont appliquées à des anatomies inhabituelles, par exemple après une chirurgie crânienne ou un traumatisme. Ces approches dépendent également fortement de la qualité de la coregistration du modèle aux IRM individuelles. Il existe également des approches multi-atlas. Les techniques de corrections d’atténuation par segmentation segmentent à partir de séquences IRM les 4 principales classes de tissus (air, os, graisse, tissu mou). Ces classes sont ensuite remplacées avec des coefficients d’atténuation fixes et prédéfinis. La séquence la plus utilisée aujourd’hui pour la segmentation est la DIXON 2 points. Elle peut être complétée par des modèles de TDM par atlas pour l’os ou par une séquence IRM à temps d’échos très courts (séquences UTE / ZTE) qui fournit des informations osseuses supplémentaires. L’avantage des séquences à temps d’écho court type UTE ou ZTE est de fournir un signal dans les tissus avec des temps $T2^*$ courts tels que l’os [45].

Ces dernières années l’apprentissage profond a rencontré un large intérêt parmi la communauté scientifique notamment dans les domaines de la vision par ordinateur. Ces récents progrès ont été translatés à l’imagerie médicale par exemple pour la segmentation [31], le débruitage, pour la détection ou la classification des lésions et ont été récemment revues par Litjens [46]. Des applications de l’apprentissage profond ont trouvé leurs places dans le problème spécifique de la correction d’atténuation. Ce problème peut être formulé comme un problème de régression où l’on cherche à prédire par voxel à partir de séquences IRM la densité en unités Hounsfields ou directement le coefficient linéaire d’atténuation μ . Certaines équipes ont été pionnières dans le domaine en évaluant dans de petites cohortes la qualité des pseudoCT générés à partir de réseaux de neurones convolutifs et de séquences IRM [47, 48, 49]. Liu a entraîné un auto-encodeur convolutif pour segmenter des IRM pondérées en T1 en trois classes (air, os, tissu mou) puis a étudié l’impact sur

la reconstruction. Enfin, à notre connaissance, seules deux équipes ont utilisé la ZTE comme entrée afin de prédire les pseudoCT. Tout d’abord Gong et al. [50] ont construit et entraîné un 2D U-net à partir de séquences IRM cérébrales DIXON et ZTE en analysant l’impact sur la reconstruction chez 14 patients. Enfin Leynes a utilisé des IRM pelviennes multiparamétriques incluant la ZTE pour l’estimation de la correction d’atténuation [51].

3.1.4 Objectifs

De ces études sortent des performances variables, et la plupart n’ont pas utilisé de larges cohortes de validations. Ainsi, dans cette partie nous essaierons d’évaluer les performances de deux architectures de réseaux de neurones s’appuyant sur la séquence ZTE pour prédire la cartographie d’atténuation cérébrale dans une large cohorte de validation de 47 patients adressés dans le cadre d’un déclin cognitif en la comparant au "*silver standard*" (la TDM) et la technique actuellement implémentée en pratique clinique à savoir l’Atlas.

3.2 Matériels et méthodes

3.2.1 Population

Soixante treize patients consécutifs provenant du CHU Pitié-Salpêtrière, Paris, France entre juillet 2016 et décembre 2017 remplissaient les critères d’inclusions suivants : avoir réalisé une TEP/IRM (SIGNA PET/MR, GE Healthcare, Waukesha, WI, USA) dans un contexte de pathologie neurodégénérative, avoir réalisé une TDM sans injection de produit de contraste iodé, pas de chirurgie de la face ou de traumatisme entre la TDM et l’IRM. Vingt-trois patients ont servi pour l’entraînement, quarante-sept pour la validation. Trois patients des cinquantes initialement prévus ont été exclus pour les raisons suivantes (la co-registation a échoué chez un patient entre la TDM et la ZTE, un patient présentait sur sa TDM un hématome paren-

chymateux et un patient présentait sur son IRM un hématome sous dural). Les données de cette étude ont été approuvées par la commission nationale de l'informatique et des libertés (CNIL No. 2111722) et a été réalisée en accord avec les principes de la déclaration d' Helsinki.

3.2.2 Acquisition TEP/IRM

Une activité de 2 MBq/kg de FDG était injectée 30 à 45 minutes avant l'acquisition TEP/IRM. Les patients étaient mis au repos dans un environnement calme avec les yeux fermés pendant au moins 20 minutes après l'injection. Une TEP/IRM (SIGNA PET/MR, GE Healthcare, Waukesha, WI, USA) simultanée a en suite été acquise. L'imagerie par résonance magnétique durait 30 minutes et incluait entre autres une séquence 3DT1, 3D FLAIR, T2* et une diffusion. Le temps de balayage incluait un balayage d'émission TEP de 20 minutes. Pour corriger l'atténuation des photons, une séquence Dixon à deux points (LAVA-Flex) en pondération T1 a été acquise produisant des images d'eau, de graisse, en phase et en déphasage. Les paramètres suivants ont été utilisés : acquisition axiale ; TR = 4 ms ; TE = 1,12 et 2,23 ms ; angle de bascule 5 ; tranche d'épaisseur de 5,2 mm avec un chevauchement de 2,6 mm ; 120 coupes ; Taille de pixel de 1.95x1.95 mm². De plus, une IRM ZTE pondérée en densité protonique a été acquise avec les paramètres suivants : acquisition radiale centrée 3D ; taille de voxel 2.4x2.4x2.4 mm³ ; champ de vision (FOV) 26,4 x 26,4 cm² ; angle de bascule 0,8 ; bande passante \pm 62.5 kHz ; TR=390 ms ; TE=0 ms ; temps d'acquisition 40 s.

3.2.3 Génération des AC map et reconstruction des images TEP

Quatres cartographies d'atténuation étaient générées à partir des pseudoCT et ont été utilisées pour la reconstruction des images TEP. Tout d'abord un pseudoCT à partir d'un Atlas (pCT_{Atlas}) était créé par une co-registation non rigide d'une TDM et implémentée à partir de la console

de TEP/IRM [52]. Un pseudoCT était produit par chaque technique d'apprentissage profond respectivement pCT_{U-net} et pCT_{Resnet} . Les TDM disponibles chez tous les patients ont servi de référence afin d'évaluer les performances des différentes techniques de corrections d'atténuation. Les TDM provenaient de précédents examens TEP/TDM ou TEMP/TDM. Le délai moyen entre la TDM et la TEP/IRM était de 13.3 ± 18.7 mois. Les TDM ont été coregistrées sur la ZTE en maximisant l'information mutuelle dans le logiciel Statistical Parametric Mapping SPM12 (<http://www.fil.ion.ucl.ac.uk/>) implémenté dans MATLAB R2015b.

3.2.3.1 Architecture U-net et génération des pCT_{U-net}

Vingt-trois patients de notre institution ayant à la fois une ZTE et un CT co-registrés ont été ajoutés à la population d'entraînement du constructeur, les quarante-septs autres ayant servi à la population de validation. Des patchs de $64 \times 64 \times 64$ ont été extraits à la fois des ZTE et des CT pour des raisons computationnelles et ont été utilisées pour l'entraînement. L'architecture U-net suivait l'architecture classique décrite par Ronneberger [31] et supportait les opérations de convolution 3D et comprenait une voie d'encodage (4 couches) suivie d'une voie de décodage (4 couches) associée à des connection entre les deux voies. L'erreur quadratique moyenne entre le pseudoCT prédit et le CT a servi de fonction de coût et était rétropropagée au réseau.

3.2.3.2 Architecture Resnet et génération des pCT_{Resnet}

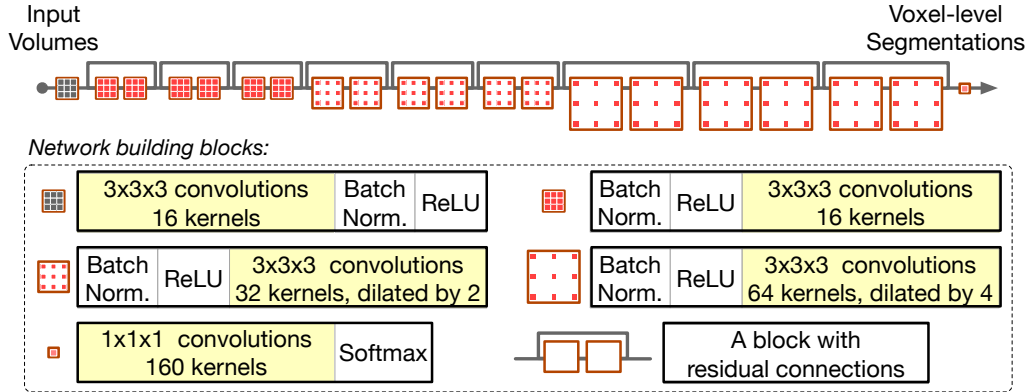
La plupart des architectures de réseau de neurones appliquées à l'imagerie médicale suivent un chemin classique d'encodage puis de décodage et font référence à un auto-codeur convolutif. Néanmoins, pour capturer des représentations globales du volume, des sous-échantillons successifs des carte d'activations intermédiaires sont réalisées en utilisant des opérations de re-

groupement (maxPooling), ce qui réduit la résolution spatiale. Cette résolution est partiellement récupérée par la partie de décodage mais à un coût de calcul élevé. Nous avons donc essayé une seconde architecture de réseau de neurones et exploitant des connexions résiduelles et adapté de Li et al. [53] où des convolutions dilatées sont également utilisées. Elles permettent d’augmenter significativement les champs de réception tout en limitant le nombre de paramètres à apprendre [54].

Les connexions résiduelles créent des connexions de mappage d’identité qui contournent les couches intermédiaires du réseau. Les entrées des blocs résiduels sont directement fusionnées aux sorties par concaténation. Les connexions résiduelles ont montré que la propagation de l’information était plus fluide et améliorait également la vitesse d’entraînement [55].

Le second réseau utilisé (Resnet) se composait donc de 20 couches de convolutions. Pour les sept premières couches convolutives, des convolutions successives de 3x3x3-voxels ont été employées. Ces couches sont conçues pour capturer des fonctions d’image de bas niveau telles que les arêtes et les coins. Dans les couches convolutionnelles suivantes, les filtres sont dilatés d’un facteur deux ou quatre. Les connexions résiduelles sont utilisées pour regrouper deux couches convolutives. Une représentation schématique du réseau résiduel original peut être vue dans la figure 3.1. Ce dernier réseau a été légèrement modifié en ajoutant une unité de biais à la dernière couche et en retirant la normalisation pour prédire une fonction continue s’étalant sur des grands écarts (de -1000 à +1000 pour les unités Hounsfields). L’erreur quadratique moyenne entre le pseudoCT prédit et le CT a aussi servi de fonction de coût et était rétropropagée au réseau proportionnellement aux poids. Une stratégie d’augmentation des données a été utilisée en utilisant des rotations et des agrandissements. Le réseau a été entraîné pendant 5 000 époques sur un GPU NVIDIA TITANx 1080Ti.

FIGURE 3.1 – Architecture Resnet utilisée par Li et al. [53]



3.2.3.3 Génération des cartes de μ , post traitement et reconstruction TEP

Tous les pseudoCT ont ensuite été processées en utilisant une solution interne dans MATLAB (MathWorks Inc., Natick, MA, USA) pour y inclure les antennes, le lit et afin de convertir les valeurs en unités Hounsfields (HU) des TDM (pseudoCT obtenus par les différentes techniques et CT) en coefficients d'atténuation linéaires μ à 511 keV à l'aide d'une transformation bilinéaire. Les images TEP ont ensuite été reconstruites avec les 5 cartes d'atténuation et les paramètres suivants : OSEM (algorithme de maximisation par sous-ensembles) 8 itérations et 28 sous-ensembles, correction TOF, modélisation de PSF, correction d'atténuation et du diffusé, taille de voxel de 1.17x1.17x2.78 mm³. Enfin 4 séries d'images TEP (TEP_{Atlas} , TEP_{U-net} , TEP_{Resnet} , and TEP_{CT}) étaient obtenues par patient.

3.2.4 Évaluation de la performance

3.2.4.1 Exactitude des pseudoCT

Pour évaluer la qualité des pseudoCT (pCT), nous avons tout d'abord étudié les performances de prédiction des structures osseuses car si erronées peuvent introduire d'importantes altérations de la quantifications car ce sont ces dernières qui sont les plus atténuantes. Nous avons donc d'abord utilisé une stratégie de seuillage fixe pour segmenter les structures osseuses à la fois des CT et des pCT. La mesure de similarité de Dice (*Dice Similarity Coefficient*, DSC) a été utilisé comme mesure de performance. Le DSC a été calculé comme suit où P correspondait au masque de prédiction de l'os par le pCT et T le masque osseux provenant du CT :

$$\text{DSC} = \frac{2 \times P \cap T}{P + T}, \quad (3.1)$$

Le DSC est une valeur numérique continue variant entre 0 et 1 et plus la valeur du DSC est proche de 1, meilleure est la segmentation.

3.2.4.2 Impact des MRAC sur la quantification TEP

La quantification des erreurs a été réalisée à la fois dans l'espace natif mais aussi dans le référentiel du Montreal Neurological Institute (MNI, McGill University, Montreal, Canada). La normalisation spatiale a été réalisée à l'aide de SPM12 implémentée dans Matlab version R2015b (Mathworks Inc., Sherborn, MA). Tous les volumes TEP ont été normalisés à partir de leur séquence IRM T1 dans la version non linéaire du MNI-ICBM 152. Les dimensions résultantes des voxels étaient de 2x2x2 mm³. Un masque a été appliqué à l'ensemble des voxels à partir de la cartographie paramétrique statistique de probabilité corticale de SPM et dont la probabilité d'être de la matière grise était > 20%. La métrique utilisée dans ce travail pour évaluer l'impact de la correction d'atténuation sur la quantification TEP a été

l'erreur TEP relative ΔSUVR définie comme :

$$\text{Erreur TEP relative : } \Delta\text{SUVR} = \frac{\text{TEP}_{\text{pCT}} - \text{TEP}_{\text{CT}}}{\text{TEP}_{\text{CT}}}, \quad (3.2)$$

où TEP_{pCT} était le volume the TEP reconstruit à partir des différents pCTs, and TEP_{CT} était le volume TEP reconstruit à partir du "*silver standard*" (ici la TDM). Les ΔSUVR ont été analysées à la fois dans le référentiel natif et dans le MNI sur tous les voxels dont la valeur de TEP_{CT} étaient différents de 0 et dans le MNI après normalisation spatiale et application du masque précédemment décrits. Nous avons également étudié les corrélations entre les CTAC et les différentes MRAC et mesurées par le coefficient de détermination : R^2 et l'erreur quadratique moyenne (*Root Mean Square Error*, RMSE).

3.2.4.3 Quantification régionale des erreurs

Nous avons calculé la moyenne et l'écart type de l'erreur TEP relative dans des régions spécifiques en utilisant les masques prédéfinis de l'AAL (*Automatic Atlas Labeling* : <http://cyceron.fr/freeware>) [56]) Les moyennes et écarts-types de l'erreur TEP relative chez les patients et pour chacune des méthodes ont été calculés pour toutes les ROI. À partir des 116 régions originelles de l'AAL, les régions cérébrales de même territoire anatomique ont été regroupées et les valeurs rapportées sont la moyenne pondérée par la taille du volume de la région étiquetée et résultant en un total de 70 régions cérébrales corticales et sous corticales. Les différences entre les MRAC et le CTAC ont été analysées par t-test apparié afin de rechercher les régions corticales dont le métabolisme était statistiquement sur ou sous-estimé. Le seuil de significativité α a été défini à $p < 0.05$ et $p < 0.01$, et corrigé des tests multiples par la technique de Bonferroni.

3.2.4.4 Processing

Toutes les données ont été processées en python 3.5 avec comme librairies principales numpy, scipy [57], pandas [58], matplotlib pour les graphiques [59] et nilearn pour la manipulation des volumes cérébraux et pour les images [60].

3.3 Résultats

3.3.1 Exactitude des pseudoCT

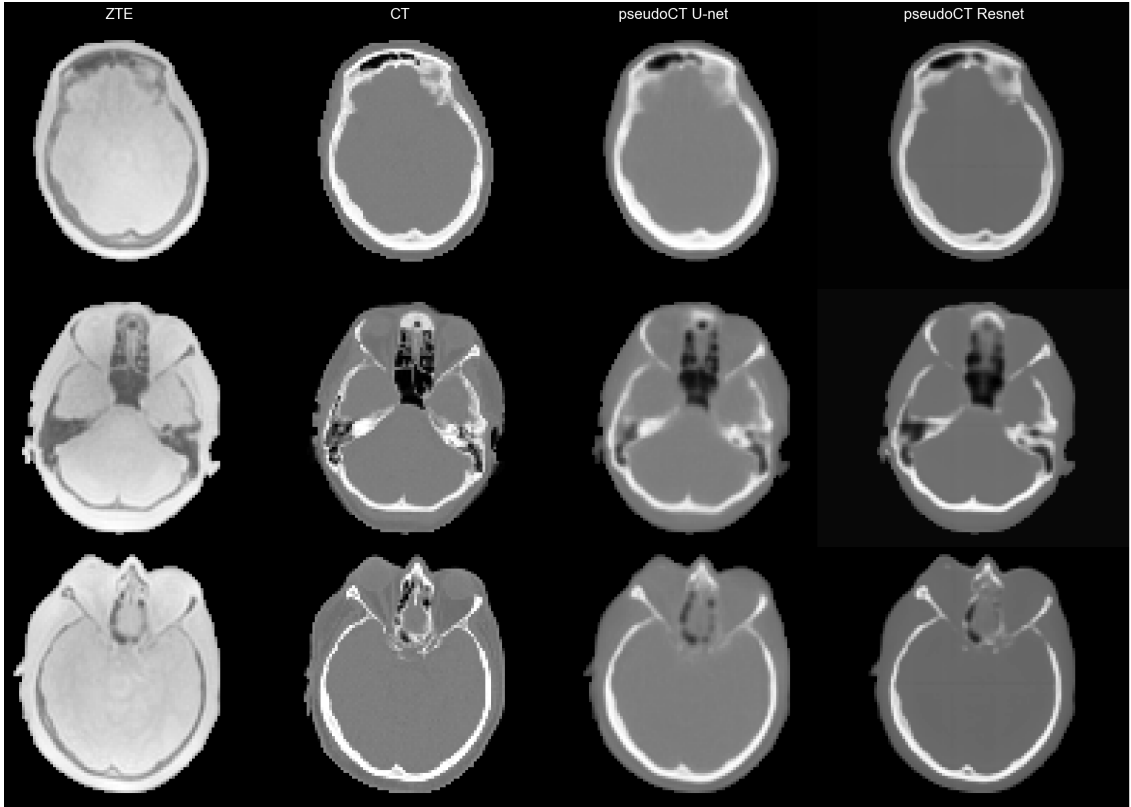
Dans la population de validation, 47 patients agé de 69,5 ans IQ75[63.4 ; 77.8], min - max [23.6 ; 90.7] ont finalement été inclus. Des exemples de coupes et de patients aléatoires sont montrés en figure 3.2.

Les pseudoCT pour l’Atlas n’étaient pas disponibles car étaient intégrés à la machine du constructeur et n’étaient donc pas exportables de manière exploitable. En terme de qualité de la segmentation osseuse, pour un seuil fixe de 200UH, le DSC pour pour le Resnet était de 0.823 IQ75[0.792 ; 0.851] et pour l’U-net était de 0.801 IQ75[0.778 ; 0.834] . Comme montré dans la figure 3.3 qui montre l’évolution du DSC en fonction du seuil osseux défini pour la segmentation, le Resnet réussi mieux que l’U-net pour la segmentation osseuse à des seuils plus élevés.

Impact des MRAC sur la quantification TEP

Les histogrammes joints entre les différentes MRAC et les CTAC sont représentés en figure 3.4. Toutes ces méthodes sont centrées autour de la ligne d’identité. Ainsi étaient retrouvés respectivement pour le R^2 et le RMSE : 0.98 et 553.0 pour (TEP_{Atlas} and TEP_{CT}), 0.99 et 327.0 pour (TEP_{Resnet} et TEP_{CT}), et 1.0 et 259.4 pour (TEP_{U-net} et TEP_{CT}) .

FIGURE 3.2 – Exemples de coupes axiales de ZTE et des pseudoCT ZTE, U-net, Resnet utilisés pour la correction d’atténuation de trois patients



Les pCT_{U-net} ont conduit à la fois à moins de biais, mais également moins de variance que les pCT_{Resnet} ou les pCT_{Atlas} . En effet dans l’espace natif avant normalisation spatiale dans le MNI, les médianes et écarts interquartiles ($IQR = IQ75\% - IQ25\%$) des erreurs TEP relatives étaient de $-1.4\% \pm 13.5$ pour TEP_{Atlas} ; $-1,4\% \pm 7.2$ pour TEP_{Resnet} et de $-0.2\% \pm 5.6$ pour TEP_{U-net} . Les distributions des erreurs relatives peuvent être vues en figure 3.5.

FIGURE 3.3 – Boîtes à moustaches du Coefficient de Dice en fonction du seuil de segmentation retenu pour les pCT_{ZTE} , pCT_{U-net} et pCT_{Resnet}

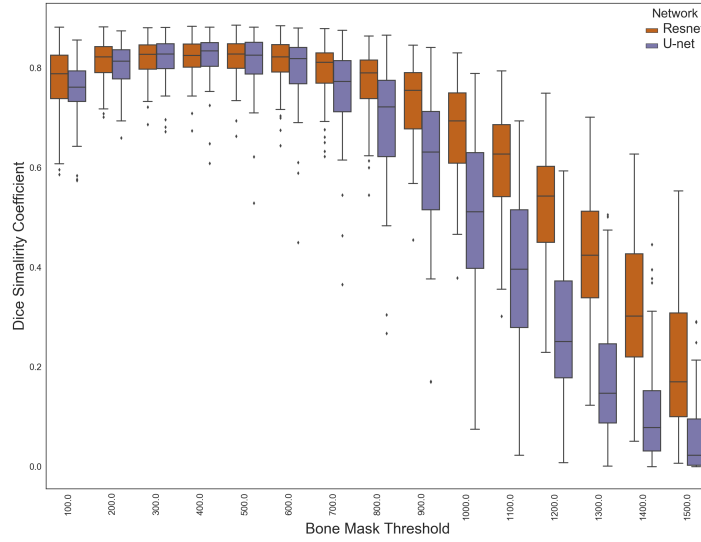


FIGURE 3.4 – Histogrammes joints des TEP_{Atlas} , $Resnet$, $U-net$ et des TEP_{CT}

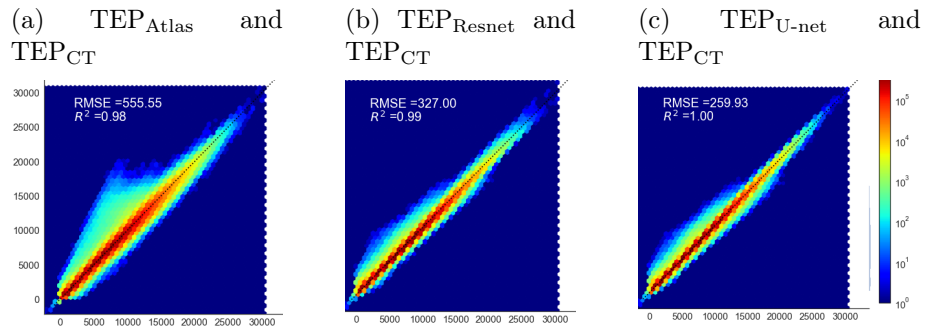
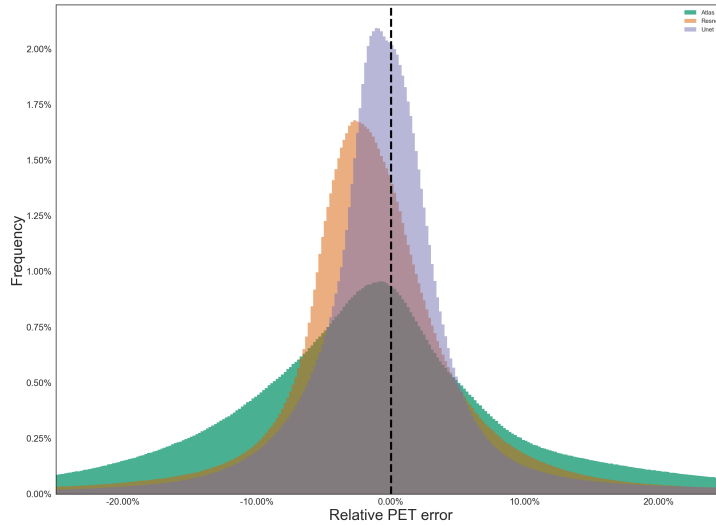


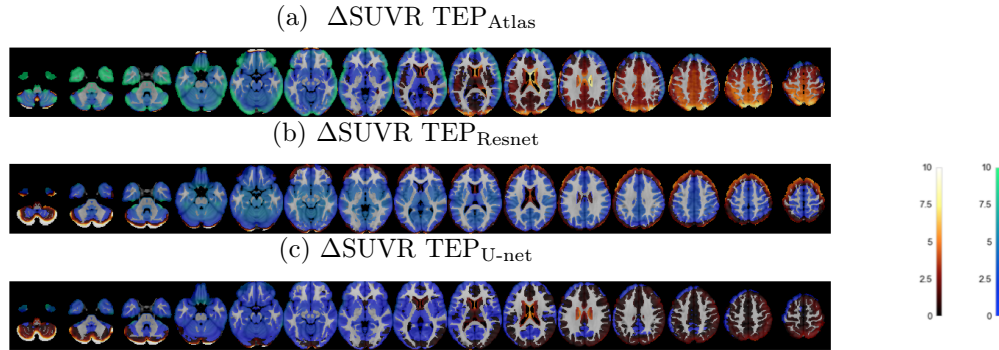
FIGURE 3.5 – Erreurs relatives TEP (Δ SUVR) dans l’espace natif par les différentes techniques de MRAC. La ligne en pointillée réfère à 0% d’erreur.



3.3.2 Quantification régionale des erreurs

Les cartographies paramétriques des erreurs TEP relatives sont montrées en figure 3.6. L’analyse par région d’intérêt (ROI) à partir des régions anatomiques regroupées de l’AAL est détaillée dans la Table 3.1 mais également sous la forme d’un radar dans la figure 3.7. On retrouvait pour l’Atlas une sur-estimation au dessus de la ligne passant par la commissure antérieure et la commissure postérieure (CA-CP) atteignant en moyenne en pariétal supérieur droit $+4.7\% \pm 7.1$ (p-value corrigée < 0.01), ainsi qu’une sous-estimation en dessous de cette ligne incluant notamment les lobes occipitaux, les lobes temporaux et minimale aux pôles atteignant $-6.6\% \pm 5.7$ (p-value corrigée < 0.01) et dans le cervelet $-5.2\% \pm 7.3$ (p-value corrigée < 0.01). Le Resnet quant à lui, était à l’origine d’une sous-estimation relativement diffuse minimale en amygdalien bilatéral de $-4.7\% \pm 1.9$ (p-value corrigée < 0.01) et une sur-estimation du vertex maximale en pariétal supé-

FIGURE 3.6 – Partie gauche de la figure : Coupes axiales des erreurs relatives TEP moyennes par voxel des différentes MRAC, avec de haut en bas : $\Delta\text{SUVR TEP}_{\text{Atlas}}$, $\Delta\text{SUVR TEP}_{\text{Resnet}}$ et $\Delta\text{SUVR TEP}_{\text{U-net}}$. Partie droite de la figure : Échelles de couleurs utilisées : hotmetal (sur-estimation, à gauche), winter (sous-estimation, à droite).



rieur droit de $+1.7\% \pm 2.6$ (p-value corrigée < 0.05) et plus modérément du cervelet. Pour l'U-net une discrète sur-estimation des aires corticales hautes et du cervelet était retrouvée, maximale dans le cortex pariétal supérieur droit de $+1.7\% \pm 2.6$, (p-value corrigée < 0.05) . En revanche, l'on pouvait observer une sous-estimation dans les régions temporo-occipitales, minimale en amygdalien gauche de $-1.8\% \pm 1.9$ (p-value corrigée < 0.01) . Les variabilités inter-régionales mesurées par les écarts-types des erreurs moyennes des différentes régions AAL s'élevaient à 3.1 pour l'Atlas, à 1.7 pour le Resnet et à 0.8 pour l'U-net.

3.4 Discussion

Dans cette partie nous avons évalué les performances de deux algorithmes d'apprentissage profond utilisant des architectures 3D basées sur des séquences ZTE pour l'estimation des cartographies d'atténuation chez 47 patients adressés pour des troubles cognitifs par dans le cadre d'examens

FIGURE 3.7 – Radar des erreurs relatives moyennes et des déviations standards des 70 régions d'intérêts regroupées d'AAL pour les TEP_{Atlas} , TEP_{U-net} et TEP_{Resnet}

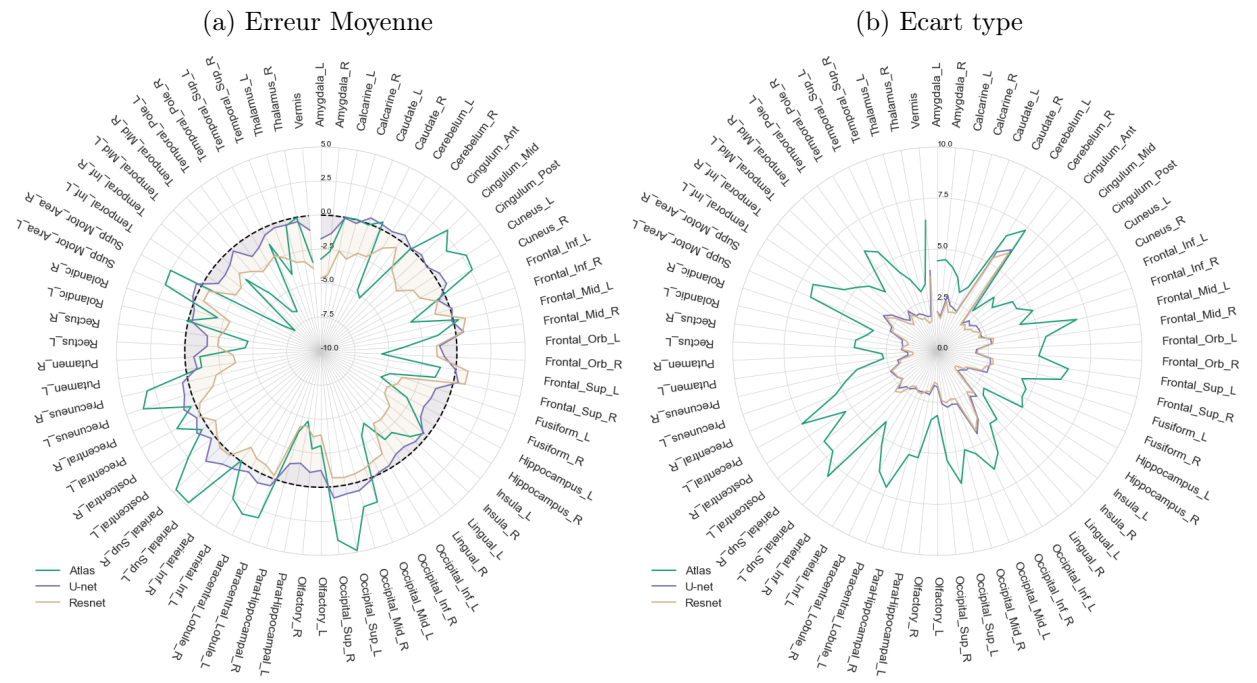


TABLE 3.1 – Moyennes et écarts-types des erreurs relatives des 70 régions d'intérêts regroupées d'AAL pour les TEP_{Atlas} , TEP_{U-net} et TEP_{Resnet} . Les niveaux de significativité des t-test appariés sont rapportés et corrigés des comparaisons multiples par la technique de Bonferroni comme suit : ** p-value corrigée < 0.01 et * p-value corrigée < 0.05

	$\Delta SUV_{Atlas} \pm SD$	$\Delta SUV_{Resnet} \pm SD$	$\Delta SUV_{U-net} \pm SD$
Amygdala_L	-3.24±4.43**	-4.63±1.91**	-1.75±1.94**
Amygdala_R	-2.65±4.5**	-4.37±1.61**	-1.31±1.72**
Calcarine_L	-0.0±4.18	-2.51±2.51**	-0.06±2.81
Calcarine_R	0.03±3.74	-2.98±1.97**	-0.26±2.33
Caudate_L	-0.09±3.07	-2.32±2.04**	0.46±2.25
Caudate_R	0.53±3.48	-2.29±2.06**	0.35±2.15
Cerebellum_L	-4.76±6.59**	-1.18±5.33	0.22±5.67
Cerebellum_R	-5.17±7.32**	-0.59±5.96	0.52±6.14
Cingulum_Ant	-0.12±2.62	-2.07±1.7**	0.05±1.81
Cingulum_Mid	2.88±3.83**	-2.47±1.88**	0.23±2.16
Cingulum_Post	1.8±3.42	-2.9±1.69**	-0.07±2.01
Cuneus_L	2.8±4.41**	-1.88±1.95**	0.43±2.27
Cuneus_R	2.63±4.4*	-1.74±2.01**	0.48±2.29
Frontal_Inf_L	-0.98±5.07	-0.68±2.33	-0.58±2.3
Frontal_Inf_R	-3.06±4.24**	-1.19±2.13*	0.14±2.26
Frontal_Mid_L	0.37±6.99	0.89±2.82	-0.13±2.68
Frontal_Mid_R	-1.33±5.36	0.51±2.73	0.58±2.67
Frontal_Orb_L	-3.95±5.13**	-1.5±1.99**	-1.27±1.92**
Frontal_Orb_R	-5.52±4.97**	-1.45±1.95**	-0.91±1.88
Frontal_Sup_L	-1.13±6.47	0.91±2.74	-0.04±2.61
Frontal_Sup_R	-1.39±6.18	0.87±2.8	0.36±2.66
Fusiform_L	-4.66±4.34**	-3.75±2.46**	-1.13±2.63*
Fusiform_R	-5.2±4.71**	-3.89±2.38**	-1.2±2.64*
Hippocampus_L	-2.83±4.7**	-4.47±1.62**	-1.41±1.73**
Hippocampus_R	-1.78±4.98	-4.4±1.54**	-1.23±1.68**
Insula_L	-0.32±2.91	-3.53±1.16**	-0.41±1.46
Insula_R	-0.92±3.1	-3.43±1.17**	-0.39±1.39
Lingual_L	-2.07±4.42**	-3.32±2.71**	-0.7±2.93
Lingual_R	-1.98±4.38*	-3.43±2.45**	-0.64±2.73
Occipital_Inf_L	-3.84±5.78**	-1.21±3.79	-0.22±3.88
Occipital_Inf_R	-3.1±6.41*	-1.09±4.32	-0.13±4.48
Occipital_Mid_L	1.82±5.36	-0.82±2.5	0.21±2.76
Occipital_Mid_R	1.85±5.01	-0.65±2.42	0.62±2.69
Occipital_Sup_L	4.89±6.58**	-0.57±2.57	0.65±2.79
Occipital_Sup_R	3.94±5.39**	-0.67±2.43	0.83±2.6
Olfactory_L	-3.06±3.15**	-3.85±1.61**	-1.24±1.76**
Olfactory_R	-2.79±3.37**	-3.69±1.53**	-1.06±1.68*
ParaHippocampal_L	-4.77±5.3**	-4.42±2.02**	-1.64±2.16**
ParaHippocampal_R	-3.85±5.25**	-4.25±2.03**	-1.43±2.11**
Paracentral_Lobule_L	3.06±7.11	-0.25±2.23	0.55±2.12
Paracentral_Lobule_R	3.28±6.53	-0.72±2.19	0.73±2.14
Parietal_Inf_L	2.61±5.29	-0.81±2.01	0.34±2.11
Parietal_Inf_R	-0.08±4.27	-1.27±2.11**	0.73±2.28
Parietal_Sup_L	4.77±8.15*	0.17±2.67	1.01±2.3
Parietal_Sup_R	4.71±7.07**	0.25±2.85	1.69±2.63*
Postcentral_L	2.41±7.08	-0.23±2.14	0.27±2.17
Postcentral_R	0.46±5.37	-0.74±2.29	0.91±2.32
Precentral_L	2.0±7.51	0.13±2.34	0.34±2.39
Precentral_R	0.37±5.42	-0.34±2.41	0.93±2.48
Precuneus_L	3.71±4.69**	-1.67±1.9**	0.53±1.97
Precuneus_R	3.08±4.03**	-2.11±1.84**	0.42±1.99
Putamen_L	-1.17±2.65	-3.68±1.19**	-0.83±1.46
Putamen_R	-1.11±2.71	-3.56±1.2**	-0.66±1.45
Rectus_L	-4.44±4.06**	-2.42±1.7**	-1.64±1.76**
Rectus_R	-4.6±4.03**	-2.44±1.68**	-1.59±1.7**
Rolandic_L	0.1±3.54	-2.82±1.45**	-0.3±1.77
Rolandic_R	-2.02±3.29**	-2.98±1.58**	-0.23±1.8
Supp_Motor_Area_L	2.41±6.74	-0.62±2.38	0.23±2.5
Supp_Motor_Area_R	2.53±6.9	-0.67±2.44	0.4±2.53
Temporal_Inf_L	-6.46±5.47**	-2.5±3.0**	-0.97±3.16
Temporal_Inf_R	-7.57±5.02**	-2.51±2.57**	-1.01±2.84
Temporal_Mid_L	-2.37±4.61*	-1.71±2.0**	-0.78±2.26
Temporal_Mid_R	-4.42±3.86**	-2.22±1.72**	-0.23±2.0
Temporal_Pole_L	-6.39±6.06**	-2.77±2.2**	-1.4±2.43**
Temporal_Pole_R	-6.61±5.66**	-2.53±1.96**	-1.13±2.01**
Temporal_Sup_L	-1.28±4.31	-2.2±1.72**	-0.53±2.03
Temporal_Sup_R	-3.98±3.72**	-2.72±1.73**	-0.31±1.9
Thalamus_L	-0.44±3.02	-3.22±1.43**	-0.38±1.76
Thalamus_R	-0.03±3.34	-3.32±1.5**	-0.33±1.76
Vermis	-3.44±6.44**	-3.9±3.67**	-1.08±3.97

TEP/IRM et nous les avons comparés à une TDM et à la technique actuellement implémentée en pratique clinique par Atlas. À notre connaissance c'est aussi le premier article à évaluer et à valider dans une grande cohorte indépendante ce type de réseaux de neurones. La MRAC s'appuyant sur un réseau de type U-net et développée par le constructeur semble être la méthode avec le moins de biais et le moins de variabilité inter ou intra-patient comparée à celles par l'Atlas ou par Resnet.

Par ailleurs, notre étude confirme les limites déjà connues en ce qui concerne la correction d'atténuation par Atlas [52]. En effet pour l'Atlas, un biais positif (sur-estimation) était retrouvée dans les régions au dessus d'une ligne rejoignant la commissure antérieure et la commissure postérieure (ligne CA-CP) spécifiquement dans les régions proche du vertex et un biais négatif (sous-estimation) dans les régions inférieures en excluant le cervelet. Il est à noter que ces régions n'apparaissent pas forcément sur le radar des erreurs moyennes, car ces sur-estimations sont moyennées avec les sous-estimations. En annexe, le radar des 116 régions AAL est présenté.

Les pseudoCT des deux réseaux de neurones convolutifs semblent visuellement appropriés sans ossifications hétérotopiques. L'U-net produit des résultats qui apparaissent comme plus lissés comparativement au Resnet. Néanmoins dans ces deux techniques les petites structures osseuses étaient souvent manquées comme les cellules mastoïdiennes ou ethmoïdiennes, les ailes du sphénoïde ou la selle turcique, probablement en raison de la faible résolution de la séquence ZTE et de la compression de l'information dans les voies d'encodage.

Il existait pour l'U-net et le Resnet un biais positif dans les régions corticales situées au-dessus de la ligne CA-CP, plus important pour le Resnet que pour l'U-net mais qui restait inférieur à celui retrouvé par l'Atlas. À l'inverse, un biais négatif dans les régions caudales à l'exception de certaines des

régions du cervelet était retrouvé. Cette sur-estimation peut être en partie expliquée par une sur-estimation de l'atténuation des cellules mastoïdiennes, probablement par un manque de capacité à restaurer la partie aérienne des mastoïdes qui était visuellement le plus souvent partiellement remplie de structures tissulaires. Il existe aussi possiblement, comme dans les approches qui tentent de segmenter automatiquement la ZTE [61], une sur-estimation de l'épaisseur corticale en rapport avec des effets de volume partiel liés à la faible résolution de la ZTE et à l'origine d'une mauvaise assignation de la classe de ces voxels d'air en os [61, 62]. Les sous-estimations caudales (temporales, occipitales, cingulaires) peuvent être expliquées par l'incapacité des deux réseaux à restituer les valeurs hautes de densités des structures osseuses comme montré dans la figure 3.3. De manière paradoxale, alors que le Resnet peut sembler meilleur au premier abord à segmenter ces structures osseuses, il existait néanmoins pour le Resnet une sous-estimation des régions caudales plus importante comparativement à l'U-net où sont pourtant situés les os les plus denses (rocher, os occipital) sous-entendant que le Resnet sous-évalue le plus au final l'épaisseur corticale dans ces régions. Visuellement, dans ces régions en particulier, les images U-net apparaissent plus lissées, expliquant la chute du DSC aux valeurs hautes de densité et dont l'origine pourrait à la fois provenir d'un post-traitement du constructeur (par exemple par l'application d'un filtre Gaussien) ou en rapport avec l'architecture même du réseau de neurones et des opérations de regroupements (*maxPooling*) dans la voie d'encodage. Aussi, les deux CNN présentaient des variabilités inter-individuelles très proches, améliorant sensiblement une des principales limites de l'Atlas.

Au total, l'analyse visuelle, l'analyse par région d'intérêt et les t-test appariés par région AAL sur les images TEP reconstruites ont montré que la méthode de correction d'atténuation en utilisant une architecture 3D U-net était celle avec le biais le plus faible et le moins de variabilité inter-individuelle ou inter-régionale comparativement aux deux autres techniques

évaluées. Peu d'études ont évalué des CNN pour l'estimation de la correction d'atténuation à partir de la ZTE. À notre connaissance, Gong et al.[50] sont les premiers et seuls à avoir étudié un CNN incluant la séquence ZTE pour l'estimation de la correction d'atténuation cérébrale. Notre U-net retrouvait des performances similaires dans la prédiction du masque osseux à un seuil identique de 200 UH avec un DSC atteignant 0.80 contre 0.77 pour leur DIXON + ZTE U-net et 0.80 pour DIXON + ZTE *Grouped U-net*. Nous avons également trouvé que l'U-net réduisait à la fois les variabilités inter-individuelles ou inter-régionales. Nos deux études présentent néanmoins des différences notables. En effet, ces derniers retrouvaient une sur-estimation du métabolisme glucidique relativement diffuse du cortex cérébral de l'ordre de 2% par leur CNN. Néanmoins, ils ont entraîné leur réseau sur une population plus petite (14 patients) et la validation croisée les expose à un risque de sur-apprentissage. De plus ils n'ont pas utilisé les ZTE seules mais en association avec la séquence DIXON. Enfin leur architecture de réseau de neurone était également différente et utilisait des coupes axiales en 2 dimensions et n'exploitaient donc pas l'information contextuelle tri-dimensionnelle. Toutes ces différences peuvent expliquer pourquoi les cartographies d'erreurs relatives sont différentes.

Notre étude présente des avantages et des limitations. Comparé aux études antérieures dans le domaine, à notre connaissance nous utilisons la plus large cohorte que ce soit pour l'entraînement ou pour la validation. Nous utilisons également des patients inclus dans le cadre du soin courant pouvant présenter de larges variations de métabolisme notamment dans les aires associatives. Les TEP/IRM et les protocoles de reconstruction étaient les mêmes pour tous les patients. Néanmoins notre travail possède des limites. Tout d'abord la collecte des données était de nature retrospective et les TDM pouvaient provenir de constructeurs différents. En outre, le lissage des cartes d'atténuations pour se rapprocher de la résolution de la TEP, la normalisation spatiale et le post-filtre Gaussien de 10mm sur les images TEP

réduisent les variabilités liées aux différentes acquisitions scanner. Une autre limitation de cette étude était le délai entre les TEP/IRM et les TDM. Cependant, un contrôle visuel entre les séquences ZTE et les TDM co-registrées étaient réalisé chez tous les patients pour s’assurer qu’aucune modification significative de l’anatomie n’était apparue pendant cet intervalle. Enfin la comparabilité entre U-net et Resnet est discutable car la population d’entraînement était très différente. En effet, comme une plus grande taille d’échantillon conduit généralement à une amélioration des performances [32], il est probable que les meilleures performances du réseau U-net soient en partie liée à une plus grande taille de l’échantillon. Néanmoins il est intéressant de voir que malgré une population plus petite, des réseaux de neurones peuvent avoir d’excellentes performances et peuvent trouver leurs applications cliniquement. Enfin, nous avons fais le choix de ne comparer ces deux réseaux qu’à l’Atlas, en effet, des techniques plus récentes sont en cours d’évaluations, mais ne sont pas couramment utilisées en pratique clinique. De plus l’évaluation de quatres techniques de correction d’atténuation aurait rendu la lecture de ce travail plus compliqué encore. Il est à noter, données non montrées, que la technique de correction d’atténuation à partir de la séquence ZTE [61], qui même si produisent des erreurs de l’ordre de -2% semblent présenter une variabilité inter-régionale plus faible que le Resnet la rendant en pratique clinique probablement plus appropriée.

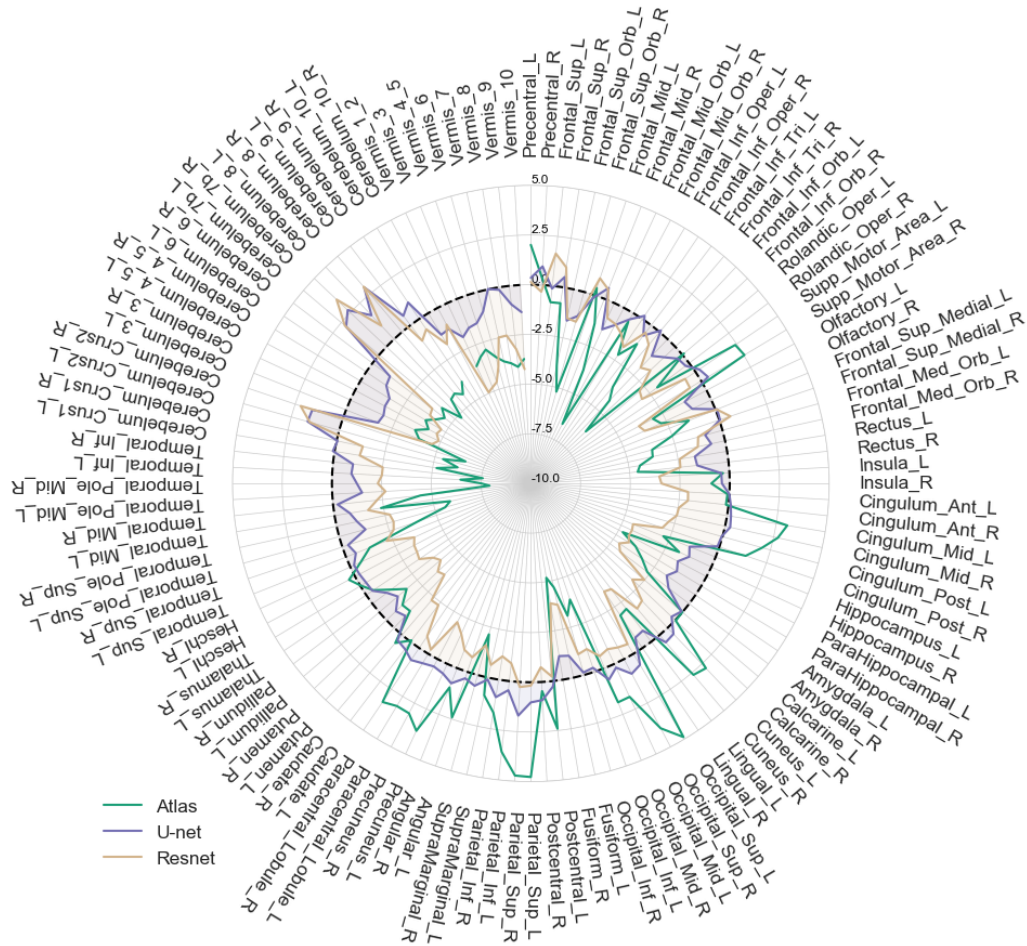
3.5 Conclusion

Nous avons donc évalué les performances de deux architectures de réseaux de neurones convolutifs basées sur la séquence ZTE pour générer des pseudoCT et des cartographies d’atténuation pour l’imagerie TEP cérébrale. Ces résultats suggèrent que la correction d’atténuation par l’U-net est plus appropriée et mieux adaptée pour l’exploration des troubles cognitifs car possède moins de biais et de variabilité inter ou intra-patient comparative-ment à l’Atlas et au Resnet. Dans l’avenir, nous allons tenter d’évaluer ces

réseaux de neurones sur d'autres séquences type UTE. Nous pourrions aussi tenter de valider la séquence ZTE sur des machines d'autres constructeurs afin de tester la reproductibilité et la robustesse de ces algorithmes.

Annexe

FIGURE 3.8 – Radar des erreurs relatives moyennes des 116 régions d'intérêts regroupées d'AAL pour les TEP_{U-net} et TEP_{Resnet}



Conclusion

Au travers de ces deux exemples nous avons donc tenté de montrer que les réseaux de neurones convolutifs pouvaient apprendre des représentations internes utiles que ce soit pour la segmentation des tumeurs gliales en ^{18}F -FET ou pour l'estimation de la cartographie d'atténuation à l'aide de la séquence IRM ZTE. Nous avons également vu, contrairement à l'idée populaire, qu'il n'était pas forcément nécessaire d'avoir de larges populations pour l'entraînement de ces réseaux, au coût de l'utilisation de stratégies d'augmentation de données et au risque de sur-apprentissage. Les réseaux de neurones convolutifs vont considérablement modifier nos pratiques dans les années à venir. Les applications sont nombreuses et pourront impacter toutes les étapes de la génération et du traitement de nos images : réduction des doses injectées, diminution des temps d'examens, accélération de la reconstruction, amélioration de la résolution, détection des lésions, segmentation, caractérisation, classification et amélioration des performances prédictives de nos images... Néanmoins même si ces réseaux semblent apprendre les tâches qui leurs sont affectées n'en sont pas intelligents pour autant. Ils agissent de plus de manière opaque et sont soumis à des biais qu'il sera important d'évaluer à l'échelle de chaque institution, de chaque machine et dans chaque indication. Les réseaux de neurones bien qu'inspirés du fonctionnement de notre cerveau ne représentent au final que l'ultime étape de traitement de l'information à savoir la restitution de l'apprentissage. Celui ci apparaît comme immuable, sans plasticité ni rétrocontrôle. Inversement il est bon de rappeler que le système nerveux ne se comporte pas comme un réseau de neurones convolutif et pour reprendre Alan Turing,

n'est pas une machine à poids discrets.

Néanmoins, jusqu'à peu, les réseaux utilisés n'étaient pas développés pour l'imagerie médicale et ses particularités. Des solutions émergent pour répondre à nos problématiques et nos caractéristiques d'images. En effet une de ses caractéristiques importantes est le caractère extrêmement déséquilibré des classes. Par exemple pour la segmentation des tumeurs gliales, plus de 99% des voxels n'étaient pas tumoraux. Un algorithme qui ne prédirait donc que des voxels non tumoraux aurait donc raison dans 99% des cas. Pour le cas de la segmentation, ce problème peut être résolu par le choix d'une fonction de coût appropriée [63] ou par l'utilisation de patches qui ne sont pas tirés aléatoirement de l'image mais de manière balancée entre les différentes classes [64]. Compte tenu de la numérisation et de l'archivage systématique des images, un des défis ne sera pas de l'accessibilité des données mais plutôt son caractère supervisé avec des annotations et des labels de qualité, fiables et reproductibles. En outre, même si les comptes rendus radiologiques à l'APHP sont associés aux images et archivés également dans le PACS, les convertir de manière automatique en labels structurés est un champ de recherche à part entière qui aujourd'hui n'apporte pas de solutions probantes. De plus, il est déjà souvent laborieux et compliqué d'obtenir des labels d'un seul imageur, mais quand plusieurs sont disponibles, le consensus peut être compliqué. Par exemple, pour la détection des micronodules pulmonaires, 4 radiologues ont participé à la construction d'une base de donnée : le LIDC-IDRI. Le nombre de micronodules où un consensus n'était pas atteint était trois fois supérieur à celui où un consensus était atteint [65]. Des approches non supervisées sont donc en cours d'évaluation pour tenter de s'affranchir de la labélisation, mais aussi des approches prenant en compte le fait que le manque de consensus est déjà une information utile.

Enfin, il va de soit qu'un examen ne s'interprète pas sans son contexte clinique et sans les antécédents du patient obtenus à l'aide d'un interroga-

toire qui doit rester comme un moment privilégié entre le médecin et son malade, et même si des applications vont probablement chercher à intégrer l'information contextuelle clinique, il paraît compliqué à une courte échéance que ce genre de process puisse être raisonnablement déployé.

Serment d'Hippocrate

Au moment d'être admis à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité.

Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences. Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences.

Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admis(e) dans l'intimité des personnes, je tairai les secrets qui me seront confiés. Reçu à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs. Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité.

Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré(e) et méprisé(e) si j'y manque.

Table des figures

1.1	Émission de positon, annihilation et détection de coïncidence par TEP (d'Iris Verel, The Promise of Immuno PET, JNM 2005)	8
1.2	Les γ -caméras sont composées successivement d'un collimateur, de détecteurs et d'une électronique	9
1.3	Les tenseurs sont des objets mathématiques issus de l'algèbre linéaire permettant de généraliser les scalaires et les vecteurs, Wikipédia	11
1.4	Intégration synaptique des potentiels dendritiques En haut à gauche : trois axones de neurones A,B,C interagissent avec un quatrième neurone. En haut à droite : la sommation temporelle deux potentiels post-synaptique excitateurs dépassant le seuil d'activation est à l'origine d'un potentiel d'action. En bas à gauche : la sommation spatiale de deux potentiels post-synaptique excitateurs et d'un potentiel post-synaptique inhibiteur ne dépasse pas le seuil d'activation. En bas à droite : la sommation spatiale de deux potentiels post-synaptique excitateurs dépassant le seuil d'activation est à l'origine d'un potentiel d'action. Issu de JM Mienville, unice.fr/JMM/Cours/Neuro	13
1.5	Une unité neuronale typique calcule la combinaison linéaire z de ses entrées x_i en utilisant les poids w_i . La sortie y de l'unité est une fonction de z , appelée fonction d'activation. Les fonctions d'activation populaires incluent les fonctions sigmoïdes (tangente hyperbolique représentée ici en rouge) ou les fonctions linéaires rectifiées (ReLU, turquoise).	14
1.6	Le compromis biais-variance.	20

1.7	Architecture classique d'un réseau de neurones convolutif. Une image est fournie en donnée d'entrée (input) et est convoluée avec des filtres (première couche de convolution) et dont les cartes d'activation sont regroupées et concaténées. . . .	24
1.8	96 filtres de convolution de taille 11x11x3 appris par la première couche de Krizhevsky et al[16].	25
2.1	Architecture d'un auto-encodeur convolutif de type U-net comme décrit par Ronneberger et al [31]	33
2.2	Courbes d'apprentissage avec évolution du DSC (en haut) et de la fonction de coût (en bas) au fil des époques	36
2.3	Exemples de segmentations avec à chaque fois : les images natives 18F-FET (en haut), la vérité terrain (milieu haut), le masque prédit (milieu bas) et la différence entre le masque prédit et la vérité terrain (en bas, avec les faux négatifs en rouge et les faux positifs en bleu)	37
3.1	Architecture Resnet utilisée par Li et al. [53]	50
3.2	Exemples de coupes axiales de ZTE et des pseudoCT ZTE, U-net, Resnet utilisés pour la correction d'atténuation de trois patients	54
3.3	Boîtes à moustaches du Coefficient de Dice en fonction du seuil de segmentation retenu pour les pCT _{ZTE} , pCT _{U-net} et pCT _{Resnet}	55
3.4	Histogrammes joints des TEP _{Atlas, Resnet, U-net} et des TEP _{CT}	55
3.5	Erreurs relatives TEP (Δ SUVR) dans l'espace natif par les différentes techniques de MRAC. La ligne en pointillée réfère à 0% d'erreur.	56
3.6	Partie gauche de la figure : Coupes axiales des erreurs relatives TEP moyennes par voxel des différentes MRAC, avec de haut en bas : Δ SUVR TEP _{Atlas} , Δ SUVR TEP _{Resnet} et Δ SUVR TEP _{U-net} . Partie droite de la figure : Échelles de couleurs utilisées : hotmetal (sur-estimation, à gauche), winter (sous-estimation, à droite).	57
3.7	Radar des erreurs relatives moyennes et des déviations standards des 70 régions d'intérêts regroupées d'AAL pour les TEP _{Atlas} , TEP _{U-net} et TEP _{Resnet}	58
3.8	Radar des erreurs relatives moyennes des 116 régions d'intérêts regroupées d'AAL pour les TEP _{U-net} et TEP _{Resnet}	65

Liste des tableaux

2.1	Caractéristiques de la population	35
2.2	Table de contingence résumant les performances pour la segmentation des voxels tumoraux ou non la population de validation.	38
3.1	Moyennes et écarts-types des erreurs relatives des 70 régions d'intérêts regroupées d'AAL pour les TEP_{Atlas} , TEP_{U-net} et TEP_{Resnet} . Les niveaux de significativité des t-test appariés sont rapportés et corrigés des comparaisons multiples par la technique de Bonferroni comme suit : ** p-value corrigée < 0.01 et * p-value corrigée < 0.05	59

Bibliographie

- [1] Stefaan Vandenberghe and Paul K Marsden. PET-MRI : A review of challenges and solutions in the development of integrated multimodality imaging. *Physics in Medicine and Biology*, 60(4) :R115–R154, February 2015.
- [2] McCulloch and Pitts. A logical calculus of the ideas immanent in nervous activity. 5(5) :115–133.
- [3] Anton Maximilian Schäfer and Hans-Georg Zimmermann. Recurrent Neural Networks are universal approximators. *International Journal of Neural Systems*, 17(4) :253–263, August 2007.
- [4] H. Bourslard and Y. Kamp. Auto-association by multilayer perceptrons and singular value decomposition. *Biological Cybernetics*, 59(4-5) :291–294, September 1988.
- [5] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Networks. *arXiv :1406.2661 [cs, stat]*, June 2014.
- [6] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. *arXiv :1512.03385 [cs]*, December 2015.
- [7] Marine Beurdeley, Julien Spatazza, Henry H. C. Lee, Sayaka Sugiyama, Clémence Bernard, Ariel A. Di Nardo, Takao K. Hensch, and Alain Prochiantz. Otx2 binding to perineuronal nets persistently regulates plasticity in the mature visual cortex. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, 32(27) :9429–9437, July 2012.
- [8] Yann Bernardinelli, Jerome Randall, Elia Janett, Irina Nikonenko, Stéphane König, Emma Victoria Jones, Carmen E. Flores, Keith K. Murai, Christian G. Bochet, Anthony Holtmaat, and Dominique Müller. Activity-Dependent Structural Plasticity of Perisynaptic Astrocy-

-
- tic Domains Promotes Excitatory Synapse Stability. *Current Biology*, 24(15) :1679–1688, August 2014.
- [9] D. J. Felleman and D. C. Van Essen. Distributed Hierarchical Processing in the Primate Cerebral Cortex. *Cerebral Cortex*, 1(1) :1–47, January 1991.
- [10] William A. Press, Alyssa A. Brewer, Robert F. Dougherty, Alex R. Wade, and Brian A. Wandell. Visual areas and spatial summation in human visual cortex. *Vision Research*, 41(10) :1321–1332, May 2001.
- [11] Brian A. Wandell, Serge O. Dumoulin, and Alyssa A. Brewer. Visual Field Maps in Human Cortex. *Neuron*, 56(2) :366–383, October 2007.
- [12] E. L. Schwartz. Spatial mapping in the primate sensory projection : Analytic structure and relevance to perception. *Biological Cybernetics*, 25(4) :181–194, December 1977.
- [13] D. H. Hubel and T. N. Wiesel. Receptive fields of single neurones in the cat’s striate cortex. *The Journal of Physiology*, 148(3) :574–591, October 1959.
- [14] Carolyn Jeane Perry and Mazyar Fallah. Feature integration and object representations along the dorsal stream visual hierarchy. *Frontiers in Computational Neuroscience*, 8, 2014.
- [15] D. Marr and H. K. Nishihara. Representation and Recognition of the Spatial Organization of Three-Dimensional Shapes. *Proceedings of the Royal Society B : Biological Sciences*, 200(1140) :269–294, February 1978.
- [16] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6) :84–90, May 2017.
- [17] Therese A. Dolecek, Jennifer M. Propp, Nancy E. Stroup, and Carol Kruchko. CBTRUS Statistical Report : Primary Brain and Central Nervous System Tumors Diagnosed in the United States in 2005–2009. *Neuro-Oncology*, 14(Suppl 5) :v1–v49, November 2012.
- [18] Karl-Josef Langen, Kurt Hamacher, Matthias Weckesser, Frank Floeth, Gabriele Stoffels, Dagmar Bauer, Heinz H. Coenen, and Dirk Pauleit. O-(2-[18F]fluoroethyl)-l-tyrosine : Uptake mechanisms and clinical applications. *Nuclear Medicine and Biology*, 33(3) :287–294, April 2006.
- [19] C. la Fougere, B. Suchorska, P. Bartenstein, F.-W. Kreth, and J.-C. Tonn. Molecular imaging of gliomas with PET : Opportunities and limitations. *Neuro-Oncology*, 13(8) :806–819, August 2011.

-
- [20] Nathalie L. Jansen, Christoph Schwartz, Vera Graute, Sabina Eigenbrod, Jürgen Lutz, Rupert Egensperger, Gabriele Pöpperl, Hans A. Kretschmar, Paul Cumming, Peter Bartenstein, Jörg-Christian Tonn, Friedrich-Wilhelm Kreth, Christian la Fougère, and Niklas Thon. Prediction of oligodendroglial histology and LOH 1p/19q using dynamic [18F]FET-PET imaging in intracranial WHO grade II and III gliomas. *Neuro-Oncology*, 14(12) :1473–1480, December 2012.
- [21] Garry Ceccon, Philipp Lohmann, Gabriele Stoffels, Natalie Judov, Christian P. Filss, Marion Rapp, Elena Bauer, Christina Hamisch, Maximilian I. Ruge, Martin Kocher, Klaus Kuchelmeister, Bernd Sellhaus, Michael Sabel, Gereon R. Fink, Nadim J. Shah, Karl-Josef Langen, and Norbert Galldiks. Dynamic O -(2-¹⁸F-fluoroethyl)-L-tyrosine positron emission tomography differentiates brain metastasis recurrence from radiation injury after radiotherapy. *Neuro-Oncology*, page now149, July 2016.
- [22] Bogdana Suchorska, Nathalie Lisa Jansen, Theo Kraus, Armin Giese, Peter Bartenstein, and Joerg Tonn. Correlation of dynamic 18FET-PET with IDH 1 mutation for prediction of outcome in anaplastic astrocytoma WHO° III independently from tumor vascularisation. *Journal of Clinical Oncology*, 33(suppl; abstr 2037), 2015.
- [23] Gabriele Pöpperl, Friedrich W. Kreth, Jan H. Mehrkens, Jochen Herms, Klaus Seelos, Walter Koch, Franz J. Gildehaus, Hans A. Kretschmar, Jörg C. Tonn, and Klaus Tatsch. FET PET for the evaluation of untreated gliomas : Correlation of FET uptake and uptake kinetics with tumour grading. *European Journal of Nuclear Medicine and Molecular Imaging*, 34(12) :1933–1942, December 2007.
- [24] Stephen L. Breen, Julia Publicover, Shiroma De Silva, Greg Pond, Kristy Brock, Brian O’Sullivan, Bernard Cummings, Laura Dawson, Anne Keller, John Kim, Jolie Ringash, Eugene Yu, Aaron Hendler, and John Waldron. Intraobserver and interobserver variability in GTV delineation on FDG-PET-CT images of head and neck cancers. *International Journal of Radiation Oncology, Biology, Physics*, 68(3) :763–770, July 2007.
- [25] Mathieu Hatt, John A. Lee, Charles R. Schmidlein, Issam El Naqa, Curtis Caldwell, Elisabetta De Bernardi, Wei Lu, Shiva Das, Xavier Geets, Vincent Gregoire, Robert Jeraj, Michael P. MacManus, Osama R. Mawlawi, Ursula Nestle, Andrei B. Pugachev, Heiko Schöder, Tony Shepherd, Emiliano Spezi, Dimitris Visvikis, Habib Zaidi,

-
- and Assen S. Kirov. Classification and evaluation strategies of auto-segmentation approaches for PET : Report of AAPM task group No. 211. *Medical Physics*, 44(6) :e1–e42, June 2017.
- [26] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *arXiv preprint arXiv :1702.05747*, 2017.
- [27] C. Nioche, F. Orlhac, and I. Buvat. LIFEx : Un logiciel gratuit pour caractériser l’hétérogénéité intra-tumorale en imagerie multimodale. *Médecine Nucléaire*, 40(3) :208, May 2016.
- [28] Christophe Nioche, Fanny Orlhac, Sarah Boughdad, Sylvain Reuze, Michael Soussan, Charlotte Robert, Claire Barakat, and Irene Buvat. A freeware for tumor heterogeneity characterization in PET, SPECT, CT, MRI and US to accelerate advances in radiomics. *Journal of Nuclear Medicine*, 58(supplement 1) :1316–1316, January 2017.
- [29] Stefanie Bette, Jens Gempt, Claire Delbridge, Jan S. Kirschke, Juergen Schlegel, Stefan Foerster, Thomas Huber, Thomas Pyka, Claus Zimmer, Bernhard Meyer, and Florian Ringel. Prognostic Value of O-(2-[18F]-Fluoroethyl)-L-Tyrosine-Positron Emission Tomography Imaging for Histopathologic Characteristics and Progression-Free Survival in Patients with Low-Grade Glioma. *World Neurosurgery*, 89 :230–239, May 2016.
- [30] Gabriele Popperl, Claudia Gotz, Walter Rachinger, Franz-Josef Gildehaus, Jorg-Christian Tonn, and Klaus Tatsch. Value of O-(2-[18F]fluoroethyl)-l-tyrosine PET for the diagnosis of recurrent glioma. *European Journal of Nuclear Medicine and Molecular Imaging*, 31(11) :1464–1470, November 2004.
- [31] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net : Convolutional networks for biomedical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 234–241. Springer, 2015.
- [32] Junghwan Cho, Kyewook Lee, Ellie Shin, Garry Choy, and Synho Do. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? *arXiv preprint arXiv :1511.06348*, 2015.

-
- [33] Jason Wang and Luis Perez. The effectiveness of data augmentation in image classification using deep learning. Technical report, Technical report, 2017.
- [34] Ian J. Goodfellow, Abadie, Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative Adversarial Nets. June 2014.
- [35] Christian P. Filss, Nathalie L. Albert, Guido Böning, Elena Rota Kops, Bogdana Suchorska, Gabriele Stoffels, Norbert Galldiks, Nadim J. Shah, Felix M. Mottaghy, Peter Bartenstein, Jörg C. Tonn, and Karl-Josef Langen. O-(2-[18F]fluoroethyl)-l-tyrosine PET in gliomas : Influence of data processing in different centres. *EJNMMI Research*, 7 :64, August 2017.
- [36] Henryk Barthel, Matthias L. Schroeter, Karl-Titus Hoffmann, and Osama Sabri. PET/MR in dementia and other neurodegenerative diseases. *Seminars in Nuclear Medicine*, 45(3) :224–233, May 2015.
- [37] Lisa M. Bloudek, D. Eldon Spackman, Michael Blankenburg, and Sean D. Sullivan. Review and meta-analysis of biomarkers and diagnostic imaging in Alzheimer’s disease. *Journal of Alzheimer’s disease : JAD*, 26(4) :627–645, 2011.
- [38] Daniela Perani, Chiara Cerami, Silvia Paola Caminiti, Roberto Santangelo, Elisabetta Coppi, Laura Ferrari, Patrizia Pinto, Gabriella Passerini, Andrea Falini, Sandro Iannaccone, Stefano Francesco Cappa, Giancarlo Comi, Luigi Gianolli, and Giuseppe Magnani. Cross-validation of biomarkers for the early differential diagnosis and prognosis of dementia in a clinical setting. *European Journal of Nuclear Medicine and Molecular Imaging*, 43(3) :499–508, March 2016.
- [39] Ciprian Catana, Alexander R. Guimaraes, and Bruce R. Rosen. PET and MR imaging : The odd couple or a match made in heaven ? *Journal of Nuclear Medicine : Official Publication, Society of Nuclear Medicine*, 54(5) :815–824, May 2013.
- [40] Jonathan P. J. Carney, David W. Townsend, Vitaliy Rappoport, and Bernard Bendriem. Method for transforming CT images for attenuation correction in PET/CT imaging : Transforming CT images for attenuation correction in PET/CT. *Medical Physics*, 33(4) :976–983, March 2006.
- [41] Thorsten Heußner, Christopher M. Rank, Yannick Berker, Martin T. Freitag, and Marc Kachelrieß. MLAA-based attenuation correction of

-
- flexible hardware components in hybrid PET/MR imaging. *EJNMMI Physics*, 4, March 2017.
- [42] Yannick Berker and Yusheng Li. Attenuation correction in emission tomography using the emission data—A review. *Medical Physics*, 43(2) :807–832, February 2016.
- [43] R. Boellaard, M. B. M. Hofman, O. S. Hoekstra, and A. A. Lam-mertsma. Accurate PET/MR Quantification Using Time of Flight MLAA Image Reconstruction. *Molecular Imaging and Biology*, 16(4) :469–477, August 2014.
- [44] Abolfazl Mehranian, Martin A Belzunce, Flavia Niccolini, Marios Politis, Claudia Prieto, Federico Turkheimer, Alexander Hammers, and Andrew J Reader. PET image reconstruction using multi-parametric anatofunctional priors. *Physics in Medicine & Biology*, 62(15) :5975–6007, July 2017.
- [45] Maya Khalifé, Brice Fernandez, Olivier Jaubert, Michael Soussan, Vincent Brulon, Irène Buvat, and Claude Comtat. Subject-specific bone attenuation correction for brain PET/MR : Can ZTE-MRI substitute CT scan accurately ? *Physics in Medicine & Biology*, 62(19) :7814–7832, September 2017.
- [46] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42 :60–88, December 2017.
- [47] Dong Nie, Xiaohuan Cao, Yaozong Gao, Li Wang, and Dinggang Shen. Estimating CT Image from MRI Data Using 3D Fully Convolutional Networks. *Deep learning and data labeling for medical applications : first International Workshop, LABELS 2016, and second International Workshop, DLMIA 2016, held in conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings. LA.*, 2016 :170–178, 2016.
- [48] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 417–425. Springer, 2017.

-
- [49] Xiao Han. MR-based synthetic CT generation using a deep convolutional neural network method. *Medical Physics*, 44(4) :1408–1419, April 2017.
- [50] Kuang Gong, Jaewon Yang, Kyungsang Kim, Georges El Fakhri, Youngho Seo, and Quanzheng Li. Attenuation correction for brain PET imaging using deep neural network based on dixon and ZTE MR images. *Physics in Medicine and Biology*, May 2018.
- [51] Andrew P Leynes, Jaewon Yang, Florian Wiesinger, Sandeep S Kaushik, Dattesh D Shanbhag, Youngho Seo, Thomas A Hope, and Peder E. Z. Larson. Direct PseudoCT Generation for Pelvis PET/MRI Attenuation Correction using Deep Convolutional Neural Networks with Multi-parametric MRI : Zero Echo-time and Dixon Deep pseudoCT (ZeDD-CT). *Journal of Nuclear Medicine*, page jnumed.117.198051, October 2017.
- [52] Tetsuro Sekine, Alfred Buck, Gaspar Delso, Edwin E. G. W. Ter Voert, Martin Huellner, Patrick Veit-Haibach, and Geoffrey Warnock. Evaluation of Atlas-Based Attenuation Correction for Integrated PET/MR in Human Brain : Application of a Head Atlas and Comparison to True CT-Based Attenuation Correction. *Journal of Nuclear Medicine : Official Publication, Society of Nuclear Medicine*, 57(2) :215–220, February 2016.
- [53] Wenqi Li, Guotai Wang, Lucas Fidon, Sebastien Ourselin, M. Jorge Cardoso, and Tom Vercauteren. On the compactness, efficiency, and representation of 3D convolutional networks : Brain parcellation as a pretext task. In *International Conference on Information Processing in Medical Imaging*, pages 348–360. Springer, 2017.
- [54] Shuchang Zhou, Jia-Nan Wu, Yuxin Wu, and Xinyu Zhou. Exploiting local structures with the kronecker layer in convolutional networks. *arXiv preprint arXiv :1512.09194*, 2015.
- [55] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
- [56] N. Tzourio-Mazoyer, B. Landeau, D. Papathanassiou, F. Crivello, O. Etard, N. Delcroix, B. Mazoyer, and M. Joliot. Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *NeuroImage*, 15(1) :273–289, January 2002.

-
- [57] Wes McKinney. Data structures for statistical computing in python. In Stéfan van der Walt and Jarrod Millman, editors, *Proceedings of the 9th Python in Science Conference*, pages 51 – 56, 2010.
- [58] Wes McKinney. Data Structures for Statistical Computing in Python. page 6, 2010.
- [59] J. D. Hunter. Matplotlib : A 2D Graphics Environment. *Computing in Science Engineering*, 9(3) :90–95, May 2007.
- [60] Alexandre Abraham, Fabian Pedregosa, Michael Eickenberg, Philippe Gervais, Andreas Mueller, Jean Kossaifi, Alexandre Gramfort, Bertrand Thirion, and Gaël Varoquaux. Machine learning for neuroimaging with scikit-learn. *Frontiers in Neuroinformatics*, 8, February 2014.
- [61] T. Sekine, E. E. G. W. ter Voert, G. Warnock, A. Buck, M. Huellner, P. Veit-Haibach, and G. Delso. Clinical Evaluation of Zero-Echo-Time Attenuation Correction for Brain 18F-FDG PET/MRI : Comparison with Atlas Attenuation Correction. *Journal of Nuclear Medicine*, 57(12) :1927–1932, December 2016.
- [62] G. Delso, F. Wiesinger, L. I. Sacolick, S. S. Kaushik, D. D. Shanbhag, M. Hullner, and P. Veit-Haibach. Clinical Evaluation of Zero-Echo-Time MR Imaging for the Segmentation of the Skull. *Journal of Nuclear Medicine*, 56(3) :417–422, March 2015.
- [63] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-Net : Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 565–571, Stanford, CA, USA, October 2016. IEEE.
- [64] M. J. J. P. van Grinsven, B. van Ginneken, C. B. Hoyng, T. Theelen, and C. I. Sánchez. Fast Convolutional Neural Network Training Using Selective Data Sampling : Application to Hemorrhage Detection in Color Fundus Images. *IEEE Transactions on Medical Imaging*, 35(5) :1273–1284, May 2016.
- [65] Samuel G. Armato, Geoffrey McLennan, Luc Bidaut, Michael F. McNitt-Gray, Charles R. Meyer, Anthony P. Reeves, Binsheng Zhao, Denise R. Aberle, Claudia I. Henschke, Eric A. Hoffman, Ella A. Kazerooni, Heber MacMahon, Edwin J. R. van Beek, David Yankelevitz, Alberto M. Biancardi, Peyton H. Bland, Matthew S. Brown, Roger M. Engelmann, Gary E. Laderach, Daniel Max, Richard C. Pais, David P.-Y. Qing, Rachael Y. Roberts, Amanda R. Smith, Adam Starkey, Poonam Batra, Philip Caligiuri, Ali Farooqi, Gregory W. Gladish, C. Matilda

Jude, Reginald F. Munden, Iva Petkovska, Leslie E. Quint, Lawrence H. Schwartz, Baskaran Sundaram, Lori E. Dodd, Charles Fenimore, David Gur, Nicholas Petrick, John Freymann, Justin Kirby, Brian Hughes, Alessi Vande Castele, Sangeeta Gupte, Maha Sallam, Michael D. Heath, Michael H. Kuhn, Ekta Dharaiya, Richard Burns, David S. Fryd, Marcos Salganicoff, Vikram Anand, Uri Shreter, Stephen Vastagh, Barbara Y. Croft, and Laurence P. Clarke. The Lung Image Database Consortium (LIDC) and Image Database Resource Initiative (IDRI) : A Completed Reference Database of Lung Nodules on CT Scans. *Medical Physics*, 38(2) :915–931, February 2011.

Résumé

Les réseaux de neurones convolutifs regroupent un ensemble d'architectures dont les unités élémentaires sont inspirées des neurones biologiques. Ils permettent d'estimer et d'extraire par un ensemble de techniques d'apprentissage, des représentations internes (filtres) qui sont ensuite convoluées avec nos images pour classer, prédire... Notre objectif au travers de deux exemples est de tenter de mieux comprendre les potentialités de ces derniers en médecine nucléaire et d'en appréhender les limites. Nous étudierons donc dans un premier temps une procédure de segmentation qui peut être assimilée à un problème de classification où l'on cherche à prédire pour chaque voxel une catégorie. Pour cela nous nous appuyerons sur une cohorte de 37 patients avec des tumeurs gliales explorées dans leur bilan initial en TEP à la 18-F-Fluoro-ethyl-thyrosine. Dans un second temps nous nous intéresserons à estimer des TDM à partir de la séquence IRM ZTE et d'évaluer son impact sur la correction d'atténuation lors de la reconstruction tomographique chez 47 patients adressés dans le cadre de pathologies neuro-dégénératives.

Mots clés : Réseaux de neurones convolutifs, neuro-imagerie, tomographie par émission de positon, apprentissage profond, médecine nucléaire

Abstract

Convolutional neural networks group together a set of architectures whose elementary units are inspired by biological neurons. They are able to estimate and extract internal representations (filters) by learning from annotated data. They are then convoluted with our images to classify or predict. Our objective through two examples is to try to better understand the potential of the latter in nuclear medicine and to understand their limits. We will first study a segmentation procedure that can be assimilated to a classification problematic where we try to predict for each voxel a category. For this we will rely on a cohort of 37 patients with glial tumors explored within initial staging by 18-F-Fluoro-ethyl-thyrosine PET. In a second time, we will focus on estimating CT scans from ZTE MRI sequences and to evaluate its impact on attenuation correction during tomographic reconstruction in 47 patients referred for neurodegenerative disease.

Key words : Convolutional neural networks, neuro-imaging, positron emission tomography, nuclear medicine

Faculté de Médecine Paris Descartes
15, rue de l'Ecole de Médecine
75270 Paris cedex