



HAL
open science

Apprentissage d'un CNN sur la segmentation des lésions de sclérose en plaques avec auto-supervision

Alexandre Fenneteau, Pascal Bourdon, David Helbert, Christine Fernandez-Maloigne, Christophe N Habas, Rémy Guillevin

► To cite this version:

Alexandre Fenneteau, Pascal Bourdon, David Helbert, Christine Fernandez-Maloigne, Christophe N Habas, et al.. Apprentissage d'un CNN sur la segmentation des lésions de sclérose en plaques avec auto-supervision. RFIAP, Jun 2020, Vannes, France. hal-03541662

HAL Id: hal-03541662

<https://hal.science/hal-03541662v1>

Submitted on 4 Apr 2023

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.

Apprentissage d'un CNN sur la segmentation des lésions de sclérose en plaques avec auto-supervision.

A. Fenneteau^{1 2 3}

P. Bourdon^{2 3}

D. Helbert^{2 3}

C. Fernandez-Maloigne^{2 3}

C. Habas^{1 3}

R. Guillevin^{3 4 5}

¹ Service neuroradiologie, Hôpital des Quinze Vingts ; Paris, France

² Institut XLIM, Université de Poitiers, UMR CNRS 7252 ; Poitiers, France

³ I3M, Laboratoire commun CNRS-Siemens, Université et CHU de Poitiers ; Poitiers, France

⁴ CHU de Poitiers ; Poitiers, France

⁵ Laboratoire DACTIM-MIS/LMA Université de Poitiers, UMR CNRS 7348 ; Poitiers, France

alexandre.fenneteau01@univ-poitiers.fr

Résumé

La sclérose en plaques (SEP) est une maladie auto-immune chronique, souvent invalidante, affectant le système nerveux central et caractérisée par une démyélinisation et des altérations neuropathiques. Les images par résonance magnétique (IRM) jouent un rôle central dans le diagnostic et le dépistage de la SEP.

Nous proposons une nouvelle méthode de segmentation des lésions de SEP pour les images IRM, basée sur les réseaux neuronaux convolutifs (CNN) et l'auto-supervision partielle et avons étudié les avantages et les inconvénients de l'utilisation de l'auto-supervision pour la tâche de segmentation. En étudiant la transférabilité en gelant les premières couches convolutives, nous avons découvert que des améliorations sont obtenues lorsque le CNN est réappris à partir des premières couches.

Mots Clef

Auto-supervision, imagerie médicale, apprentissage profond, segmentation en IRM, sclérose en plaques.

Abstract

Multiple Sclerosis (MS) is a chronic, often disabling, auto-immune disease affecting the central nervous system and characterized by demyelination and neuropathic alterations. Magnetic Resonance (MR) images plays a pivotal role in the diagnosis and the screening of MS.

We propose a novel MS lesions segmentation method for MR images, based on Convolutional Neural Networks (CNNs) and partial self-supervision and studied the pros and cons of using self-supervision for the current segmentation task. Investigating the transferability by freezing the firsts convolutional layers, we discovered that improvements are obtained when the CNN is retrained from the first layers.

Keywords

Self-supervision, Medical Imaging, Deep learning, MRI Segmentation, Multiple sclerosis.

1 Introduction

La sclérose en plaques (SEP) est une maladie auto-immune du système nerveux central. Elle affecte 1 à plus de 200 personnes sur 100 000 selon la région [28], elle apparaît généralement vers 30 ans [1] et peut rapidement induire un handicap élevé [19]. L'imagerie par résonance magnétique (IRM) est l'un des examens les plus précieux pour le diagnostic, le pronostic et le suivi de MS [34]. Les images IRM permettent d'identifier, de localiser, de compter et de déterminer l'activité des lésions démyélinisantes ; cette procédure semble être une tâche répétitive et longue, et est souvent accomplie avec une assistance virtuelle basée sur la vision par ordinateur avec une variabilité inter-observateur possible [5].

L'intérêt pour la segmentation des lésions automatiques de la substance blanche (WM) et en particulier des lésions de SEP a augmenté de manière significative au cours de la dernière décennie. Plusieurs challenges, tels que le *Medical Imaging Computing & Computer Assisted Intervention (MICCAI) 2016* [7] ont été menés pour une meilleure évaluation des performances au sein de la communauté de la vision par ordinateur. Des ensembles de données patient annotés ont également été rendus publics, permettant d'explorer plus facilement la capacité des algorithmes d'apprentissage automatique tels que les CNN à synthétiser la sémantique d'images médicales. Les dernières recherches dans ce domaine étudient l'importance de certains paramètres et suggèrent différentes techniques pour améliorer la segmentation, la plupart utilisent des CNN. Nair *et al.* proposent dans [25] de recourir au *Montecarlo dropout* pour accéder à des indicateurs de segmentation lors d'un

prediction tels que la variabilité. Roy *et al.* présentent dans [31] leur réseau neuronal convolutionnel et montrent des améliorations en augmentant la taille des patches. Hashemi *et al.* ont adapté un U-net 3D dans [13] avec des couches denses dans les voies de l'encodeur et du décodeur obtenant de meilleurs résultats, dans leur article, ils étudient l'influence des paramètres de la fonction de coût et la stratégie de fusion des patches. L'équipe Valverde *et al.* dans [38] s'est intéressé à la réutilisation de leur CNN sur des images médicale provenant d'autres scanners avec différents protocoles et ont montré que de bons résultats peuvent être obtenus avec peu de nouvelles annotations et de réapprentissage. McKinley *et al.* [24] montrent que la segmentation simultanée des lésions de substance blanche et des tissus cérébraux améliore la qualité de la segmentation et Brosch *et al.* [4] pré-entraînent leur CNN avec une machine retrainée de Boltzmann pour améliorer les performances de segmentation.

Toutes les études susmentionnées utilisent des CNN sur des images 2D ou des volumes 3D, certaines utilisant des patches et d'autres des images ou des volumes complets à partir d'images IRM multimodales. La plupart d'entre elles utilisent une architecture de réseau neuronal conçue comme un encodeur-décodeur plus ou moins proches d'un U-net [30]. Seul Valverde *et al.* [38] considèrent la tâche de segmentation comme une classification voxel par voxel.

Alors que les réseaux de neurones convolutionnels sont progressivement devenus une référence en matière de vision par ordinateur en tant que technique supervisée fortement axée sur les données [39, 29], son utilisation en imagerie médicale est souvent limitée par la faible quantité d'échantillons disponibles. Une autre difficulté consiste à obtenir des annotations d'experts de la part des radiologues car cela nécessite beaucoup de temps, ce qui n'est pas toujours possible [21]. Heureusement, des techniques telles que l'augmentation des données ou le transfert d'apprentissage ont été proposés pour surmonter ces limitations [22].

En raison de la grande quantité d'images IRM non annotées disponibles dans un hôpital, nous proposons d'améliorer simplement les résultats de segmentation en les exploitant conjointement avec des images annotées. Ce processus peut être adapté à presque toutes les techniques et tâches en IRM et en tomographie assistée par ordinateur (CT) en utilisant l'auto-supervision.

La technique d'auto-supervision introduite par Doersch *et al.* dans [8] vise à réduire la nécessité d'un grand nombre d'échantillons annotés. Cette technique est basée sur l'apprentissage contextuel [33] et le transfert d'apprentissage [27]. Elle repose sur la formation d'un réseau de neurones sur une tâche non supervisée (c'est-à-dire qui n'a pas besoin d'être annotée) et apparemment sans utilité pour apprendre le contexte, et pour réutiliser partiellement le réseau résultant pour apprendre l'inférence sur la tâche cible initiale supervisée. Comme le réseau est déjà formé pour comprendre les données d'image entrantes sur une tâche

similaire mais plus simple, moins d'échantillons annotés sont nécessaires.

À notre connaissance, seules quelques études comme [6, 2] ont fait état de l'utilisation de l'auto-supervision dans l'analyse de l'imagerie médicale. Nous proposons une nouvelle stratégie pour apprendre le contexte des images avec l'auto-supervision en imagerie médicale et pour évaluer ses performances en segmentation des lésions de la sclérose en plaques.

Nous supposons que la mise en œuvre de l'auto-supervision sur un CNN peut guider la tâche d'apprentissage supervisée au sein d'un CNN déjà conçu pour l'analyse d'images IRM, réduisant ainsi le taux d'erreur et améliorant la qualité globale de la segmentation.

Il est important d'étudier le potentiel de l'auto-supervision en imagerie médicale car ce domaine évolue très rapidement, entraînant une augmentation exponentielle des espaces de données déjà de grande dimension pour lesquels seuls quelques échantillons sont disponibles. Dans notre cas, nous prévoyons de travailler sur de multiples séquences IRM dans les années à venir pour étudier la SEP, en commençant par très peu de patients. Dans ce contexte, nous aurons besoin de nouvelles stratégies pour l'augmentation des données ou l'apprentissage de modèles avancés.

2 Matériel et méthode

Le réseau de neurones décrit par Isensee *et al.* [15] a été choisi en raison de son efficacité prouvée en matière de segmentation. Il a obtenu la troisième place au MICCAI Brain Tumor Segmentation (BraTS) challenge 2017 et utilise une architecture d'encodeur-décodeur comme dans la plupart des articles récents sur la segmentation de l'imagerie médicale. Nous l'avons adapté pour prendre cinq patches 3D en entrée, respectivement extraits des images pondérées en T1 (T1W), T1 avec un agent de contraste (T1Wc), T2 (T2W), T2 en *Fluid Attenuated Inversion Recovery* (FLAIR) et en densité de protons (PDW), car ces séquences sont généralement analysées par le radiologue pour évaluer la présence de lésions de sclérose en plaques.

La tâche cible initiale non supervisée, consiste à prédire la position d'un patch dans son volume d'extraction. Nous espérons que cette tâche de régression à sorties multiples partage des caractéristiques ou *features* en commun avec la tâche de segmentation des lésions cérébrales. Pour réaliser la première tâche cible non supervisée nous n'avons appris qu'une version tronquée du réseau de neurone de segmentation, uniquement la partie encodeur, car la tâche de régression ne nécessite pas reconstruction. Ce premier réseau a été entraîné pour prédire les coordonnées d'extraction d'un patch 3D (x , y et z). Les poids de l'encodeur entraîné ont été transférés à l'encodeur de segmentation pour la tâche supervisée de segmentation des lésions de la sclérose en plaques. Pendant l'entraînement, la segmentation a été effectuée et évaluée sur des patches. La méthode est résumée en Figure 1.

Nous avons travaillé avec des images normalisées ali-

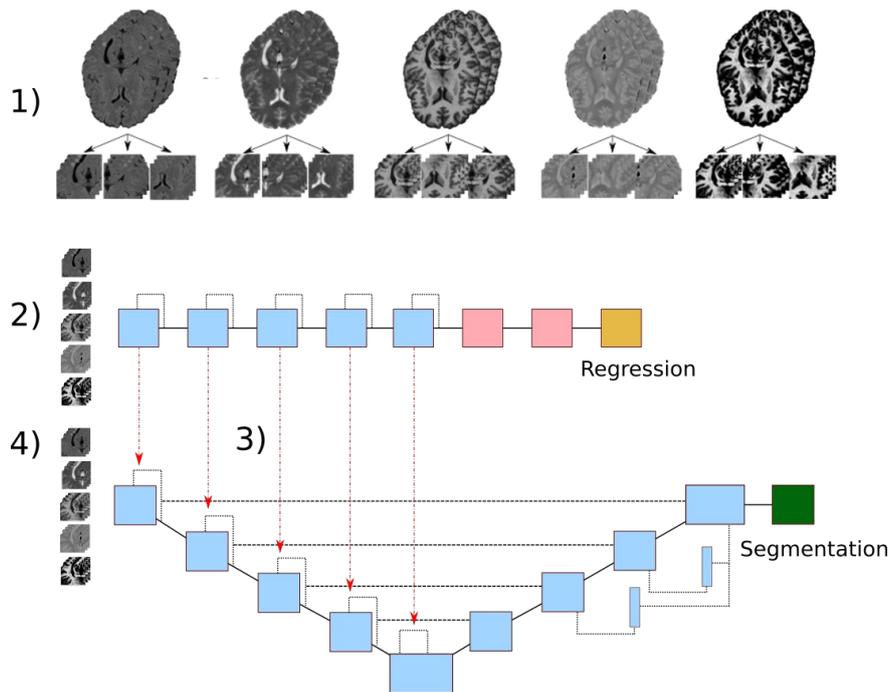


FIGURE 1 – *Workflow* de segmentation des lésions de la sclérose en plaques : 1) Les patches 3D de plusieurs séquences sont extraits à partir des volumes entiers. 2) Les patches sont envoyés au premier réseau neuronal pour l'apprentissage de la tâche de localisation. 3) Les poids des couches sont transférés au CNN de segmentation. 4) Apprentissage de la tâche de segmentation.

gnées les unes par rapport aux autres et redimensionnées à $128 \times 128 \times 128$ voxels avec des patch de $32 \times 32 \times 32$ voxels. L'objectif est garder un bon compromis pour avoir des patch assez grands pour capturer les lésions avec leur environnement et à la fois assez petits pour avoir le plus de sous-échantillon possible afin de bénéficier d'un maximum d'éléments d'apprentissage. Un tel choix a déjà donné de bons résultats avec Roy *et al.* [31]. Notre CNN n'a pas été entraîné avec un autre forme d'augmentation des données car cela augmente considérablement la durée des entraînements bien que cela puisse augmenter les performances. Pour les examens ne disposant pas de toutes les séquences d'acquisition utilisées (par exemple T1Wc ou PDW), celles-ci ont été remplacées lors de l'apprentissage par des volumes « vides » comme extension de la technique du *sequence dropout* détaillée dans [9]. Cette technique consiste à fixer aléatoirement les modalités d'entrée à zéro afin de garantir que le modèle d'ajustement CNN est suffisamment générique pour fournir des prévisions fiables même lorsque les modalités sont manquantes. Nous appliquons la méthode de validation croisée par sous-échantillonnage aléatoire répété pendant la formation. Un lot est constitué de 12 patches sélectionnés au hasard du même patient avec au moins 75 % du cerveau ce qui évite de prendre en compte les patches sans cerveau. Le CNN a été entraîné à maximiser le score de Dice (équation 3) qui mesure le taux de similarité entre deux échantillons. Son calcul est un compromis entre la précision (équation

$$Sensibilite = \frac{VP}{P} \quad (1)$$

$$Precision = \frac{VP}{VP + FP} \quad (2)$$

$$Dice = \frac{2VP}{2VP + FP + FN} \quad (3)$$

$$Dice = 2 \cdot \frac{Precision \cdot Sensibilite}{Precision + Sensibilite} \quad (4)$$

FIGURE 2 – Décomposition du calcul du score de Dice. VP pour le nombre de vrai positif, FP pour le nombre de faux positif, P pour le nombre de positif.

2) et la sensibilité 1 comme détaillé dans l'équation 4 en Figure 2 visant à réduire l'impact d'une distribution déséquilibrée des données sur l'évaluation des résultats. L'optimiseur Adam [18] est utilisé pour aider à réduire le temps de convergence.

3 Résultats

Quatre ensembles de données publiques différents de segmentation des lésions de la sclérose en plaques ont été recueillies provenant des challenges de segmentation des lésions de la sclérose en plaques du MICCAI 2008 [32] et 2016 [7], le challenge de segmentation de la sclérose

en plaques 2015 du Symposium international d'imagerie biomédicale (ISBI) [5] consistant en 21 examens de 5 patients différents à différents moments, et la base de données publiques de Lesjak *et al.* [20]. Deux autres ensembles de données publiques provenant d'un sous-ensemble sélectionné manuellement d'OASIS3 [23] avec des patients cognitivement normaux et en déclin et du MICCAI BraTS challenge 2017 [3] avec des tumeurs cérébrales ont été utilisés comme source d'images supplémentaires. Le tableau 1 donne des informations supplémentaires sur les données. L'ensemble de données MICCAI 2016 a été choisi comme test pour avoir des résultats avec des examens complets, sans séquences manquantes.

TABLE 1 – Détails concernant les ensembles de données

Données	Examens	Séquence	SEP
MICCAI 2008	20	T1W, T2W, FLAIR	OUI
MICCAI 2016	15	T1W, T1Wc, T2W, FLAIR, PDW	OUI
ISBI 2015	21	T1W, T2W, FLAIR and PDW	OUI
Lesjak <i>et al.</i>	30	T1W, T1Wc, T2W, FLAIR	OUI
OASIS3	82	T1W, T2W, FLAIR	NON
BraTS 2017	285	T1W, T1Wc, T2W, FLAIR	NON

Toutes les images des ensembles de données ont été prétraitées pour être aussi comparables que possible. Le prétraitement est très proche de celui effectué par [5] pour le challenge ISBI. Lorsque c'était possible, les images non traitées ont été utilisées. Après la première correction biais de champ N4 [35], nous avons choisi d'aligner toutes les séquences sur le FLAIR car la plupart des lésions de nos jeux de données ont été segmentées dans l'espace FLAIR. L'alignement a été effectué par l'outil FSL FLIRT [17, 12, 16]. Pour chaque modalité l'« histogram matching » [26] est appliqué séparément. Les images de référence pour les histogrammes ont été choisies parmi les images de bonne qualité de nos jeux de données à cette étape du prétraitement. L'image T1 est ensuite dépouillée du crâne [14] et enregistrée sur un modèle de cerveau MNI de 1 mm [10]. Le masque cérébral et la transformation sont ensuite appliqués aux autres séquences. Le pipeline global est illustré dans la figure 3. Toutes les images sont redimensionnées à $128 \times 128 \times 128$ voxels avec une résolution de $1,422 \times 1,703 \times 1,422$ mm³ pour travailler avec une taille d'image raisonnable et une résolution acceptable pour distinguer les lésions.

L'apprentissage préalable de la tâche de régression non supervisée a été effectuée avec tous les examens d'entraînement, y compris ceux ne contenant aucun patient atteint

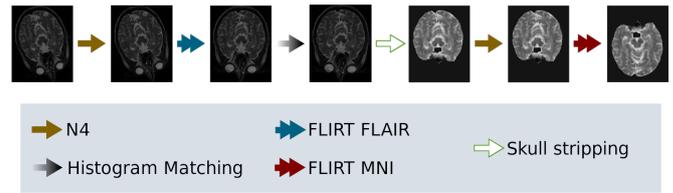


FIGURE 3 – Step of the pipeline illustrated with a T2 image.

de SEP. Un ratio de 50 % des examens de SEP a été respecté afin de garantir que le réseau puisse construire des cartes de caractéristiques générales à partir de différentes textures d'IRM du cerveau, en évitant le déséquilibre entre les examens de SEP et ceux qui n'en ont pas. Au cours du second apprentissage supervisé sur la segmentation des lésions, seuls les ensembles de données de la sclérose en plaques ont été utilisés. Lors de l'apprentissage, la segmentation cible est choisie au hasard à partir des différentes segmentations manuelles disponibles, certains examens pouvant en contenir jusqu'à quatre différentes selon le nombre d'experts médicaux impliqués. Comme la délimitation manuelle des lésions peut varier d'un expert à l'autre, nous supposons que l'utilisation de ces cartes individuelles au lieu de combinaisons peut aider la CNN à établir son propre consensus.

Pour chaque patient, 150 patchs choisis au hasard dans la zone du cerveau ont été segmentés. La segmentation finale a été obtenue en faisant la moyenne de toutes les prédictions qui se chevauchaient. Elle a permis d'obtenir une prédiction consensuelle couvrant plus de deux fois le volume du cerveau.

Cinq stratégies de transfert ont été évaluées :

- Transferts de poids avec les six premières couches convolutives gelées (6CF)
- Transferts de poids avec les trois premières couches convolutives gelées (3CF)
- Transferts de poids avec les deux premières couches convolutives gelées (2CF)
- Transferts de poids avec la première couche convolutive gelée (1CF)
- Transferts de poids sans gel de couche (0CF)

Pour comparaison, nous avons entraîné une version sans auto-supervision avec des initialisés selon la méthode uniforme glot [11], uniquement à la segmentation des lésions de la sclérose en plaques (VAN). La mesure d'évaluation des similitudes utilisée pour la comparaison est le score de Dice.

Les résultats moyens sont présentés dans le tableau 2 et le diagramme en boîte montrant la distribution des résultats dans l'ensemble de test est présenté dans la figure 4. L'évaluation a été réalisée sur la base de tests MICCAI 2016 acquise avec un scanner IRM et des protocoles différents des bases d'entraînement. La segmentation de référence pour l'évaluation est formée par le consensus de la segmentation des différents radiologues. Nous avons obtenu d'excel-

TABLE 2 – Évaluation des apprentissages.

Apprentissage	Dice	Sensibilité	Précision
VAN	66.94%	60.63%	78.84%
0CF	67.63%	61.47%	79.30%
1CF	66.89%	59.79%	80.31%
2CF	64.44%	57.50%	77.53%
3CF	62.69%	55.86%	75.06%
6CF	57.93%	49.60%	75.03%

lentes performances par rapport aux publications récentes, tout en gardant à l'esprit que le test n'est pas le même et ne peut pas être entièrement comparé à d'autres études telles que celle de Valverde *et al.* [37] qui a obtenu un score de Dice de 53,5%, Valcarcel *et al.* [36] avec 56% et 57%, ou Roy *et al.* [31] a obtenu 56,39%. La bonne qualité de la segmentation pour toutes les versions indique que l'architecture du CNN, la méthode d'apprentissage globale sont pertinents pour notre tâche de segmentation. Le meilleur score de Dice et la meilleure sensibilité sont obtenus avec la méthode 0CF et la meilleure précision est atteinte avec la méthode 1CF.

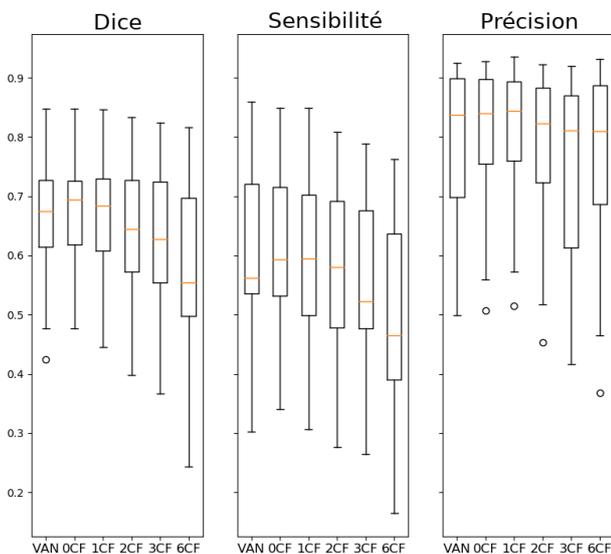


FIGURE 4 – Diagramme en boîte comparant les résultats obtenus pour chaque méthode d'apprentissage. La médiane est représenté par la ligne horizontale orange.

La Figure 4 montre que l'augmentation du nombre de couches gelées diminue le score de Dice et la sensibilité et augmente la dispersion des résultats. La précision est également diminuée avec le nombre de couches gelées, mais moins que les autres mesures.

4 Discussion

Seule la version 0CF de l'auto-supervision a surpassé pour chaque mesure la méthode VAN. Notre hypothèse initiale et notre motivation pour l'auto-supervision étaient que les tâches de régression non supervisée et de segmentation supervisée partageraient les premières couches convolutives du modèle, comme cela est souvent le cas en vision par ordinateur. Les résultats indiquent cependant que ce n'est pas le cas et que les toutes premières couches doivent être ré-apprenties pour une meilleure segmentation. Une hypothèse possible est que la tâche de segmentation des lésions de la SEP est très spécialisée et nécessite un modèle très spécialisé, par opposition aux tâches quotidiennes de reconnaissance ou de classification des objets visuels qui sont supposées partager un terrain commun.

Le diagramme en boîte de la Figure 4 montre que le gel des couches convolutives a diminué le score de Dice et la sensibilité dès le début. La perte de précision semble n'apparaître qu'après le gel des deux premières couches. Ces résultats suggèrent que le gel des premières couches convolutionnelles affecterait principalement la sensibilité de la prédiction. Ceci est donc en faveur de l'hypothèse selon laquelle les premières couches du CNN captureraient une grande partie des informations utiles de l'image dans les toutes premières couches. Cela suggère également que la sélection et la collecte de la plupart des caractéristiques discriminantes sont réalisées progressivement après la première couche.

La Figure 5 illustre une visualisation de la segmentation réalisée avec notre technique. Nous pouvons voir sur cette image que l'augmentation du nombre de couches gelées diminue la taille de la zone segmentée et augmente la taille de la zone correspondant au faux positif dans l'ellipse au bord frontal gauche du ventricule. Cela peut révéler un sous-apprentissage, car cette zone contient généralement des lésions.

Nos résultats confirment l'idée que la régression et la segmentation sont des tâches vraiment différentes, même si elles sont effectuées avec des images similaires. Une des raisons peut être que le nombre d'échantillon d'apprentissage est trop important pour observer l'avantage du pré-apprentissage. Les observations concordent avec l'hypothèse énoncée dans l'article de Yosinski *et al.* [40]. Il y est avancé que la spécialisation du réseau neuronal augmente avec la profondeur et que l'apprentissage par transfert semble toujours apporter des améliorations même après ré-apprentissage dès la première couche. Contrairement à la méthode que nous proposons, dans l'article de Yosinski *et al.* les deux premières couches sont assez générales pour les deux tâches, cependant leur évaluation porte sur deux tâches de classification très proches alors que notre étude porte sur la tâche de régression et de segmentation.

5 Conclusion

Nous proposons une nouvelle façon de mettre en œuvre et d'utiliser les méthodes d'auto-supervision en imagerie

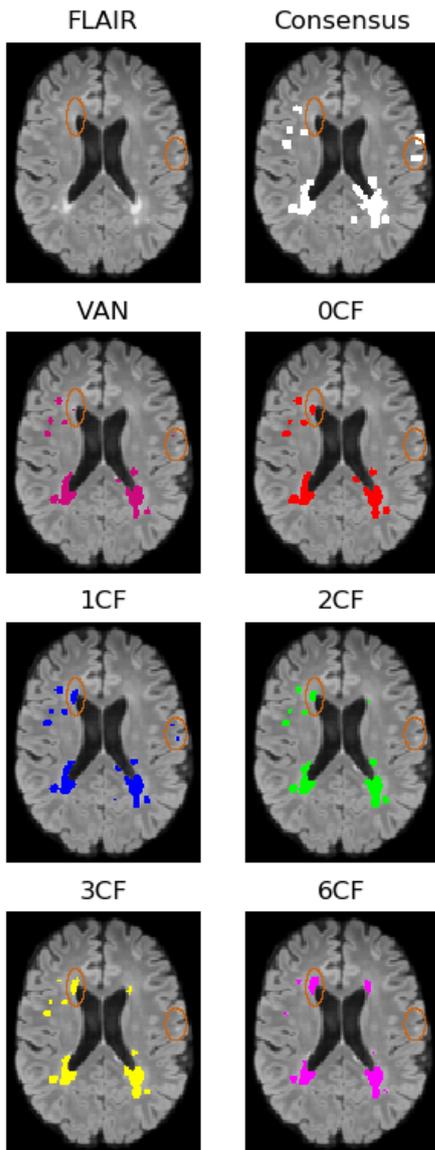


FIGURE 5 – Exemple de visualisation des segmentations prédites. Les principales zones de différence sont entourées par des ellipses oranges.

médicale en utilisant la localisation de patch au sein d'une image comme une tâche non supervisée. L'objectif est de réutiliser le modèle appris pour affiner une tâche supervisée d'intérêt. Nous avons obtenu de bons résultats globaux par rapport à l'état de l'art. Notre technique améliore la qualité de la segmentation des lésions de la SEP, bien que l'amélioration soit plus faible qu'espérée et dû à des facteurs différents de ceux escomptés. Nous avons observé que même les premières couches de notre CNN semblent être spécialisées pour la tâche de segmentation et ne sont pas aussi générales qu'on pourrait le penser. Les premières couches semblent être plus impliquées dans la sensibilité de notre CNN, ce qui laisse penser que la discrimination des informations significatives serait effectuée après ces couches.

Nos conclusions soutiennent l'idée que l'utilisation de l'auto-supervision pour des tâches d'analyse d'image médicale ne semble pas être aisée et nécessite des recherches plus approfondies pour définir les limites de la similarité et de la hiérarchie des tâches de vision. Cependant, nous pensons que les techniques d'auto-supervision pourraient être davantage utilisées et que le fait de trouver une bonne première tâche non supervisée à apprendre peut conduire à des améliorations dans le domaine de l'imagerie médicale.

Références

- [1] E D Acheson. Epidemiology of multiple sclerosis. *British Medical Bulletin*, 33(1) :9–14, 01 1977.
- [2] Wenjia Bai, Chen Chen, Giacomo Tarroni, Jinming Duan, Florian Guitton, Steffen E Petersen, Yike Guo, Paul M Matthews, and Daniel Rueckert. Self-supervised learning for cardiac mr image segmentation by anatomical position prediction. *arXiv preprint arXiv :1907.02757*, 2019.
- [3] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4 :170117, 2017.
- [4] T. Brosch, L. Y. W. Tang, Y. Yoo, D. K. B. Li, A. Troubleshoot, and R. Tam. Deep 3d convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation. *IEEE Transactions on Medical Imaging*, 35(5) :1229–1239, May 2016.
- [5] Aaron Carass, Snehashis Roy, Amod Jog, Jennifer L Cuzzocreo, Elizabeth Magrath, Adrian Gherman, Julia Button, James Nguyen, Ferran Prados, Carole H Sudre, et al. Longitudinal multiple sclerosis lesion segmentation : resource and challenge. *NeuroImage*, 148 :77–102, 2017.
- [6] Liang Chen, Paul Bentley, Kensaku Mori, Kazunari Misawa, Michitaka Fujiwara, and Daniel Rueckert. Self-supervised learning for medical image analysis using image context restoration. *Medical image analysis*, 58 :101539, 2019.
- [7] Olivier Commowick, Frederic Cervenansky, and Roxana Ameli. Msseg challenge proceedings : Multiple sclerosis lesions segmentation challenge using a data management and processing infrastructure, 2016.
- [8] Carl Doersch, Abhinav Gupta, and Alexei A Efros. Unsupervised visual representation learning by context prediction. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1422–1430, 2015.
- [9] Yushan Feng, Huitong Pan, Craig H. Meyer, and Xue Feng. A self-adaptive network for multiple sclerosis lesion segmentation from multi-contrast mri with

- various imaging protocols. *CoRR*, abs/1811.07491, 2018.
- [10] Vladimir S Fonov, Alan C Evans, Robert C McKinsty, CR Alml, and DL Collins. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. *NeuroImage*, (47) :S102, 2009.
- [11] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256, 2010.
- [12] Douglas N Greve and Bruce Fischl. Accurate and robust brain image alignment using boundary-based registration. *Neuroimage*, 48(1) :63–72, 2009.
- [13] Seyed Raein Hashemi, Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, Sanjay P Prabhu, Simon K Warfield, and Ali Gholipour. Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation : Application to multiple sclerosis lesion detection. *IEEE Access*, 7 :1721–1735, 2019.
- [14] J. E. Iglesias, C. Liu, P. M. Thompson, and Z. Tu. Robust brain extraction across datasets and comparison with publicly available methods. *IEEE Transactions on Medical Imaging*, 30(9) :1617–1634, Sep. 2011.
- [15] Fabian Isensee, Philipp Kickingereder, Wolfgang Wick, Martin Bendszus, and Klaus H Maier-Hein. Brain tumor segmentation and radiomics survival prediction : Contribution to the brats 2017 challenge. In *International MICCAI Brainlesion Workshop*, pages 287–297. Springer, 2017.
- [16] Mark Jenkinson, Peter Bannister, Michael Brady, and Stephen Smith. Improved optimization for the robust and accurate linear registration and motion correction of brain images. *Neuroimage*, 17(2) :825–841, 2002.
- [17] Mark Jenkinson and Stephen Smith. A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2) :143–156, 2001.
- [18] Diederik P Kingma and Jimmy Ba. Adam : A method for stochastic optimization. *arXiv preprint arXiv :1412.6980*, 2014.
- [19] Ilya Kister, Eric Chamot, Amber R Salter, Gary R Cutter, Tamar E Bacon, and Joseph Herbert. Disability in multiple sclerosis : a reference for patients and clinicians. *Neurology*, 80(11) :1018–1024, 2013.
- [20] Žiga Lesjak, Alfiia Galimzianova, Aleš Koren, Matej Lukin, Franjo Pernuš, Boštjan Likar, and Žiga Spiclin. A novel public mr image dataset of multiple sclerosis patients with lesion segmentations based on multi-rater consensus. *Neuroinformatics*, 16(1) :51–63, Jan 2018.
- [21] 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, 2017.
- [22] Alexander Selvikvåg Lundervold and Arvid Lundervold. An overview of deep learning in medical imaging focusing on mri. *Zeitschrift für Medizinische Physik*, 29(2) :102–127, 2019.
- [23] Daniel S Marcus, Anthony F Fotenos, John G Csernansky, John C Morris, and Randy L Buckner. Open access series of imaging studies : longitudinal mri data in nondemented and demented older adults. *Journal of cognitive neuroscience*, 22(12) :2677–2684, 2010.
- [24] Richard McKinley, Rik Wepfer, Fabian Aschwend, Lorenz Grunder, Raphaela Muri, Christian Rummel, Rajeev Verma, Christian Weisstanner, Mauricio Reyes, Anke Salmen, et al. Simultaneous lesion and neuroanatomy segmentation in multiple sclerosis using deep neural networks. *arXiv preprint arXiv :1901.07419*, 2019.
- [25] Tanya Nair, Doina Precup, Douglas L Arnold, and Tal Arbel. Exploring uncertainty measures in deep networks for multiple sclerosis lesion detection and segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 655–663. Springer, 2018.
- [26] László G Nyúl, Jayaram K Udupa, and Xuan Zhang. New variants of a method of mri scale standardization. *IEEE transactions on medical imaging*, 19(2) :143–150, 2000.
- [27] SJ Pan and Q Yang. A survey on transfer learning. *IEEE transaction on knowledge discovery and data engineering*, 22 (10), 2010.
- [28] Maura Pugliatti, Stefano Sotgiu, and Giulio Rosati. The worldwide prevalence of multiple sclerosis. *Clinical neurology and neurosurgery*, 104(3) :182–191, 2002.
- [29] Waseem Rawat and Zenghui Wang. Deep convolutional neural networks for image classification : A comprehensive review neural computation, 2017.
- [30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net : Convolutional networks for biomedical image segmentation. In *MICCAI*, volume 9351 of *LNCS*, pages 234–241. Springer, 2015. (available on arXiv :1505.04597 [cs.CV]).
- [31] Snehashis Roy, John A Butman, Daniel S Reich, Peter A Calabresi, and Dzung L Pham. Multiple sclerosis lesion segmentation from brain mri via fully convolutional neural networks. *arXiv preprint arXiv :1803.09172*, 2018.
- [32] Martin Styner, Joohee Lee, Brian Chin, Matthew S Chin, Olivier Commowick, Hoai-Huong Tran, Valerie Jewells, and Simon Warfield. 3d segmentation

in the clinic : A grand challenge ii : Ms lesion segmentation. *MIDAS Journal*, 11 2007.

- [33] Jin Sun and David W Jacobs. Seeing what is not there : Learning context to determine where objects are missing. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5716–5724, 2017.
- [34] Alan J Thompson, Brenda L Banwell, Frederik Barkhof, William M Carroll, Timothy Coetzee, Giancarlo Comi, Jorge Correale, Franz Fazekas, Massimo Filippi, Mark S Freedman, et al. Diagnosis of multiple sclerosis : 2017 revisions of the mcdonald criteria. *The Lancet Neurology*, 17(2) :162–173, 2018.
- [35] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee. N4itk : Improved n3 bias correction. *IEEE Transactions on Medical Imaging*, 29(6) :1310–1320, June 2010.
- [36] Alessandra M Valcarcel, Kristin A Linn, Simon N Vandekar, Theodore D Satterthwaite, John Muschelli, Peter A Calabresi, Dzung L Pham, Melissa Lynne Martin, and Russell T Shinohara. Mimosa : an automated method for intermodal segmentation analysis of multiple sclerosis brain lesions. *Journal of Neuroimaging*, 28(4) :389–398, 2018.
- [37] Sergi Valverde, Mariano Cabezas, Eloy Roura, Sandra González-Vilà, Deborah Pareto, Joan C Vilanova, Lluís Ramió-Torrentà, Àlex Rovira, Arnau Oliver, and Xavier Lladó. Improving automated multiple sclerosis lesion segmentation with a cascaded 3d convolutional neural network approach. *NeuroImage*, 155 :159–168, 2017.
- [38] Sergi Valverde, Mostafa Salem, Mariano Cabezas, Deborah Pareto, Joan Carles Vilanova, Lluís Ramió-Torrentà, Alex Rovira, Joaquim Salvi, Arnau Oliver, and Xavier Lladó. One-shot domain adaptation in multiple sclerosis lesion segmentation using convolutional neural networks. *NeuroImage. Clinical*, page 101638, 2018.
- [39] Athanasios Voulodimos, Nikolaos Doulamis, Anastasios Doulamis, and Eftychios Protopapadakis. Deep learning for computer vision : A brief review. *Computational intelligence and neuroscience*, 2018, 2018.
- [40] Jason Yosinski, Jeff Clune, Yoshua Bengio, and Hod Lipson. How transferable are features in deep neural networks ? In *Advances in neural information processing systems*, pages 3320–3328, 2014.