

Segmentation d'IRM de cerveaux de nouveau-nés en quelques secondes à l'aide d'un réseau de neurones convolutif *pseudo-3D* et de transfert d'apprentissage

Yongchao XU¹, Thierry GÉRAUD¹, Isabelle BLOCH²

¹EPITA Research and Development Laboratory (LRDE)
Le Kremlin-Bicêtre, France

²LTCI, Télécom ParisTech
Université Paris-Saclay, France

yongchao.xu@lrde.epita.fr, thierry.geraud@lrde.epita.fr
isabelle.bloch@telecom-paristech.fr

Résumé – L'imagerie par résonance magnétique (IRM) du cerveau est utilisée sur les nouveau-nés pour évaluer l'évolution du cerveau et diagnostiquer des maladies neurologiques. Ces examens nécessitent souvent une analyse quantitative des différents tissus du cerveau, de sorte qu'avoir une segmentation précise est essentiel. Dans cet article, nous proposons une méthode automatique rapide de segmentation en différents tissus des images IRM 3D de cerveaux de nouveau-nés ; elle utilise un réseau de neurones totalement convolutif (FCN) et du transfert d'apprentissage. Par rapport aux approches similaires qui reposent soit sur des patches 2D ou 3D, soit sur des FCN totalement 3D, notre méthode est beaucoup plus rapide : elle ne prend que quelques secondes, et une seule modalité (T2) est nécessaire. Afin de prendre les informations 3D en compte, trois coupes 2D successives sont empilées pour former une image 2D en couleurs, dont l'ensemble sur tout le volume sert d'entrée à un FCN, pré-entraîné sur ImageNet pour la classification d'images naturelles. Nos expériences sur un ensemble de données de référence montrent que notre méthode obtient des résultats du niveau de ceux de l'état de l'art.

Abstract – Magnetic resonance imaging (MRI) is widely used to assess brain development in neonates and to diagnose a wide range of neurological diseases. Such studies often require a quantitative analysis of different brain tissues, so it is essential to be able to classify them accurately. In this paper, we propose a fast automatic method that segments 3D brain MR images into different tissues using fully convolutional network (FCN) and transfer learning. As compared to existing deep learning-based approaches that rely either on 2D patches or on fully 3D FCN, our method is way much faster: it only takes a few seconds, and only a single modality (T2) is required. In order to take the 3D information into account, all three successive 2D slices are stacked to form a set of 2D color images, which serve as input for the FCN pre-trained on ImageNet for natural image classification. To the best of our knowledge, this is the first method that applies transfer learning to segment neonatal brain 3D MR images. Our experiments on a reference public dataset show that our method achieves state-of-the-art results.

1 Introduction

Une segmentation automatique et précise en différents tissus des images par résonance magnétique (IRM) du cerveau des nouveau-nés est une étape importante permettant d'évaluer l'évolution néonatale du cerveau et de diagnostiquer de nombreuses pathologies neurologiques. Obtenir une segmentation de qualité est difficile dans le cas des nouveau-nés car l'effet de volume partiel est significatif et le contraste est faible entre les matières blanches et grises.

Dans la suite, nous nous intéressons aux tissus suivants : matière grise corticale (CoGM), ganglions de la base et thalamus (BGT), substance blanche myélinisée (MWM) ou non myélinisée (UWM), tronc cérébral (BS), cervelet (CB), ventricules (Vent), et liquide céphalorachidien dans les espaces extracérébraux (CR).

Dans cet article, nous proposons une méthode rapide associant un réseau de neurones totalement convolutif (FCN) [2],

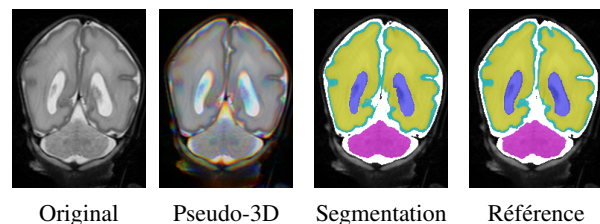


FIGURE 1 – Un exemple de résultat de segmentation (obtenu pour un volume 3D entier en quelques secondes).

du transfert d'apprentissage, et des informations *pseudo-3D*. En prenant trois coupes 2D successives d'un volume 3D nous formons une image 2D en couleurs (voir la 2e colonne de la figure 1); le volume 3D entier donne donc un ensemble d'images 2D en couleurs. Ces images sont passées en entrée sur le réseau de base d'un FCN, le réseau VGG [1]; ce dernier a été pré-entraîné sur l'ensemble de données ImageNet. Nous écartons

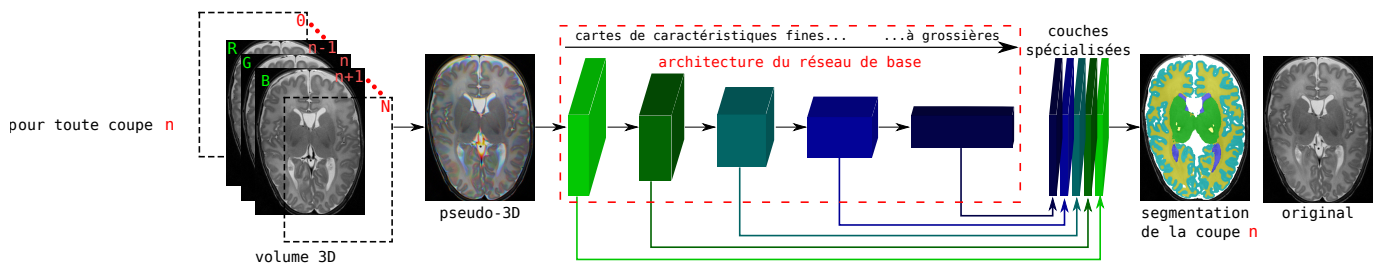


FIGURE 2 – Architecture proposée, construite sur le réseau pré-entraîné VGG [1].

les couches entièrement connectées, et nous ajoutons des couches convolutionnelles spécialisées à la fin de chacune des cinq couches de convolution de VGG. Une combinaison linéaire de ces couches spécialisées (des cartes de caractéristiques de fines à grossière) donne la segmentation finale.

Les principales contributions de cet article sont les suivantes : **1.** une méthode automatique qui segmente, grâce à l’approche pseudo-3D, tout un volume IRM 3D en quelques secondes (entre 2 et 6 s) ; **2.** l’utilisation du transfert d’apprentissage pour segmenter l’ensemble d’un volume IRM 3D, ce qui ne nécessite que très peu de volumes d’apprentissage. Par exemple, un seul volume d’apprentissage a été utilisé pour obtenir le résultat donné en figure 1.

2 État de l’art

Les méthodes de segmentation IRM du cerveau peuvent être grossièrement divisées en trois catégories : les méthodes à base d’atlas, très populaires dans le cas néonatal [3, 4] ; les méthodes de traitement d’images et de reconnaissance des formes classiques [5, 6, 7] ; les méthodes par apprentissage profond, et plus particulièrement les réseaux neuronaux convolutifs (CNN) [8, 9, 10]. Ces dernières méthodes utilisent des patches 2D, des patches 3D [11], ou une technique hybride 2,5D (3 patches pris respectivement dans les trois plans) [12, 13]. Plus de références sont données dans [14] et [15].

L’utilité pour des problèmes d’imagerie médicale d’un réseau pré-entraîné sur l’ensemble de données ImageNet a déjà été démontrée ; citons [16] pour la détection et la classification de ganglions lymphatiques, [17] pour la détection de pathologies des poumons, et [18] pour la segmentation des nerf et disque optiques en imagerie rétinienne. Il faut également noter que nous utilisons une architecture de réseau (décrite dans la section suivante) semblable à cette dernière référence.

Enfin, l’idée de former des images en couleurs (pour les entrer dans un réseau) se retrouve dans [12] et dans [19]. Cependant, formées à partir de trois patches (respectivement axial, coronal et sagittal), ce sont donc des images de patches.

3 Méthode proposée

Pré-traitement. Notre méthode de segmentation ne nécessite qu’une seule modalité d’IRM cérébrale de nouveau-né (une séquence T2). Nous commençons par une égalisation d’histogramme basique du volume entier.

Architecture pour la segmentation. Une segmentation performante d’images naturelles peut être obtenue grâce à un réseau de neurones totalement convolutif (FCN) et à l’aide de transfert d’apprentissage [2]. Ici, nous proposons d’utiliser cette même méthode pour la segmentation d’images IRM 3D du cerveau, bien que ces images soient très différentes de images naturelles. Nous nous appuyons sur le réseau VGG à 16 couches [1] pré-entraîné sur des millions d’images naturelles de la base ImageNet, et dédié à la classification d’images. Pour notre application, nous écartons les couches entièrement connectées de la fin du réseau VGG, et nous gardons les 5 étages qui forment le *réseau de base*. Ce réseau de base est composé de couches de convolution (CONV) : $z_i = W_i \times x + b_i$, de couches de correction (ReLU) : $f(z_i) = \max(0, z_i)$, et de couches de *max pooling* entre deux des étages successifs, où x est l’entrée de chaque couche de convolution, W_i est le paramètre de convolution, et b_i est le terme de polarisation. Ainsi le réseau de base produit des cartes de caractéristiques de grain fin à grossier. Inspirés par les travaux dans [2, 18], nous ajoutons une couche de convolution spécialisée (avec un noyau de taille 3×3) avec $K = 16$ cartes de caractéristiques après les couches de convolution à la fin de chaque étage. Nous re-dimensionnons toutes les couches spécialisées à la taille de l’image originale et nous les concaténons. Une dernière couche de convolution avec une taille de noyau 1×1 est ajoutée toute à la fin pour combiner linéairement les cartes de caractéristiques et produire la segmentation résultat. L’architecture de réseau proposée est représentée dans la figure 2.

Contrairement à la méthode dans [18] qui utilise une architecture similaire sur des images de rétine, qui sont naturellement 2D et en couleurs, nous avons des volumes (images 3D) en niveaux de gris. Il nous faut donc “préparer” les entrées du réseau. Pour former une image 2D en couleurs, qui servira d’entrée au réseau pour segmenter la $n^{\text{ième}}$ coupe du volume 3D, nous utilisons les coupes $n-1$, n et $n+1$, pour ses canaux respectifs rouge, vert et bleu. Si la coupe $n-1$ ou $n+1$ est hors du volume, une image constante à 0 est utilisée à sa place. Ce processus est représenté dans la figure 2 (à gauche). Chaque image 2D en couleurs forme ainsi une représentation *pseudo-3D* d’une partie (3 coupes) du volume 3D. Cette représentation permet ainsi d’intégrer des informations 3D, tout en évitant les coûts en mémoire et en calcul de l’approche 2,5D à base de patches et, *a fortiori*, des réseaux entièrement en 3D. Cette idée originale est une contribution majeure de l’article.

Pour la phase d’apprentissage, nous utilisons une fonction de perte logistique avec une loi multinomiale de classification un-

Code	Coupes	Âge	Nbr. de voxels	Taille des voxels (mm ³)	Expériences (images)	Temps
EXP1	axiales	40 semaines	512×512×50	0,35×0,35×2,0	2 apprentissage / 5 test T2	3,5 s
EXP2	coronales	30 semaines	384×384×50	0,34×0,34×2,0	2 apprentissage / 5 test T2	2,2 s
EXP3	coronales	40 semaines	512×512×110	0,35×0,35×1,2	2 + 2 apprentissage / 5 test T2	6,5 s

TABLE 1 – Données d’apprentissage et de test de NeoBrain12 [20], et expériences menées.

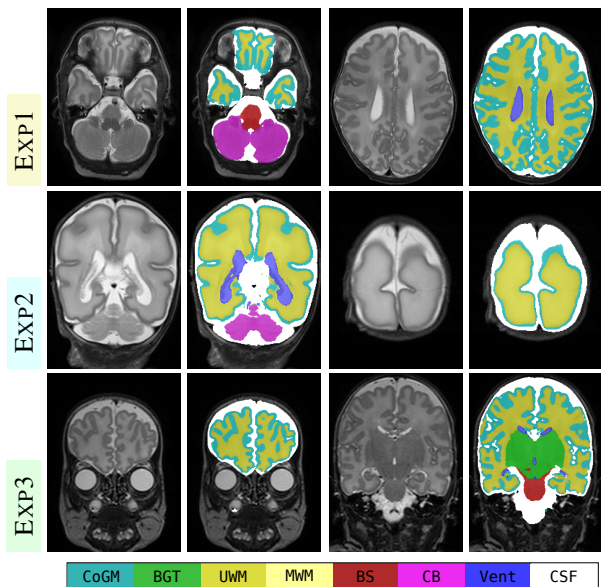


FIGURE 3 – Quelques résultats qualitatifs.

parmi-plusieurs, et les prédictions réelles sont passées à un *softmax* pour obtenir une probabilité sur les classes. Nous utilisons également la stratégie classique d’augmentation de données par mise à l’échelle et rotation, et nous soustrayons la valeur moyenne de chaque canal dans les images d’entraînement. Nous affinons l’ensemble du réseau pendant 100K itérations en utilisant un très faible taux d’apprentissage ($lr = 10^{-10}$). La fonction de perte est minimisée par une descente de gradient stochastique avec un terme d’inertie ($momentum = 0,99$ et $weight_decay = 0,005$). La fonction de perte est moyennée sur 20 images.

Pour la segmentation, après avoir pré-traité le volume 3D (par égalisation d’histogramme), nous préparons l’ensemble des images *pseudo-3D* en couleurs. Ensuite, nous soustrayons les valeurs moyennes de chaque canal, et nous passons chaque image 2D à travers le réseau.

Post-traitement. Au cours de l’inférence, une région R de chaque tranche qui est faible, ou qui ne se superpose pas de manière importante avec le masque du cerveau M (par exemple $|R| < 5mm^2$ ou $|R \cap M|/|R| < 0,8$), est automatiquement considérée comme arrière-plan. Pour chaque volume 3D, le masque M est donné par la plus grande région du résultat inféré pour la tranche médiane du volume.

4 Résultats expérimentaux

Nous avons évalué la méthode proposée sur l’ensemble de données du défi MICCAI “Neonatal Brain Segmentation 2012

(NEOBRAINS12)” [20]. Ce jeu de données, qui sert constamment à l’évaluation de nouveaux algorithmes, est considéré comme une référence. Étant données une segmentation automatique S d’un tissu et sa segmentation de référence G , l’évaluation quantitative utilise le coefficient de Dice $DC = 2 \times |S \cap G| / (|S| + |G|)$, et la distance surfacique moyenne (MSD) entre les frontières de S et G .

Nous avons réalisé trois types d’expériences, décrites dans le tableau 1, et dont les résultats quantitatifs sont donnés dans le tableau 2; quelques résultats qualitatifs de segmentation sont montrés dans la figure 3. Plus de détails et de résultats sont disponibles sur la page dédiée [21].

Pour les images axiales acquises sur des prématurés à 40 semaines d’âge corrigé (expérience EXP1) et les images coronales à 30 semaines (expérience EXP2), nous avons utilisé les deux images d’apprentissage qui étaient données. Pour les images coronales acquises à 40 semaines (EXP3), il n’y avait pas d’images d’apprentissage disponibles; nous avons donc entraîné notre réseau sur les images d’apprentissage des deux premières expériences. Le temps total d’apprentissage pour chaque modèle est d’environ cinq heures pour 100K itérations en utilisant une carte GPU NVIDIA GeForce GTX 1080.

Chaque expérience a été évaluée avec cinq images de test en utilisant le système d’évaluation en ligne de NEOBRAINS12 (voir l’URL donnée dans [20]). Les résultats de segmentation, donnés dans le tableau 2, sont au niveau des meilleurs résultats de l’état de l’art. Lors de la phase de test, le temps moyen d’exécution pour la segmentation d’un volume IRM entier est de quelques secondes (cf. la colonne la plus à droite du tableau 1). À titre de comparaison, les méthodes concurrentes ont des temps de segmentation allant de 2-3 minutes à 1 heure.

5 Conclusion

Nous avons présenté une méthode de segmentation *automatique* et *rapide* des volumes IRM de cerveaux de nouveau-nés. Nos deux contributions principales sont d’utiliser le transfert d’apprentissage pour ce type d’images, et la technique de *pseudo-3D*: des images en couleurs obtenues par empilement de coupes successives du volume. Grâce au transfert d’apprentissage, la méthode proposée permet d’obtenir de bons résultats même avec une unique image d’entraînement. Cela est particulièrement précieux dans certains problèmes de segmentation en imagerie médicale, pour lesquels un grand ensemble de données avec des annotations précises peut être difficile à obtenir. De plus, la technique de *pseudo-3D* permet d’éviter les demandes coûteuses en mémoire et en calcul des approches à base de patches, qu’ils soient 2D, 3D ou 2,5D. Dans les expérimentations, nous avons atteint l’état de l’art sur une base de données

Code	Méthode	CoGM		BGT		UWM		BS		CB		Vent		CSF	
		DC	MSD	DC	MSD	DC	MSD	DC	MSD	DC	MSD	DC	MSD	DC	MSD
EXP1	méthode proposée	0.87	0.11	0.91	0.51	0.93	0.11	0.85	0.49	0.94	0.33	0.87	0.24	0.83	0.20
	UPF_SIMBioSys [4]	0.85	0.15	0.93	0.29	0.91	0.17	0.85	0.15	0.94	0.28	0.83	0.44	0.79	0.29
	UNC-IDEA [7]	0.86	0.11	0.92	0.33	0.92	0.13	0.83	0.27	0.92	0.45	0.79	0.25	0.79	0.25
	médianes des 5 suivants	0.84	0.18	0.88	0.62	0.88	0.25	0.79	0.69	0.91	0.53	0.81	0.32	0.73	0.54
EXP2	méthode proposée	0.79	0.14	0.89	0.42	0.95	0.14	0.84	0.37	0.91	0.30	0.87	0.33	0.89	0.13
	UPF_SIMBioSys [4]	0.75	0.16	0.90	0.38	0.93	0.22	0.86	0.32	0.92	0.31	0.88	0.25	0.85	0.17
	CIMAT_Team	0.69	0.26	0.89	0.41	0.93	0.28	-	-	-	-	0.82	0.22	0.82	0.22
	médianes des 5 suivants	0.60	0.38	0.82	0.77	0.87	0.46	0.71	0.88	0.87	0.40	0.86	0.41	0.74	0.46
EXP3	méthode proposée	0.79	0.21	0.86	0.98	0.91	0.18	0.68	1.13	0.89	0.65	0.82	0.41	0.82	0.30
	MorphoSeg [5]	0.77	0.21	0.86	0.96	0.89	0.24	0.72	0.95	0.91	0.55	0.78	0.39	0.78	0.39
	UPF_SIMBioSys [4]	0.73	0.27	0.89	0.52	0.87	0.30	0.76	0.53	0.91	0.59	0.85	0.34	0.72	0.55
	médianes des 5 suivants	0.72	0.28	0.87	0.87	0.85	0.35	0.73	0.84	0.91	0.62	0.81	0.48	0.71	0.59

TABLE 2 – Évaluation quantitative de la méthode présentée sur les données NEOBRAIN12 [20].

de référence, tout en ayant réduit considérablement le temps d'exécution de la segmentation. Dans l'avenir, nous souhaitons tester la méthode proposée sur des images IRM de cerveaux d'adultes afin de vérifier sa polyvalence. Une autre perspective réside dans d'autres façons de former des images en couleurs à partir de volumes 3D, avec éventuellement la prise en compte de données multi-modales.

Recherche reproductible. Toutes les informations permettant de reproduire nos résultats sont accessibles à partir d'une page dédiée [21], ainsi que des images de résultats supplémentaires et des tableaux comparatifs.

Références

- [1] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.
- [2] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *CVPR*, 2015, pp. 3431–3440.
- [3] M. J. Cardoso *et al.*, "AdaPT : an adaptive preterm segmentation algorithm for neonatal brain MRI," *NeuroImage*, vol. 65, pp. 97–108, 2013.
- [4] G. Sanroma *et al.*, "Building an ensemble of complementary segmentation methods by exploiting probabilistic estimates," in *Intl. Work. on Machine Learning in Medical Imaging*. Springer, 2016, pp. 27–35.
- [5] L. Gui *et al.*, "Morphology-driven automatic segmentation of MR images of the neonatal brain," *Medical Image Analysis*, vol. 16, no. 8, pp. 1565–1579, 2012.
- [6] P. Moeskops *et al.*, "Evaluation of an automatic brain segmentation method developed for neonates on adult MR brain images," in *SPIE Med. Ima.*, 2015, p. 941315.
- [7] L. Wang *et al.*, "LINKS : Learning-based multi-source IntegratioN framework for segmentation of infant brain images," *NeuroImage*, vol. 108, pp. 160–172, 2015.
- [8] W. Zhang *et al.*, "Deep convolutional neural networks for multi-modality iso-intense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.
- [9] M. Lai, "Deep learning for medical image segmentation," <https://arxiv.org/pdf/1505.02000>, 2015.
- [10] P. Moeskops *et al.*, "Automatic segmentation of MR brain images with a convolutional neural network," *IEEE Trans. on Med. Imaging*, vol. 35, no. 5, pp. 1252–1261, 2016.
- [11] H. Chen, Q. Dou, L. Yu, and P. Heng, "VoxResNet : Deep voxelwise residual networks for volumetric brain segmentation," <https://arxiv.org/abs/1608.05895>, 2016.
- [12] A. Prason *et al.*, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *MICCAI*, 2013, pp. 246–253.
- [13] K. Fritscher *et al.*, "Deep neural networks for fast segmentation of 3D medical images," in *MICCAI*, vol. 2, 2016, pp. 158–165.
- [14] C. N. Devi *et al.*, "Neonatal brain MRI segmentation : A review," *Computers in Biology and Medicine*, vol. 64, pp. 163–178, 2015.
- [15] I. Despotović, B. Goossens, and W. Philips, "MRI segmentation of the human brain : Challenges, methods, and applications," *Computational and Mathematical Methods in Medicine*, vol. 2015, 2015.
- [16] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection : CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. on Medical Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [17] Y. Bar *et al.*, "Chest pathology detection using deep learning with non-medical training," in *ISBI*, 2015, pp. 294–297.
- [18] K.-K. Maninis *et al.*, "Deep retinal image understanding," in *MICCAI*, 2016, pp. 140–148.
- [19] H. R. Roth *et al.*, "A new 2.5D representation for lymph node detection using random sets of deep convolutional neural network observations," in *MICCAI*, 2014, pp. 520–527.
- [20] I. Išgum *et al.*, "Evaluation of automatic neonatal brain segmentation algorithms : the NeoBrainS12 challenge," *Med. Image Analysis*, vol. 20, no. 1, pp. 135–151, 2015.
- [21] Y. Xu, T. Géraud, and I. Bloch, "Extra materials for this paper." <http://publications.lrde.epita.fr/xu.17.gretsi>, 2017.