

Impact du contraste des images scanner et IRM FLAIR dans les réseaux profonds pour la segmentation de la lésion d'AVC

Juliette MOREAU¹, David ROUSSEAU², Laura MECHTOUFF^{3,4}, Tae-Hee CHO^{3,4}, Carole FRINDEL¹

¹CREATIS, Université Lyon1, CNRS UMR5220, INSERM U1206, INSA-Lyon, 69621 Villeurbanne, France

²LARIS, UMR IRHS INRAe, Université d'Angers, 49100 Angers, France

³Département de neurologie, Hospices Civils de Lyon, 69500 Bron, France

⁴CarMeN, INSERM U1060, INRAe U1397, Université Lyon 1, INSA de Lyon, 69621 Villeurbanne, France

juliette.moreau@creatis.insa-lyon.fr

Résumé – L'imagerie médicale est centrale dans la prise en charge des patient-es admis-es pour un AVC. L'IRM cérébrale est faite en urgence pourvu qu'elle ne retarde pas le traitement. Si l'IRM n'est pas disponible ou en cas de contre-indication, un scanner cérébral est réalisé. Le but de ces travaux est de déterminer l'impact du contraste de ces images lors d'une segmentation de lésion d'AVC par des réseaux profonds. Le réseau de référence U-Net est comparé à deux autres architectures : le cGAN qui tente de renforcer les résultats du U-Net avec un réseau discriminatoire et le Mask R-CNN qui utilise des boîtes englobantes. Ce dernier segmente plus précisément sur les scanners, alors que le U-net offre un meilleur Dice pour les images IRM.

Abstract – Brain imaging plays a central role in the management of stroke patients. Cerebral MRI is the key imaging carried out in emergency in those patients provided that it does not delay the therapy. If MRI is not available or if contraindication, CT imaging will be performed. The objective of this work is to determine the impact of the contrast of these images for an accurate segmentation of the stroke lesion by deep networks. The U-Net reference network was compared to two others architectures: the cGAN which tries to reinforce the U-Net results with a discriminatory network and the Mask R-CNN which uses bounding boxes. The latter is more precise on scans, while U-Net offers a better Dice with MRI images.

1 Introduction

L'accident vasculaire cérébral ischémique (AVC) est une cause majeure de handicap acquis, de démence et de mortalité en France. Les traitements de reperfusion sont recommandés pour traiter ces patient-es; l'imagerie cérébrale est centrale dans leur prise en charge [1]. L'imagerie par résonance magnétique (IRM) et le scanner sont deux modalités utilisées. Elles sont nécessaires pour poser un diagnostic et préciser la localisation et le volume de la lésion. L'IRM permet d'obtenir des images avec un bon contraste. Disposer d'un outil de segmentation robuste sur les images serait utile pour les clinicien-nes lors de la prise en charge des patient-es pour évaluer leur pronostic et la réponse au traitement.

Dans la littérature, la majorité des travaux de segmentation des lésions d'AVC ont été réalisés à partir d'images IRM. Cependant, en pratique clinique le scanner est souvent utilisé car il est plus accessible, mais ces images étant moins contrastées la segmentation est plus difficile. Une référence dans le domaine est l'Ischemic Stroke Lesion Segmentation Challenge (ISLES) dont les trois premières éditions étaient basées uniquement sur des images IRM avec différentes tâches de segmentation ou de prédiction. Des données de perfusion scanner n'ont été fournies qu'en 2018 pour effectuer une certaine segmentation. Cependant, il s'agit d'une tâche différente de la nôtre car plusieurs cartes sont utilisées pour segmenter la lésion. La segmentation sur l'imagerie scanner reste un défi, nous avons exploré quelques solutions potentielles pour segmenter les lésions sur ces images particulièrement difficiles et comparer leurs performances à celles obtenues sur la base d'images IRM FLAIR.

2 Méthodologie

2.1 Données et pré-traitements

Les patient-es sont issu-es de la cohorte HIBISCUS-STROKE [2] d'observation monocentrique recrutant des personnes présentant un AVC dû à une occlusion d'une artère intracrânienne proximale traitée par thrombectomie. Les critères d'inclusion sont : (1) patient-es présentant un AVC de la circulation antérieure lié à une occlusion intracrânienne proximale; (2) IRM de diffusion comme imagerie de base; (3) patient-es traité-es par thrombectomie; (4) imagerie scanner 24 heures après l'inclusion; (5) acquisition FLAIR 6 jours après l'inclusion. La vérité terrain a été produite à partir des images IRM FLAIR car la lésion d'AVC y est plus visible et donc facile à détecter. Ces images ont été segmentées manuellement par un expert (THC) aveugle aux données cliniques à l'aide d'une méthode semi-automatique¹. Cette segmentation sera la vérité terrain de nos modèles aussi bien en IRM FLAIR qu'en scanner.

Avant l'apprentissage, un pré-traitement est effectué. En premier l'extraction de la boîte crânienne est faite avec la librairie FSL² améliorée par Muschelli et al [3] pour les images scanners, et la méthode HD-BET³ pour les images IRM. Tous les volumes sont ensuite recalés dans le référentiel de l'IRM DWI via une méthode de recalage non linéaire avec ANTs⁴. Les volumes 3D sont séparés en coupes 2D afin d'augmenter le nombre d'échantillons.

1. 3D slicer, <https://www.slicer.org/>

2. <https://fsl.fmrib.ox.ac.uk>

3. <https://github.com/MIC-DKFZ/HD-BET>

4. <http://stnava.github.io/ANTs/>

Seules les coupes avec une lésion d’après la vérité terrain sont sélectionnées et une partie est retournée horizontalement afin de placer toutes les lésions sur le même hémisphère. Ces étapes se basent sur l’hypothèse que la localisation des lésions est grossièrement connue grâce à des examens préliminaires. Les images sont enfin redimensionnées en 192x192 pixels.

2.2 Modèles

L’architecture U-Net [4], un réseau multi-échelles devenu l’algorithme standard de segmentation ces dernières années, est utilisée comme méthode de référence. Des cartes de probabilité avec 2 classes (lésion et fond) sont obtenues en sortie, les lésions sont seuillées à 0.5 pour définir l’infarctus final.

L’idée est ensuite de comparer cette référence à d’autres méthodes. La piste de développement est d’apporter une modification au U-Net, voir s’il est possible de l’améliorer. Un mécanisme d’attention connu est l’ajout de portes d’attention au U-Net [5]. Nous avons cependant décidé de tester une méthode qui ne modifie pas l’architecture principale mais ajoute un réseau qui impacte la mise à jour des poids avec un conditional Generative Adversarial Network [6] (cGAN). En effet, cette méthode a déjà été considérée pour améliorer les résultats du U-Net, notamment dans le cadre de petits jeux de données [7]. Il s’agit d’un GAN particulier conditionné par un échantillon de référence, il ne peut être utilisé que lorsque des images appariées sont disponibles. Les paires utilisées sont les images natives et leurs masques de segmentation manuelle correspondant. Le générateur, un U-Net, est amélioré grâce au discriminateur qui différencie le masque manuel de la sortie du réseau impactant la mise à jour des poids du générateur.

Pour la dernière architecture, nous avons utilisé un mécanisme d’attention, basé sur la détection d’objets. Il y a deux architectures principales qui font une segmentation à partir de boîtes englobantes : le Mask R-CNN [8] et le poly-YOLO [9]. Le premier a été choisi car les segmentations en sortie sont plus précises et non pas sous forme de polygones. L’algorithme en deux étapes extrait des cartes de caractéristiques à l’aide d’un réseau convolutif afin d’obtenir des régions d’intérêt (ROI). Ces dernières sont alors utilisées pour finir la segmentation.

3 Expériences et résultats

3.1 Sélection des patient-es à l’aide du contraste

Parmi les 108 patient-es de la cohorte HIBISCUS-STROKE, une sélection basée sur le volume de la lésion et son contraste est faite. Le contraste est estimé avec le ratio de Fisher $F = \frac{(\mu_c - \mu_l)^2}{\sigma_c^2 + \sigma_l^2}$ dont la valeur reflète la différence entre la lésion et le reste du tissu cérébral en fonction des valeurs des intensités dans chaque zone et en tenant compte de l’homogénéité de chacune ; la distribution gaussienne pour les images IRM et scanner ayant déjà été démontrée [10]. Les lésions présentant un volume inférieur à 10 mL ainsi que celles avec un volume inférieur à 25 ml et un ratio de Fisher inférieur à 1 sont éliminées, car elles sont jugées comme trop difficilement visibles à l’oeil nu. Au total, 58 patient-es sont

incluses dans l’étude avec un volume lésionnel de 10 mL à 190 mL.

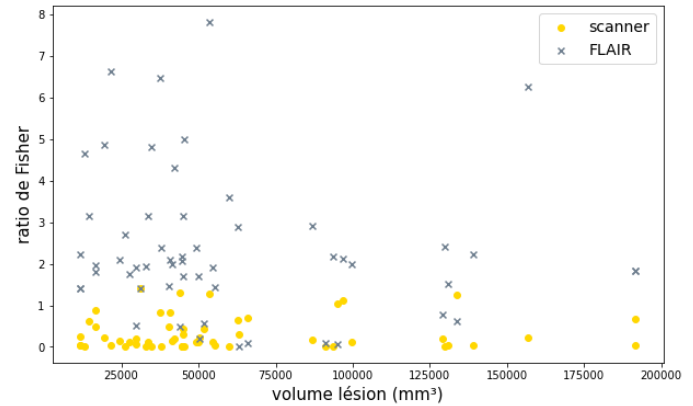


FIGURE 1 – Contraste de la lésion sur IRM FLAIR et CT normalisées (taux de variation : 2.08 en scanner, 0.80 en IRM FLAIR)

La figure 1 met en avant la différence de contraste de la lésion par rapport au tissu cérébral sain, chaque point représentant un patient. Les images ont été normalisées avant le calcul car les ordres de grandeur des valeurs des pixels sont très différents entre IRM en scanner. On observe distinctement que le contraste des IRM FLAIR est meilleur que celui des scanners, rendant la lésion plus visible et donc la segmentation plus facile. En figure 2 on observe que cette différence est surtout due à des moyennes de niveau de gris différentes pour un écart-type équivalent. En plus de cela, le taux de variation (ratio entre écart-type et moyenne) de l’indice de Fisher démontre une plus grande hétérogénéité du contraste en scanner qu’en IRM FLAIR.

3.2 Implémentation

L’architecture U-Net utilisée est celle issue de la publication originale [4]. La fonction de coût est une combinaison d’entropie croisée et de Dice entre deux classes : lésion et fond. Le réseau compte 17 millions de paramètres à entraîner. Le générateur du cGAN est le U-Net précédent avec les mêmes fonctions de coût et le discriminateur est un réseau à 5 couches, les 4 premières sont des couches convolutives 2D avec une fonction d’activation LeakyReLU [11] et la dernière est une couche convolutive 2D, la fonction de coût est une entropie croisée binaire combinée à une sigmoïde. L’ajout du second réseau ajoute 1.5 million de paramètres aux 17 du U-Net. Pour le U-Net et le cGAN, le taux d’apprentissage utilisé est de 0.0002. Nous avons utilisé le détecteur [12] pour implémenter un Mask R-CNN. Un apprentissage par transfert à partir d’ImageNet est effectué sur un ResNet50 pour l’extraction de caractéristiques. Le taux d’apprentissage est de 0.001.

Une validation croisée 5-folds a été effectuée ; les 58 patient-es sélectionnés sont séparés en différents jeux d’apprentissage (n=46) et de test (n=12). Pour mettre en évidence la difficulté du scanner par rapport à l’IRM, les modèles sont entraînés et testés avec les mêmes paramètres sur les données IRM FLAIR pré-traitées et séparées de la même façon. L’IRM étant moins bien résolue en

TABLE 1 – Résultats sur le jeu de test. Test statistique par modalité entre les modèles avec un test non-apparié unilatéral de Mann-Whitney, marque * si le résultat d’un modèle est meilleur que les deux autres.

	scanner			IRM FLAIR		
	U-Net	cGAN	Mask R-CNN	U-Net	cGAN	Mask R-CNN
Détection (%)	88 ± 5	93 ± 3 *	79 ± 3	96 ± 2.7	98 ± 0.3	91.9 ± 0.6
DSC	0.36 ± 0.04	0.35 ± 0.03	0.42 ± 0.04 *	0.68 ± 0.03	0.64 ± 0.03	0.62 ± 0.05
HD (pixel)	31 ± 2.0	31 ± 2.5	26 ± 1.6 *	21 ± 3	26 ± 1.7	21 ± 2.2

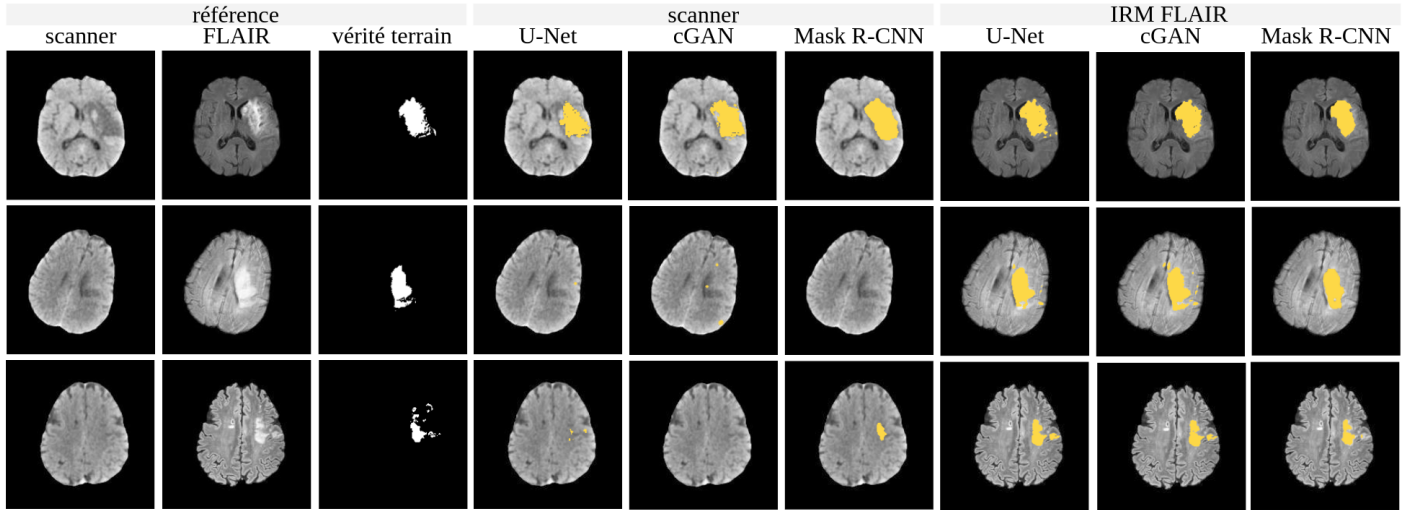


FIGURE 2 – Visualisation des segmentations proposées par les trois méthodes. De gauche à droite : scanner, IRM FLAIR, vérité terrain, segmentation sur scanner U-Net, cGAN, Mask R-CNN, segmentation sur IRM FLAIR U-Net, cGAN, Mask R-CNN. Ratio de Fisher de haut en bas pour scanner : 0.21, 0.24, 0.07 et pour IRM FLAIR : 2.85, 2.64, 3.45.

nombre de coupes que le scanner, il y a environ 3000 images en entraînement et 730 en test pour le scanner contre 580 et 170 pour l’IRM. A chaque fois, l’entraînement a lieu sur 200 époques avec 12 images par lot. L’entraînement prend environ 2h30 pour le U-Net, 3h pour le cGAN et presque 11h pour le Mask R-CNN (e.g. pour les données scanner). Une partie de cet important temps de calcul s’explique par la méthode utilisée, les ROI calculées par image sont ensuite considérées indépendamment, augmentant considérablement le nombre d’images à traiter.

L’évaluation et la comparaison quantitative entre les méthodes est faite grâce au taux de détection qui rend compte du nombre de coupes sur lesquelles la lésion est détectée et à deux métriques qui vont juger le contour de la segmentation le score de Dice (DSC) et la distance de Hausdorff (HD) qui sera plus sévère.

3.3 Résultats

Le U-Net et le GAN donnent en sortie une segmentation par coupe avec deux classes tandis que le Mask R-CNN propose parfois plusieurs segmentations par coupe; en effet cette méthode a initialement été développée pour de la segmentation multi-instance. Un post-traitement sélectionne alors le masque avec le meilleur Dice. Les résultats sont calculés pour chaque coupe du jeu de test; seules celles avec prédiction sont utilisées, hors taux de détection. Les résultats sont résumés dans le tableau 1.

D’un point de vue global, les performances sont meilleures avec les images IRM que les images scanners pour toutes les métriques.

Concernant les images scanner, la détection est particulièrement basse pour le Mask R-CNN. L’évolution du U-Net en cGAN permet d’augmenter le taux de détection, autrement les résultats sont similaires sur le scanner et un peu moins bon pour l’IRM mais de façon non significative. Le mauvais contraste du scanner ne permet pas d’obtenir des valeurs de Dice aussi hautes que sur IRM FLAIR. Ces plus faibles valeurs peuvent en partie être justifiées par la vérité terrain choisie : les images FLAIR -sur lesquelles la lésion a été détournée- ont été acquises 5 jours après les scanners, bien que la lésion soit considérée stabilisée 24 heures après la thrombectomie, une évolution de la lésion entre les deux ne peut pas être exclue. La distance d’Hausdorff est elle aussi meilleure en IRM qu’en scanner, à chaque fois le Mask R-CNN offre les plus faibles valeurs promettant des segmentations assez stables avec moins de points aberrants éloignés de la lésion. Un point retient notre attention : les deux métriques qui estiment la précision de la lésion ne s’accordent pas sur l’IRM, le Dice pointe le U-Net comme meilleur et la distance d’Hausdorff met plutôt en avant le Mask R-CNN. Cela pourrait venir de la complexité du Mask R-CNN qui procède à une segmentation en deux étapes avec plusieurs propositions de masque en sortie; les différentes composantes d’une lésion peuvent être prédites dans des masques distincts et le post-traitement élimine certaines les plus éloignées alors que les deux autres algorithmes ne proposent qu’un seul masque dès le début. Ce phénomène intervient surtout en IRM, du fait d’un contraste accru, où les plus petits détails de la lésion même excentrés y sont distinguables : 21% des coupes IRM ont

plus d'une segmentation parmi celle où une lésion est détectée contre 18% en scanner (test de Mann-Whitney unilatéral $p=0.013$). Cette complexité se voit en figure 2 où les masques proposés par le Mask R-CNN sont plus compacts et moins dentelés. Sur cette même figure, les deux premières lignes montrent des coupes où la lésion est plutôt bien contrastée en scanner. La première coupe est bien segmentée avec un Dice légèrement meilleur pour le Mask R-CNN (0.87 contre 0.78 pour le U-Net et 0.71 pour le cGAN). Mais sur la deuxième ligne aucune prédiction n'est faite pour le Mask R-CNN alors que les deux autres algorithmes font des toutes petites prédictions de quelques pixels, elles sont alors très mauvaises et vont fortement réduire les mesures de qualité de segmentation. Il y a un compromis à trouver entre la détection des lésions et leur précision qui explique en partie les meilleures performances du Mask R-CNN. Ces deux coupes ne posent pas problème pour leur segmentation en IRM. Sur la dernière ligne la lésion est très peu visible sur scanner : l'attention sous forme de boîtes englobantes du Mask R-CNN permet une segmentation que le U-Net et le cGAN ne savent pas faire alors que sur IRM FLAIR aucun modèle n'a de difficulté à segmenter.

La comparaison entre les résultats obtenus sur les images scanner et IRM FLAIR met en évidence la difficulté de la tâche de segmentation en scanner. Il est très intéressant de noter que le Mask R-CNN est moins performant que les autres sur les images IRM FLAIR pour le Dice ce qui est l'inverse avec la modalité scanner. Ce point révèle l'importance du mécanisme en deux étapes avec boîtes englobantes qui s'avère utile lorsque le contraste est faible comme en imagerie scanner, mais s'avère moins efficace lorsqu'il n'y a pas cette difficulté de contraste. Les coupes présentent sur la figure 2 mettent en avant la différence de contraste entre les deux modalités avec des valeurs de ratio de Fisher très différentes.

Chacun des trois modèles entraînés sur données IRM ont été testés sur 35 patients du challenge ISLES 2022 pour lequel on dispose d'IRM FLAIR et de masques de segmentations associés. Les résultats obtenus sont un peu moins bon que sur notre jeu de données mais dans le même ordre de grandeur : détection supérieure à 92%, Dice entre 0.53 et 0.59 et distance d'Hausdorff entre 32.3 et 20.3.

4 Conclusion

La différence de contraste entre les modalités scanner et IRM a un impact sur les résultats des modèles testés dans le cadre de ce travail. D'une part ils sont moins bons pour les scanners, qui ont un contraste réduit, d'autre part ce ne sont pas les mêmes modèles qui permettent les meilleures segmentations. Parmi les architectures testées pour le scanner, un compromis est nécessaire entre la détection des lésions et la précision de la segmentation obtenue. Même si le Mask R-CNN a un taux de détection plus faible, il fournit des résultats prometteurs pour l'imagerie scanner.

Afin de préciser l'impact du contraste, il serait pertinent de modifier artificiellement ce dernier en atténuant le signal dans la lésion d'après le masque à l'aide d'un coefficient atténuateur et appliquant un noyau de convolution pour adoucir les bords. Une autre idée serait de poursuivre dans la voie du Mask R-CNN avec

un modèle faiblement supervisé ne fournissant que des boîtes englobantes au réseau pour l'entraînement. Comme cette vérité terrain est plus facile et rapide à créer, cela permettrait d'avoir un plus grand nombre d'images en entraînement pour avoir un algorithme de segmentation robuste sur des types de lésion variés. Cela résoudrait un biais de ce travail : la vérité terrain est réalisée sur le FLAIR favorisant cette modalité, de plus 5 jours séparent les deux acquisitions et la lésion a pu évoluer même si elle est stabilisée au bout de 24h.

5 Remerciements

Ce travail a été soutenu par la RHU MARVELOUS (ANR-16-RHUS-0009) de l'Université Claude Bernard Lyon-1, et par la RHU BOOSTER (ANR-18-RHUS-0001), au sein du programme "Investissements d'Avenir" de l'Agence Nationale de la Recherche.

Références

- [1] INSERM accident vasculaire cérébral (avc). <https://www.inserm.fr/information-en-sante/dossiers-information/accident-vasculaire-cerebral-avc>. Accessed : 2021-04-15.
- [2] N. Debs et al. Impact of the reperfusion status for predicting the final stroke infarct using deep learning. *NeuroImage : Clinical*, 29, 2021.
- [3] J. Muschelli et al. Validated automatic brain extraction of head ct images. *Neuroimage*, 114, 2015.
- [4] O. Ronneberger et al. U-net : Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015.
- [5] O. Oktay et al. Attention u-net : Learning where to look for the pancreas. *arXiv preprint arXiv :1804.03999*, 2018.
- [6] M. Mirza and S. Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv :1411.1784*, 2014.
- [7] G. Heilemann et al. Can generative adversarial networks help to overcome the limited data problem in segmentation? *Zeitschrift für Medizinische Physik*, 2021.
- [8] K. He et al. Mask r-cnn. In *Proceedings of the IEEE international conference on computer vision*, 2017.
- [9] P. Hurtik et al. Poly-yolo : higher speed, more precise detection and instance segmentation for yolov3. *arXiv preprint arXiv :2005.13243*, 2020.
- [10] P. Gravel et al. A method for modeling noise in medical images. *IEEE Transactions on medical imaging*, 23(10), 2004.
- [11] A. Maas et al. Rectifier nonlinearities improve neural network acoustic models. In *Proc. icml*, volume 30. Citeseer, 2013.
- [12] R. Girshick et al. Detectron. <https://github.com/facebookresearch/detectron>, 2018.