

Utilisation d’informations géométriques pour l’amélioration de la segmentation automatique d’organes allongés

Rebeca VÉTIL^{1,2}, Alexandre BÔNE¹, Marie-Pierre VULLIERME³, Marc-Michel ROHÉ¹, Pietro GORI², Isabelle BLOCH^{4,2}

¹Guerbet Research, Villepinte, France

²LTCI, Télécom Paris, Institut Polytechnique de Paris, France

³Département de Radiologie, Hôpital d’Annecy-Genevois, Université Paris Cité, France

⁴Sorbonne Université, CNRS, LIP6, Paris, France

(rebeca.vetil, alexandre.bone, marc-michel.rohe)@guerbet.com,
(pietro.gori, isabelle.bloch)@telecom-paris.fr, mpvullierme@gmail.com

Résumé – Les réseaux de neurones profonds sont largement utilisés pour la segmentation automatique d’organes dans des images médicales. Cependant, certains organes sont plus difficiles à délimiter que d’autres, en raison par exemple du faible contraste à leurs extrémités. Dans cet article, nous proposons d’améliorer la segmentation des organes allongés grâce à des informations géométriques. Nous introduisons ces dernières pendant l’apprentissage ou lors du post-traitement. Les deux stratégies n’introduisent pas de paramètres d’apprentissage supplémentaires et peuvent être facilement appliquées à tout modèle existant. La méthode proposée est évaluée sur le problème de segmentation du pancréas. Les résultats montrent que la sous-segmentation systématique des extrémités est bien corrigée, tout en préservant la qualité globale de la segmentation.

Abstract – Deep neural networks are widely used for automated organ segmentation in medical images. However, some organs are more challenging to delineate than others, for instance due to low contrast at their extremities. In this paper, we propose to improve the segmentation of elongated organs thanks to Geometrical Priors that can be introduced during training or at post-processings. Both strategies do not introduce additional training parameters and can be easily applied to any existing network. The proposed method is evaluated on the challenging problem of pancreas segmentation. Results show that Geometrical Priors allow us to correct the systematic under-segmentation pattern of a state-of-the-art method, while preserving the overall segmentation quality.

1 Introduction

La segmentation automatique d’organes est une étape méthodologique essentielle au développement d’outils d’aide au diagnostic par ordinateur. Cependant, certains organes sont plus difficiles à segmenter que d’autres, en raison par exemple de contours ambigus ou de formes allongées. Plusieurs méthodes de segmentation automatique ont été proposées, et l’utilisation de réseaux de neurones convolutionnels profonds est progressivement reconnue comme la méthode de référence puisqu’ils obtiennent les meilleurs résultats dans beaucoup d’applications [4] en termes de coefficient de similarité de Dice (DSC). Aujourd’hui, la plupart des réseaux proposés s’inspirent du réseau U-Net [5]. Récemment, dans [3], les auteurs ont introduit le *no-new U-Net*, surnommé nnU-Net, où l’accent est mis sur les étapes de prétraitement, d’apprentissage et de post-traitement plutôt que sur l’architecture du réseau. Le nnU-Net a remporté de nombreuses compétitions de segmentation d’images médicales, dont le Medical Decathlon¹.

Si certains algorithmes de segmentation automatique atteignent des performances comparables à celles des radiologues [1], la segmentation de certains organes, tels que le pancréas ou le

côlon, demeure un problème ouvert puisque les performances sont inférieures à celles obtenues sur d’autres organes abdominaux [10]. Ces organes difficiles présentent une forme allongée ainsi qu’un faible contraste à leurs extrémités : ces caractéristiques rendent leur délimitation difficile et les algorithmes manquent les extrémités. D’un point de vue clinique, cette erreur peut être critique si la tâche de segmentation automatique fait partie d’un outil plus complexe de détection de tumeurs, car des masses aux extrémités pourraient être manquées. Pour résoudre ce problème, une approche possible consiste à effectuer une segmentation multi-organes, en délimitant toutes les structures majeures proches de l’organe cible [2, 9]. Par exemple la méthode de [11] propose une fonction de coût qui intègre des informations anatomiques calculées sur des données non annotées. Cependant, la méthode dépend fortement de la quantité de données entièrement annotées, et les informations ne tiennent pas compte de la géométrie des organes. De telles observations nous ont motivés à rechercher d’autres informations anatomiques, géométriques, et dont les définitions ne dépendraient pas de données annotées.

Nous cherchons ici à améliorer la segmentation automatique d’organes allongés. Nous illustrons nos contributions sur le problème de la segmentation du pancréas. En plus d’avoir une forme allongée, le pancréas présente une forte ambiguïté à ses

1. <http://medicaldecathlon.com/results/>

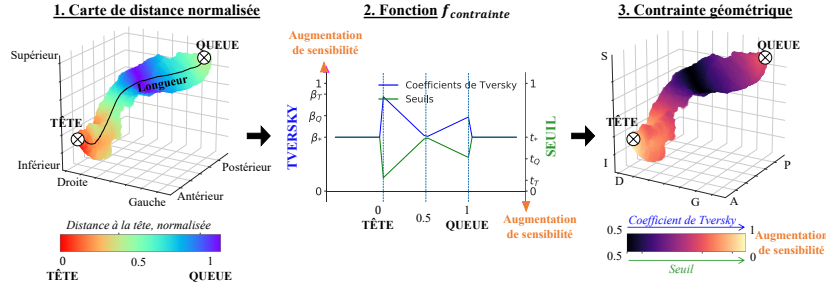


FIGURE 1 – **Construction d’une contrainte géométrique.** À partir d’un masque de segmentation binaire, une carte de distance normalisée est calculée, puis transformée en contrainte géométrique.

extrémités puisque ces dernières sont fortement intriquées avec le duodénum et la rate. De telles caractéristiques rendent la segmentation du pancréas difficile, même pour des radiologues expérimentés [4]. Nous évaluons nos méthodes sur deux bases de données publiques, *The Cancer Imaging Archive Pancreas-CT* (TCIA) [6] et *Medical Segmentation Decathlon* (MSD) [8]. Nous évaluons les performances du modèle état de l’art, le nnU-Net, pour la segmentation du pancréas et vérifions qu’il sous-estime significativement ses extrémités. Pour corriger ce problème, nous proposons deux stratégies qui contrôlent la sensibilité des prédictions en utilisant des informations géométriques, qui peuvent être introduites soit pendant l’apprentissage, en utilisant une fonction de coût de Tversky [7] locale, soit pendant le post-traitement, en utilisant des seuils locaux.

2 Méthode

Soit un organe de forme allongée O_{el} , dont les extrémités e_0, e_1 sont difficiles à délimiter. Soit $\mathcal{R} = \{r_i \in \{0, 1\}, i = 1 \dots N\}$ sa segmentation de référence, avec N le nombre de voxels de l’image 3D. Soit $\mathcal{P} = \{p_i \in [0, 1], i = 1 \dots N\}$ la carte de probabilité générée par un réseau de segmentation. Nous définissons la contrainte géométrique \mathcal{G}_{prior} qui donne pour chaque voxel i l’augmentation de sensibilité désirée ρ_i : $\mathcal{G}_{prior} = \{\rho_i \in [0, 1], i = 1 \dots N\}$. Pour l’exemple de la segmentation du pancréas, les extrémités e_0 et e_1 seront désignées par tête (T) et queue (Q) dans la suite.

Contraintes géométriques. Les contraintes géométriques attribuent à chaque voxel d’un masque de segmentation une augmentation de sensibilité. Comme l’illustre la figure 1, la construction d’une contrainte géométrique \mathcal{G}_{prior} suit trois étapes. Tout d’abord, une carte de distance euclidienne attribue à chaque voxel sa distance par rapport à la tête du pancréas. La carte de distance est normalisée de manière à ce que les valeurs de distance soient 0 et 1 à la tête et à la queue, respectivement. Ensuite, une fonction $f_{contrainte}$ attribue à chaque voxel i une augmentation de sensibilité ρ_i en fonction de sa distance à la tête. La fonction $f_{contrainte}$ est construite de sorte à donner aux régions ordinairement sous-segmentées un surcroît de sensibilité, favorisant ainsi leur segmentation. Pour le pancréas, $f_{contrainte}$ est une fonction linéaire par morceaux qui met l’ac-

cent sur les extrémités. Elle est définie par trois paramètres : ρ_T, ρ_Q, ρ_* , correspondant à l’augmentation de la sensibilité dans la tête, la queue et le reste de l’objet, respectivement. Enfin, la contrainte géométrique résultante peut être utilisée pour définir une fonction de coût de Tversky locale ou des seuils locaux.

Fonction de coût de Tversky locale. La première façon d’utiliser les contraintes géométriques est pendant l’apprentissage, au moyen d’une version locale de la fonction de coût de Tversky définie dans [7] :

$$T(\mathcal{P}, \mathcal{R}, \mathcal{G}_{prior}^{Tversky}) = \frac{\sum_{i=1}^N p_i^0 r_i^0}{\sum_{i=1}^N p_i^0 r_i^0 + \sum_{i=1}^N \alpha_i p_i^0 r_i^1 + \sum_{i=1}^N \beta_i p_i^1 r_i^0} \quad (1)$$

où p_i^0 (respectivement p_i^1) est la probabilité prédite d’appartenance du voxel i à O_{el} (respectivement, à l’arrière-plan). Les mêmes notations s’appliquent à $r_i \in \mathcal{R}$. Les paramètres (α_i, β_i) contrôlent le compromis entre la spécificité et la sensibilité pour le voxel i . En particulier, des valeurs plus élevées de β_i pénalisent davantage les erreurs de sous-segmentation. Alors que cette fonction de coût de Tversky est utilisée dans [7] avec des coefficients globaux (α, β) , notre méthode utilise des coefficients locaux, définis par la contrainte géométrique : $\mathcal{G}_{prior}^{Tversky} = \{(\alpha_i, \beta_i) \in \mathbb{R}^2, i = 1 \dots N\}$. En pratique, $\alpha_i = 1 - \beta_i$, et $\beta_i \in]0, 1[$. Les valeurs de $\mathcal{G}_{prior}^{Tversky}$ sont calculées pour chaque patient une fois, avant l’apprentissage, en utilisant leur masque de segmentation obtenu manuellement.

Seuils locaux. La seconde façon d’utiliser les contraintes géométriques est en post-traitement, via la définition de seuils locaux t_i , tels que $\mathcal{G}_{prior}^{threshold} = \{t_i \in]0, 1[, i = 1 \dots N\}$. Les seuils t_i sont appliqués sur la sortie de probabilité \mathcal{P} , et le voxel i appartient au pancréas si $p_i \geq t_i$. Les zones dont la sensibilité est élevée auront des seuils plus bas. Les informations a priori utilisées pour définir $\mathcal{G}_{prior}^{seuil}$ sont calculées à partir de la segmentation prédite obtenue avec un seuil global $t = 0, 5$ où tête et queue sont différenciées avec l’orientation des images lors de l’acquisition. Le résultat final est désigné par P_{bin} .

Par souci de simplicité, l’augmentation de sensibilité ρ sera utilisée pour faire référence aux coefficients de Tversky locaux, avec $\rho_i = \beta_i$, ou aux seuils locaux, avec $\rho_i = 1 - t_i$, selon le contexte.

3 Expériences

Données. Nous utilisons deux bases de données publiques : la base de données TCIA [6], comportant 80 sujets, et l’ensemble d’apprentissage de la base de données MSD pancréas [8], comportant 281 sujets. Les deux ensembles de données contiennent des images de tomodynamométrie en phase portale, ainsi que les masques de segmentation du pancréas réalisés manuellement par des radiologues. Les données sont réparties en trois sous-ensembles, d’apprentissage, validation et test, avec un rapport TCIA/MSD de 75 :25 dans tous les sous-ensembles.

Méthode de référence. Le prétraitement, l’architecture du réseau, l’optimisation et le post-traitement sont automatiquement sélectionnés à l’aide de la procédure de configuration automatique du réseau nnU-Net [3]. Tous les paramètres par défaut sont conservés dans une expérience de référence.

Paramètres. Nous réalisons des expériences avec différentes combinaisons d’augmentations de sensibilité (ρ_T, ρ_Q, ρ_*), où chaque combinaison représente une fonction $f_{contrainte}$ spécifique. Les valeurs de ρ_x varient dans $[0, 5, 0, 7, 0, 95]$, désignées par $[\rho_x^5, \rho_x^7, \rho_x^{95}]$. En outre, les méthodes proposées sont comparées à la fonction de coût de Tversky originale [7], avec des coefficients globaux.

Mesures de performance. La qualité de la segmentation aux extrémités est quantitativement évaluée avec deux mesures. La première est l’erreur sur la longueur de l’organe, exprimée en mm. Pour calculer la longueur, le masque de segmentation est converti en un graphe d’adjacence de voxels. L’erreur de longueur est ensuite définie comme suit : $Err Long = Long(\mathcal{P}_{bin}) - Long(\mathcal{R})$, où \mathcal{P}_{bin} et \mathcal{R} désignent respectivement les masques de segmentation prédits et manuels. Dans l’exemple du pancréas, où les extrémités ont tendance à être manquées, nous pouvons estimer que les erreurs de longueur négatives correspondent à une sous-segmentation, et les erreurs positives à une sur-segmentation. La deuxième mesure est le DSC aux extrémités, défini comme la moyenne des DSC calculés sur 5% de la longueur du pancréas à partir de chaque extrémité. Enfin, afin de s’assurer que la qualité globale de la segmentation n’est pas détériorée, la différence du DSC global par rapport à la méthode de référence est également rapportée.

4 Résultats

Au total, 10 expériences, en plus de la méthode de référence, sont réalisées. Dans une première série d’expériences A-E, la fonction de coût de référence est remplacée par une fonction de coût de Tversky locale. Dans une deuxième série d’expériences A’-E’, le post-traitement de la méthode de référence est remplacé par une approche de seuillage local, tandis que la fonction de coût de référence est conservée. Les détails des résultats quantitatifs sont présentés dans la table 1.

Résultats de la méthode de référence. Les résultats de la méthode de référence révèlent une sous-segmentation importante, caractérisée par une erreur de longueur médiane négative de

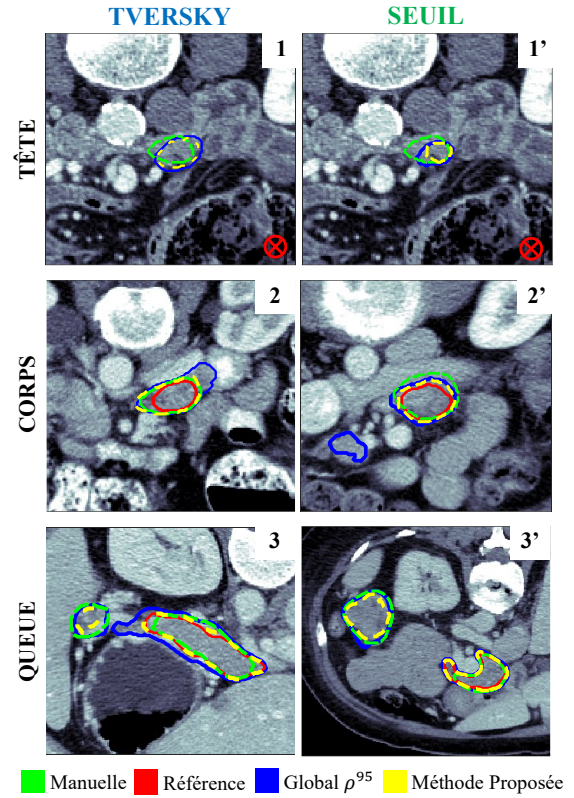


FIGURE 2 – Résultats de segmentations obtenus à la tête, le corps et la queue du pancréas. La première colonne montre les expériences avec la fonction de coût de Tversky locale, la deuxième celles avec les seuils locaux. Les contours bleus correspondent à A et A’. Les contours jaunes correspondent aux meilleurs résultats de la table 1, obtenus par D et E’. ⊗ indique une prédiction nulle pour la méthode de référence.

−3.2mm. Toutes les expériences A-E et A’-E’ parviennent à résoudre ce problème puisque l’erreur médiane est déplacée de manière significative vers des valeurs positives.

Coût de Tversky local. Les expériences A-E, utilisant une fonction de coût de Tversky globale ou locale, produisent des effets significatifs sur l’erreur de longueur. Cependant, (A) montre qu’un paramètre global β^{95} diminue le DSC global, indiquant une détérioration de la segmentation globale. En outre, (A) montre également une baisse du DSC aux extrémités. D’autre part, l’utilisation de β^7 globale, comme dans (B), n’affecte pas la qualité de la segmentation globale mais ne produit qu’une légère amélioration de l’erreur de longueur. Les deux cas (A, B) illustrent les limites de la fonction de coût de Tversky globale, qui empêchent son utilisation avec de forts coefficients. Ces inconvénients sont atténués par les contraintes géométriques : (C-E) n’entraînent pas de baisse significative du DSC global, mais corrigent fortement l’erreur de longueur. Cette capacité d’atténuation est illustrée plus en détail dans la figure 2 : aux extrémités, (D) (jaune) se comporte comme (A), entraîné avec un fort coefficient Tversky global (bleu), et surpasse la méthode de référence (rouge). En revanche, près du corps du pancréas,

TABLE 1 – Mesures d’évaluation obtenues sur les 65 cas de test. Nous rapportons la médiane et l’écart interquartile entre parenthèses. Le DSC médian pour la méthode de référence est de 88,0 (5,0). Les différences significatives (t-test au niveau $p = 0,05$) avec la méthode de référence sont indiquées par *, sauf pour l’erreur de longueur pour laquelle tous les résultats sont significatifs. Le meilleur résultat, en **gras**, est sélectionné parmi les expériences qui ne détériorent pas significativement le DSC par rapport à la méthode de référence.

A PRIORI		GLOBAL			LOCAL		
		ρ^{95}	ρ^7		$\rho_T^{95}, \rho_Q^{95}, \rho_*^5$	$\rho_T^{95}, \rho_Q^7, \rho_*^5$	$\rho_T^{95}, \rho_Q^7, \rho_*^7$
TVERSKY	Expérience	Référence	A	B	C	D	E
	Erreur de Longueur (mm)	-3,2 (6,4)	1,2 (7,8)	-2,0 (6,5)	2,3 (7,6)	-0,4 (8,2)	0,2 (6,8)
	DSC aux extrémités (%)	70,0 (3,4)	67,5 (2,1)	71,0 (3,4)	68,0 (1,7)	72,5 (2,8)*	70,5 (3,2)
	δDSC (%) p.r à la Référence	-	-4*	0,2	-1,3*	0,1	-2*
SEUIL	Expérience	Référence	A'	B'	C'	D'	E'
	Erreur de Longueur (mm)	-3,2 (6,4)	-0,3 (6,3)	-1,8 (7,2)	-1,4 (7,6)	-1,6 (7,3)	-1,0 (7,7)
	DSC aux extrémités (%)	70,0 (3,4)	72,0 (2,3)*	75,0 (2,95)*	74,0 (2,5)*	74,0 (2,7)*	74,0 (2,6)*
	δDSC (%) p.r à la Référence	-	-2*	0,3	0,1	-0,1	0,2

où l’augmentation de la sensibilité est plus faible, (D) imite la méthode de référence et évite les erreurs de sur-segmentation importantes causées par (A). En plus de réduire l’erreur de longueur, (D) augmente également de manière significative le DSC aux extrémités, ce qui suggère que les extrémités sont non seulement mieux détectées, mais aussi mieux délimitées.

Seuils locaux. Les expériences A'-E', utilisant des seuils locaux, apportent de plus faibles améliorations de l’erreur de longueur, la meilleure amélioration, dans (A'), étant obtenue au détriment du DSC global. Toutefois, la table 1 montre que le seuillage local produit une augmentation systématique et significative du DSC aux extrémités, ainsi que des améliorations qualitatives illustrées dans la figure 2. En particulier, les figures 2.1' et 2.3' montrent des échecs totaux de la méthode de référence, mais récupérés avec succès par (E').

5 Discussion et conclusion

Nous avons proposé deux méthodes pour accroître la sensibilité de la segmentation aux extrémités des organes allongés. Nos méthodes reposent sur le calcul de contraintes géométriques qui attribuent à chaque voxel une augmentation de la sensibilité. Ces contraintes géométriques peuvent être utilisées pendant l’apprentissage, via des coefficients de Tversky locaux, ou lors du post-traitement, via des seuils locaux. Les résultats quantitatifs sur le problème de la segmentation du pancréas ont démontré que les deux techniques proposées ont réussi à augmenter de manière significative la sensibilité de la segmentation aux extrémités de l’organe. Les points forts de notre travail sont triples : premièrement, les méthodes proposées sont interprétables grâce aux contraintes géométriques qui reflètent la complexité anatomique de l’organe. Ensuite, la formulation générale de ces contraintes géométriques permet d’appliquer nos méthodes à tout organe dont les régions sont difficiles à segmenter. Enfin, nos méthodes peuvent être facilement appliquées à tout réseau existant. En particulier, la stratégie des seuils locaux ne nécessite pas de ré-entraîner le modèle. La principale limite de notre travail est l’introduction d’hyper-paramètres qui, bien que guidés par les connaissances anatomiques, peuvent rendre la recherche fastidieuse. Ainsi, les

travaux futurs pourraient explorer l’apprentissage de fonctions $f_{contrainte}$, dont les paramètres seraient appris dynamiquement pendant l’apprentissage ou le post-traitement.

Références

- [1] W. Bai et al. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *J Cardiovasc Magn Reson*, 20(1) :65, 2018.
- [2] L. Chu et al. Application of deep learning to pancreatic cancer detection : lessons learned from our initial experience. *JACR*, 16(9), 2019.
- [3] F. Isensee et al. nnU-Net : a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods*, 18(2) :203–211, 2021.
- [4] H. Kumar, S.V DeSouza, and M.S Petrov. Automated pancreas segmentation from computed tomography and magnetic resonance images : a systematic review. *Comput Methods Programs Biomed*, 178 :319–328, 2019.
- [5] O. Ronneberger, P. Fischer, and T. Brox. U-Net : Convolutional networks for biomedical image segmentation. In *MICCAI*, 2015.
- [6] H. Roth et al. Data from pancreas-CT. The Cancer Imaging Archive, 2016.
- [7] S.S.M Salehi, D. Erdogmus, and A. Gholipour. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In *MICCAI*, 2017.
- [8] A.L. Simpson et al. A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv :1902.09063*, 2019.
- [9] Y. Wang et al. Abdominal multi-organ segmentation with organ-attention networks and statistical fusion. *MedIA*, 55 :88–102, 2019.
- [10] X. Yao, Y. Song, and Z. Liu. Advances on pancreas segmentation : a review. *Multimed. Tools. Appl.*, 79(9) :6799–6821, 2020.
- [11] Y. Zhou et al. Prior-aware neural network for partially-supervised multi-organ segmentation. In *CVPR*, 2019.