

Génération d'un CT synthétique à partir d'une séquence IRM avec un réseau antagoniste génératif en radiothérapie

Kevin N.D. Brou Boni^{1,4}, John Klein⁴, Ludovic Vanquin¹, David Pasquier^{2,4}, Nick Reynaert^{1,3}

¹Service de Physique Médicale, Centre Oscar Lambret, Lille, France

²Département de Radiothérapie, Centre Oscar Lambret, Lille, France

³Service de Physique Médicale, Institut Jules Bordet, Bruxelles, Belgique

⁴Univ. Lille, CNRS, Centrale Lille, UMR 9189 - CRIStAL, F-59000 Lille, France

k-brouboni@o-lambret.fr

Résumé – La planification de traitement uniquement basé sur l'IRM en radiothérapie nécessite une génération fiable d'un scanner (CT) synthétique à partir d'une IRM. Les méthodes de transcription d'un IRM en CT par apprentissage profond sur la base de paires alignées requièrent une préparation des données en amont importante. Pour alléger ces préparatifs et simplifier l'apprentissage, nous proposons d'entraîner un réseau antagoniste génératif conditionné (cGAN) avec en entrée une seule séquence pondérée en T2 (FRFSE). Ce réseau utilisera toute la plage de valeur du CT et de l'IRM. L'entraînement de ce réseau a été effectué dans la région du pelvis de 8 patients, une évaluation croisée à partir des images de ces patients représentant un total de 1298 images a montré une efficacité vraisemblablement supérieure aux cGAN présents dans la littérature.

Abstract – MRI only based treatment planning for radiotherapy requires the generation of reliable synthetic scan (CT) images from MRI ones. Several pre-processing steps are necessary to train a deep learning based image translation method using a dataset of aligned MRI/CT image pairs. To minimize the number of pre-processing steps, we propose to train a conditional generative adversarial network (cGAN) using only T2 (FRFSE) MRI sequences. This network uses the whole span of CT and MRI voxel values. This network is trained on 1298 pelvis region images from 8 patients. The proposed network achieves presumably better cross validated performances as compared to state-of-the-art cGANs.

1 Introduction

L'imagerie par résonance magnétique (IRM) présente un contraste plus élevé par rapport au scanner (CT), c'est pourquoi elle est devenue un standard en radiothérapie. Elle est aussi la première modalité d'imagerie médicale en ce qui concerne le diagnostic, le suivi de la tumeur ainsi que la segmentation des volumes cibles (VC) et des organes à risques (OAR) pour le plan de traitement. Cependant, pour réaliser le calcul de dose d'irradiation curative, le CT reste indispensable puisqu'il permet d'obtenir la densité électronique des tissus nécessaires au calcul de la dose. Actuellement lors des plans de traitements, l'IRM est réalisée puis un oncologue se charge de tracer le contour du VC et des OAR sur les images acquises. Ces contours sont transférés sur le CT grâce à un recalage entre les deux modalités. Cette étape cruciale introduit des incertitudes géométriques [1]. Il est généralement reconnu qu'un plan de traitement se basant uniquement sur l'IRM serait une avancée majeure [2, 3]. L'utilisation unique de l'IRM dans le plan de traitement améliore le confort du patient, évite les incertitudes liées au recalage tout en annulant l'expo-

sition aux rayons ionisants du CT.

Le terme de CT synthétique (sCT) désigne une image IRM dont la valeur des voxels a été convertie en unités Hounsfield (HU), la base de représentation d'une image CT. Des méthodes précédemment proposées ont utilisé des réseaux antagonistes génératifs (GAN) [4] pour la synthèse de sCTs dans la région pelvienne [5] et cérébrale [6], respectivement avec des paires d'images alignées et non alignées. Elles ont toutes les deux fourni de meilleurs résultats par rapport à celle utilisant un réseau (sans adversaire) de type CNN [7, 8]. Comparé aux autres méthodes de génération de sCT [9, 10] (Bulk density, Atlas), les réseaux de neurones permettent d'avoir une meilleure précision et sont significativement plus rapides [7, 8]. Dans ces travaux, nous présentons un GAN conditionné (cGAN) capable de générer un sCT à partir d'une seule séquence IRM (T2-Fast recovery fast spin-echo) sans sur-quantification des images IRM ou CT. En comparaison à une approche multi-séquences [5], notre approche mono-séquence nécessite moins de prétraitements (un seul recalage nécessaire) et réduit le nombre de paramètres à apprendre dans les réseaux de neurones constituant le cGAN, nous permet-

tant d’apprendre un générateur très performant sur des images de plus grande résolution que [5].

2 Données d’apprentissage

Cette étude a été menée sur un total de 8 patients, atteints d’un cancer de la prostate ou du rectum, provenant tous du jeu de données “Gold Atlas Project” [11]. Ce projet a pour but de proposer une base pour la segmentation ou la génération de sCTs. Les acquisitions du CT et de l’IRM ont été réalisées en position de traitement¹. Un recalage rigide a été effectué entre le CT et l’IRM d’un même patient selon une procédure clinique afin de comparer le CT de référence au sCT. La Table 1 reprend les principaux paramètres d’acquisition.

Table 1 – paramètres d’acquisition des images

Nombre de patients	8
Coupes sagittales par patient	74 - 91
CT	Siemens
Modèle	Somatom Definition AS+
Kernel	B30f
IRM	GE
Pondération en T2	Fast Recovery Fast Spin Echo (FRFSE)
Modèle	Discovery 750w
TR (ms)	6000-6600
TE (ms)	97

3 Méthodes et implémentation

3.1 Rappels sur les cGANs

Un GAN [4] est composé de deux réseaux de neurones : le générateur $G(z)$ avec z un vecteur de bruit et le discriminateur $D(x)$. Dans notre problématique, x représente une image CT. L’ensemble des images CT est distribué selon une loi p_x inconnue. G cherche à transformer les vecteurs z en images x de sorte à ce qu’un échantillon de taille n $\{G(z^{(1)}), \dots, G(z^{(n)})\}$ suive la loi p_x . Le discriminateur essaie de distinguer les images réellement distribuées selon p_x de celles produites par son adversaire G . La valeur $D(x)$ est la probabilité que x n’ait pas été généré par G .

Pour faire de la transcription d’images IRM vers CT, on souhaite conditionner ces réseaux avec une image IRM y . Une façon simple d’atteindre cet objectif est d’alimenter ces deux réseaux avec y . Le générateur et le discriminateur deviennent donc $G(y, z)$ et $D(x, y)$ respectivement. Au fur et à mesure de l’entraînement, G doit être capable de générer des x de plus en plus fidèles à la loi p_x rendant ainsi la tâche de classification de D de plus en plus difficile. Les réseaux G et D sont entraînés de façon alternée et attaquent la même fonction objectif. Le discriminateur

cherche à la maximiser tandis que le générateur tente de la minimiser. A la fin de l’entraînement, le générateur induit (dans le cas idéal) une distribution égale à p_x tandis que $D(x, y) = \frac{1}{2}$. La fonction de coût \mathcal{L} du cGAN est

$$\mathcal{L}(G, D) = \mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{z,y}[\log(1 - D(G(y, z), y))] \quad (1)$$

En pratique, on résout $\arg \min_G \max_D \mathcal{L}(G, D) + \lambda \mathcal{L}_1(G)$ où λ est un hyperparamètre contrôlant le terme additionnel $\mathcal{L}_1(G)$ imposant à G de produire des images CT proches en norme-1 de la vraie image CT [12].

Ce réseau est optimisé en suivant l’approche standard de Goodfellow et al. [4] en alternant les étapes de rétro-propagation du gradient entre le générateur et le discriminateur et z est induit par dropout en phase d’apprentissage comme de test.

3.2 Pré-traitement des images

Nous utilisons un nombre réduit de pré-traitements des images IRM afin d’avoir le flux de travail le plus épuré possible. Pour chaque image IRM, une transformation linéaire ramène la plage de valeur des voxels entre le 5^{ème} et 95^{ème} percentiles des valeurs de voxel de l’image initiale. On utilise une suppression du fond des images (CT et IRM) à l’aide d’une approche classique (seuillage) ou semi-supervisée (rectangle englobant). Contrairement à [5], nous n’effectuons aucune sur-quantification et travaillons avec des entiers de 16 bits.

3.3 Architecture du réseau

Nous avons utilisé un cGAN “pix2pix” [12] (voir Figure 4) à l’instar de [5]. Le port en Tensorflow a été modifié pour permettre l’utilisation en entrée et en sortie du réseau d’images en niveaux de gris codées sur 16 bits. Nous avons aussi ajouté 2 couches dans le générateur afin de pouvoir travailler avec des images de tailles (512 × 512). On dénote C_k une couche de (Convolution + Normalisation du Batch + LReLU à 0.2) et \tilde{C}_k une couche de (Convolution + Normalisation du Batch + ReLU) avec un dropout de 0.5 où k représente le nombre de filtres. On utilise pour le générateur comme pour le discriminateur des filtres de taille (4 × 4) avec un stride (2, 2).

3.3.1 Générateur : Architecture U-Net [13]

Cette architecture peut être qualifiée de multi-résolution. Elle se compose d’une première moitié (encodeur) qui conduit le réseau progressivement vers de petites échelles suivi d’une seconde moitié (décodeur) qui reconstruit une sortie de même résolution que l’entrée².

2. Chaque couche du décodeur est suivi d’une convolution transposée permettant de ré-augmenter la résolution d’un facteur de 2.

1. Table d’examen horizontale et non pas ovale

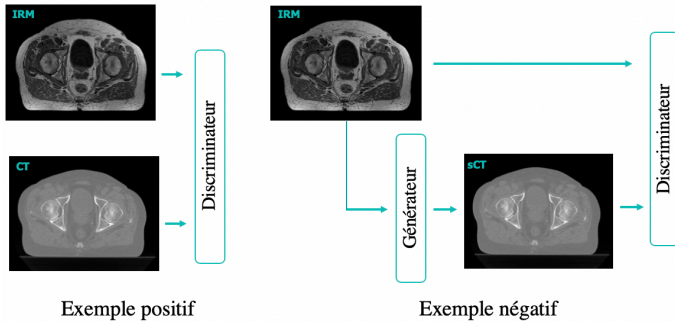


Figure 1 – Entraînement d’un cGAN. Ce réseau est composé d’un générateur de type U-Net et d’un discriminateur de type PatchGAN. Le discriminateur apprend à classifier entre un CT réel et un CT synthétique tandis que le générateur apprend à tromper le discriminateur.

Encodeur : C64³ - C128 - C256 - C512 - C512 - C1024 - C1024 - C1024.

Décodeur : $\tilde{C}1024$ - $\tilde{C}1024$ - $\tilde{C}1024$ - $\tilde{C}512$ - C512 - C256 - C128 - C64 - C1.

On applique une tangente hyperbolique en sortie du décodeur. La gamme de valeur des voxels est rétablie en sortie du générateur par simple ré-encodage en uint16.

3.3.2 Discriminateur : Architecture Patch-GAN [12]

Ce réseau prend en entrée des couples de patches de taille $N \times N$ extraits conjointement de x et y . Le réseau se termine par une couche softmax dont la sortie est la probabilité qu’un patch provient d’une image non artificielle. La probabilité $D(x, y)$ est la moyenne des probabilités sur l’ensemble des patches.

Discriminateur : C64⁴ - C128 - C256 - C512 - C1.

4 Résultats

En moyenne, 1056 coupes transversales ont été utilisées pour un entraînement d’environ 10h sur GPU (NVIDIA Quadro P6000). Une fois le générateur entraîné, il faut en moyenne 8.2 s pour générer le sCT complet d’un patient.

La performance du réseau est évaluée en utilisant une validation croisée à 8 plis. L’entraînement est réalisé sur 7 des 8 patients puis testé sur le 8^{ème}. Pour chaque pli, un seul patient correspond au jeu de validation qui sert à l’évaluation et les 7 autres au jeu d’apprentissage qui sert à optimiser les paramètres du réseau. Ce réseau a été entraîné pendant 200 epochs, avec un batch size de 1 et avec des poids initialisés à partir d’une distribution gaussienne de moyenne nulle et d’écart type de 0,02. Nous utilisons ADAM comme algorithme d’optimisation avec $\alpha =$

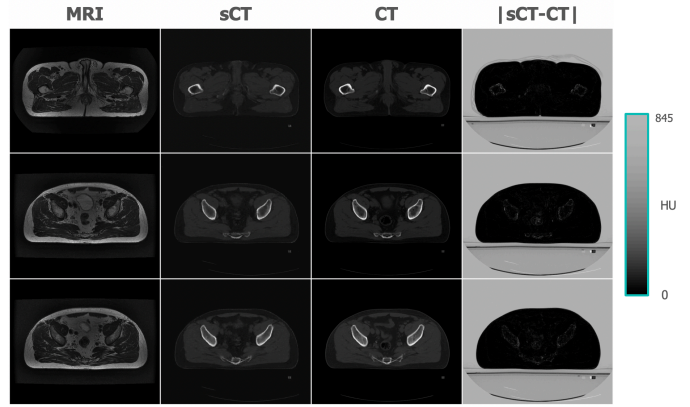


Figure 2 – De gauche à droite Image IRM, le CT généré, le CT de référence et la valeur absolue de la différence entre le CT de référence et le CT synthétique.

0.0002 et $\beta_1 = 0.5$. Une augmentation des données est effectuée en retournant horizontalement et en recadrant aléatoirement les images pour accroître la robustesse du modèle. La précision du sCT généré par rapport au CT de chaque patient est évaluée en terme d’erreur absolue moyenne (EAM) voxel par voxel :

$$EAM = \frac{1}{n} \sum_{i=1}^n |CT(i) - sCT(i)|, \quad (2)$$

où n est le nombre total de voxels appartenant à l’intersection des régions volumiques correspondant à l’intérieur du patient côté CT et sCT.

4.1 Évaluation qualitative et quantitative du sCT

La Figure 4 montre le résultat de la génération de 3 sCT (correspondant au patient #3, #6 et #7) avec notre méthode. Avec elles, l’IRM, le CT correspondant et la différence CT / sCT en valeur absolue. La différence est moins prononcée sur les tissus mous que sur les structures osseuses et les poches d’air. Cela s’explique partiellement à cause de la séquence IRM utilisée, qui ne permet pas de différencier en termes d’intensité l’os cortical de l’air. De plus, la présence d’artefacts non négligeables sur l’IRM au niveau de l’appareil digestif dégrade les conditions de génération. La Table 2 montre une comparaison quantitative entre le CT et sCT pour nos 8 patients, la variabilité des résultats s’expliquent par la présence de poches d’air et d’artefacts plus ou moins important selon les patients.

5 Conclusion

Nous avons montré qu’un cGAN peut générer un CT synthétique à partir d’une seule séquence IRM en utilisant des paires alignées (IRM et CT) pour l’entraînement. Notre architecture a l’avantage d’être plus simple que les

3. Pas de normalisation du batch pour la première couche.

4. Pas de normalisation du batch pour la première couche.

Table 2 – Valeurs de l’erreur absolue moyenne en HU entre le CT et sCT.

Ensemble de test	Nombre de coupes	EAM
Patient 1	76	45.7
Patient 2	91	78.9
Patient 3	74	70.9
Patient 4	75	47.0
Patient 5	88	61.1
Patient 6	86	47.4
Patient 7	85	61.0
Patient 8	74	71.5
Moyenne	1298	60.4

autres méthodes 2D [5, 8] et 2.5D [7] pour la mise en place d’un plan de traitement basé sur l’IRM et d’avoir montré son efficacité sur des images d’une résolution deux fois plus grande. Une évaluation quantitative sur 8 patients donne en moyenne 60.4 (± 13) HU (EAM). En comparaison, Maspero et al. [5] rapportent une EAM de 65 (± 10) HU et Dinkla et al. [7] (2.5D) une de 67 (± 11) HU. Cependant, ces études ont été menées avec des jeux de données différents ce qui rend une comparaison directe impossible. Dans la continuité de ces travaux, nous souhaitons développer un modèle séquentiel utilisant les coupes adjacentes, afin d’atténuer les discontinuités inter-coupes.

Remerciements

Ce projet a reçu des financements au titre du programme Interreg 2 Mers 2014-2020 co-financé par le Fonds Européen de Développement Régional en vertu du contrat de subvention n° 2S04-022.

Références

- [1] T. Nyholm, M. Nyberg, M. G. Karlsson, and M. Karlsson, “Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments,” *Radiation Oncology*, vol. 4, p. 54, dec 2009.
- [2] D. Pasquier, N. Betrouni, M. Vermandel, T. Lacornerie, E. Lartigau, and J. Rousseau, “MRI alone simulation for conformal radiation therapy of prostate cancer : technical aspects,” in 2006 International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 160–163, IEEE, aug 2006.
- [3] T. Nyholm and J. Jonsson, “Counterpoint : Opportunities and Challenges of a Magnetic Resonance Imaging–Only Radiotherapy Work Flow,” *Seminars in Radiation Oncology*, vol. 24, pp. 175–180, jul 2014.
- [4] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” in *Advances in neural information processing systems*, pp. 2672–2680, 2014.
- [5] M. Maspero, M. H. F. Savenije, A. M. Dinkla, P. R. Seevinck, M. P. W. Intven, I. M. Jurgenliemk-Schulz, L. G. W. Kerkmeijer, and C. A. T. van den Berg, “Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy,” *Physics in Medicine & Biology*, vol. 63, p. 185001, sep 2018.
- [6] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Išgum, “Deep mr to ct synthesis using unpaired data,” in *International Workshop on Simulation and Synthesis in Medical Imaging*, pp. 14–23, 2017.
- [7] A. M. Dinkla, J. M. Wolterink, M. Maspero, M. H. Savenije, J. J. Verhoeff, E. Seravalli, I. Išgum, P. R. Seevinck, and C. A. van den Berg, “MR-Only Brain Radiation Therapy : Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network,” *International Journal of Radiation Oncology*Biophysics*Physics*, vol. 102, pp. 801–812, nov 2018.
- [8] X. Han, “MR-based synthetic CT generation using a deep convolutional neural network method,” *Medical Physics*, vol. 44, pp. 1408–1419, apr 2017.
- [9] A. L. Chin, A. Lin, S. Anamalayil, and B.-K. K. Teo, “Feasibility and limitations of bulk density assignment in MRI for head and neck IMRT treatment planning,” *Journal of Applied Clinical Medical Physics*, vol. 15, pp. 100–111, sep 2014.
- [10] B. Demol, C. Boydev, J. Korhonen, and N. Reynaert, “Dosimetric characterization of MRI-only treatment planning for brain tumors in atlas-based pseudo-CT images generated from standard T 1-weighted MR images,” *Medical Physics*, vol. 43, pp. 6557–6568, nov 2016.
- [11] T. Nyholm, S. Svensson, S. Andersson, J. Jonsson, M. Sohlén, C. Gustafsson, E. Kjellén, K. Söderström, P. Albertsson, L. Blomqvist, B. Zackrisson, L. E. Olsson, and A. Gunnlaugsson, “MR and CT data with multiobserver delineations of organs in the pelvic area-Part of the Gold Atlas project,” *Medical Physics*, vol. 45, pp. 1295–1300, mar 2018.
- [12] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, “Image-to-image translation with conditional adversarial networks,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1125–1134, 2017.
- [13] O. Ronneberger, P. Fischer, and T. Brox, “U-net : Convolutional networks for biomedical image segmentation,” in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–241, Springer, 2015.