

Synthèse non supervisée d’images ceCT-CT sous contrainte anatomique

Giammarco LA BARBERA¹, Haithem BOUSSAID², Francesco MASO¹,
Sabine SARNACKI^{3,4}, Laurence ROUET², Pietro GORI¹, Isabelle BLOCH^{5,1,3}

¹LTCI, Télécom Paris, Institut Polytechnique de Paris, France

²Philips Research Paris, Suresnes, France

³IMAG2, Institut Imagine, Université Paris Cité, France

⁴Université Paris Cité, Chirurgie Pédiatrique Viscérale et Urologique, Hôpital Necker Enfants-Malades, APHP, France

⁵Sorbonne Université, CNRS, LIP6, Paris, France

giammarco.labarbera@telecom-paris.fr, haithem.boussaid@philips.com,
francimaso96@gmail.com, sabine.sarnacki@aphp.fr, laurence.rouet@philips.com,
pietro.gori@telecom-paris.fr, isabelle.bloch@sorbonne-universite.fr

Résumé – La segmentation de certaines structures anatomiques dans des images CT avec contraste (ceCT) peut être difficile à cause de la variabilité de la diffusion du produit de contraste. L’utilisation combinée d’images CT et ceCT atténue cette difficulté, mais au prix d’une exposition doublée aux rayonnements. Le remplacement d’une des acquisitions CT par des modèles génératifs permet de maintenir la performance de segmentation, en limitant les doses de rayons X. Nous présentons une extension de CycleGAN pour générer de telles images, à partir de bases d’apprentissage d’images non appariées, où des contraintes anatomiques sont introduites par sélection automatique de la région d’intérêt et par utilisation du score d’un *Self-Supervised Body Regressor*, qui améliore la sélection d’images anatomiquement appariées entre les deux domaines CT et ceCT, et renforce la cohérence anatomique. Les résultats montrent des améliorations qualitatives et quantitatives par rapport aux méthodes existantes.

Abstract – Segmentation of some anatomical structures in contrast-enhanced CT (ceCT) images can be difficult due to the variability in contrast medium diffusion. The combined use of CT and ceCT images alleviates this difficulty, but at the cost of a double radiation exposure. Replacing one of the CT acquisitions with generative models allows maintaining segmentation performance, while limiting radiation doses. We present an extension of CycleGAN to generate such images, from unpaired databases, where anatomical constraints are introduced by automatic selection of the region of interest and by using the score of a *Self-Supervised Body Regressor*, which improves the selection of anatomically-paired images between the two domains (CT and ceCT), and enforces anatomical consistency. Results show qualitative and quantitative improvements compared to state-of-the-art methods.

1 Introduction

La segmentation d’images médicales, en particulier d’images CT avec contraste (ceCT) à l’aide de réseaux de neurones convolutifs (CNN) peut être difficile à cause de l’hétérogénéité et de la variabilité de la diffusion du produit de contraste. Dans [6, 7, 12] les auteurs montrent que l’utilisation conjointe d’images sans contraste (CT) et ceCT est capable de traiter l’hétérogénéité des images ceCT et donc d’améliorer la segmentation. Cependant, afin de limiter les radiations ionisantes, les cliniciens n’acquièrent souvent qu’une seule modalité de CT. En apprentissage profond, l’une des approches courantes pour compenser l’absence d’une modalité d’imagerie consiste à utiliser des modèles génératifs [10] pour la synthétiser. En l’absence d’ensembles de données appariées, les méthodes de synthèse non supervisées reposant sur CycleGAN [11] et UNIT [4] sont une bonne solution [3, 10]. Certains auteurs ont déjà appliqué ces méthodes pour la synthèse d’images abdominales ceCT-CT [6, 7, 12].

Cependant, ces modèles n’ont pas la capacité de préserver les structures fines [12] et peuvent produire des artefacts [10], ce qui empêche l’application de ces méthodes aux vaisseaux sanguins. Afin d’améliorer ces méthodes, nous proposons une extension de l’approche par CycleGAN qui comprend :

- (i) la sélection automatique de la zone d’intérêt en exploitant les informations anatomiques, afin de réduire la distribution anatomique des données ;
- (ii) l’utilisation d’un régresseur sur l’anatomie (*Self-Supervised Body Regressor*, SSBR), similaire à celui de [8], pour mieux sélectionner des images anatomiquement appariées entre les domaines ceCT et CT non appariés, en aidant le discriminateur à se spécialiser dans la tâche ;
- (iii) l’utilisation du score SSBR comme fonction de coût qui contraint le générateur à produire une coupe décrivant le même contenu anatomique que l’entrée, ressemblant au classificateur auxiliaire des GAN [5] ;

(iv) l'utilisation de l'image d'entrée comme modèle pour le générateur, comme dans [1], et l'utilisation d'un masque binaire anatomique pour contraindre la sortie.

Nous appliquons notre méthode à la génération d'images abdominales CT à partir d'images ceCT et vice versa, mais elle est exploitable pour d'autres applications. Nous montrons que notre méthode améliore considérablement la qualité de la synthèse des deux modalités par rapport aux méthodes existantes. Il est important de souligner que, dans ce travail, l'utilisation d'images synthétiques vise à augmenter les modalités disponibles pour un futur réseau de segmentation et que leur utilisation n'est pas destinée au diagnostic clinique.

2 Méthode proposée

La méthode décrite dans cette section est résumée par le schéma de la figure 1.

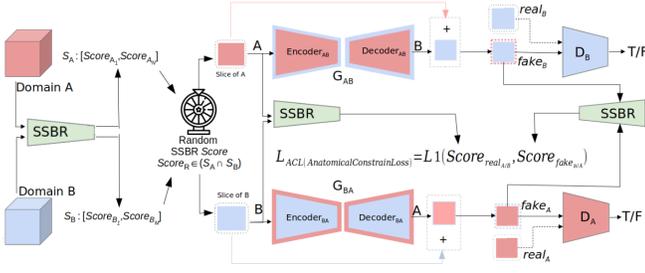


FIGURE 1 – Méthode proposée pour la sélection de coupes anatomiquement appariées par *Self-Supervised Body Regressor*, et son utilisation comme fonction de coût L_{ACL} .

Sélection des coupes anatomiquement appariées par SSBR

En raison des limites actuelles des GPU, une approche courante pour l'apprentissage d'un réseau GAN consiste à utiliser des coupes 2D, mais au prix de la perte d'informations 3D. Il est donc important de sélectionner uniquement les coupes de la région abdominale pour réduire la région d'intérêt (ROI) et donc la distribution des données, et faciliter ainsi le processus génératif. Cela pourrait être fait hors ligne et manuellement, comme dans [6, 12]. Ici, au contraire, nous proposons une méthode automatique et en ligne. Nous sélectionnons automatiquement la première coupe des poumons et la dernière coupe de la zone intestinale respectivement comme points de repère supérieur et inférieur, car ils sont faciles à détecter en raison de la forte présence de pixels noirs dans les deux acquisitions ceCT et CT.

Dans une ROI constante, la difficulté de traiter des ensembles de données médicales non appariées peut être atténuée en exploitant l'anatomie approximativement commune entre les sujets [9]. Néanmoins, dans la région abdominale, il faut tenir compte des différentes tailles et longueurs des organes, impliquant que deux coupes abdominales à la même position relative

peuvent ne pas avoir le même contenu anatomique. Pour surmonter ce problème, nous proposons l'utilisation d'un SSBR, comme le montre la figure 1. Présenté dans [8], un *body-regressor* non supervisé est un CNN qui apprend un système de coordonnées corporelles représentant l'apparence anatomique à partir d'images CT non étiquetées.

Pour l'apprentissage nous proposons l'optimisation de trois fonctions de coût. La première, comme dans [8], favorise un ordre croissant des scores du SSBR en fonction des positions des coupes :

$$L_{order} = - \sum_{k=1}^K \sum_{p=1}^{P-1} \log(h(\text{Score}_{k,p+1} - \text{Score}_{k,p})) \quad (1)$$

où $\text{Score}_{k,p}$ est la sortie du SSBR pour la coupe p du volume CT k , h est la fonction d'activation sigmoïde, K est le nombre de volumes CT dans l'ensemble (*mini-batch*) choisi et P est le nombre de coupes dans chaque volume.

La deuxième fonction de coût exploite la sélection automatique de la ROI, forçant la première et la dernière coupes à avoir un score de -1 et 1 respectivement :

$$L_{norm} = \sum_{k=1}^K (f(\text{Score}_{k,1} + 1) + f(\text{Score}_{k,P} - 1)) \quad (2)$$

où f est une norme L1 lissée.

Enfin, la troisième fonction de coût prend en compte la variabilité anatomique de la zone abdominale. En utilisant le masque binaire BM du corps pour chaque coupe (obtenu facilement en CT), nous voulons que la différence entre les scores successifs soit proportionnelle au cardinal normalisé de l'intersection des BM des coupes successives :

$$L_{anat} = \sum_{k=1}^K \sum_{p=1}^{N-1} f(\Delta_{k,p+1}^{BM} - \Delta_{k,p}^{BM}) \quad (3)$$

$$\Delta_{k,p}^{BM} = 1 - \frac{|BM_{k,p} \cap BM_{k,p-1}|}{|BM_{k,p-1}|}$$

$$\Delta_{k,p} = \text{Score}_{k,p} - \text{Score}_{k,p-1}$$

Finalement, les termes de la fonction de coût sont combinés par une moyenne pondérée et la fonction à optimiser s'écrit :

$$L_{SSBR} = \alpha L_{order} + \beta L_{anatomy} + \gamma L_{norm} \quad (4)$$

où α , β et γ sont les poids.

Une fois que le SSBR est correctement entraîné, pour extraire les coupes anatomiquement appariées pour chaque itération du CycleGAN, nous opérons comme suit :

1. un seul patient est sélectionné pour chacun des domaines ceCT et CT non appariés, appelés domaines A et B ;
2. les volumes 3D sont automatiquement limités à la région abdominale ;
3. les scores SSBR sont prédits pour chaque coupe 2D des deux ROI 3D, avec $S_A : [\text{Score}_{A_1}, \dots, \text{Score}_{A_N}]$ et $S_B : [\text{Score}_{B_1}, \dots, \text{Score}_{B_M}]$, où N est le nombre de coupes du patient du domaine A et M celui du patient du domaine B ;

4. J scores SSBR aléatoires, notés $Score_{R_j}$, sont échantillonnés dans $S_A \cap S_B$, où J est le nombre sélectionné de coupes correspondant à la taille du *mini-batch* choisi ;
5. Pour chaque $Score_{R_j}$, la coupe dont le score est le plus proche est sélectionnée dans chaque domaine, comme $\arg \min |Score_{R_j} - Score_{i,p}|$ où i est le domaine (A ou B) et p est la coupe correspondante dans $[1, N]$ (respectivement $[1, M]$).

CycleGAN sous contrainte anatomique Inspirés par [5], nous proposons l'utilisation du SSBR pré-entraîné comme classifieur auxiliaire pour renforcer la cohérence des parties du corps entre l'entrée et la sortie synthétisée. Pendant la phase d'apprentissage du générateur, nous ajoutons aux fonctions de coût du CycleGAN standard une distance L1 entre le score SSBR de l'entrée *real* A (respectivement B) et le score SSBR de la coupe générée *fake* B (respectivement A), appelée contrainte anatomique (*Anatomical Constraint Loss, ACL*), comme le montre la figure 1 :

$$L_{ACL} = \frac{1}{J} \sum_{j=1}^J |Score_{real_{A/B,j}} - Score_{fake_{B/A,j}}| \quad (5)$$

Nous proposons également de contraindre davantage les modèles d'un point de vue anatomique de deux manières. Premièrement, comme dans [1], nous utilisons l'image d'entrée comme modèle, c'est-à-dire que les générateurs doivent seulement estimer comment modifier l'image d'entrée sans estimer une image de sortie à partir de zéro. Deuxièmement, pendant l'inférence, nous supprimons les artefacts créés dans les zones vides originales (par exemple, le fond ou l'air). Nous utilisons ici un masque binaire obtenu comme pour l'équation 3.

3 Résultats et discussion

Données Pour l'apprentissage non apparié, nous utilisons les images de *The Cancer Imaging Archive* (TCIA) [2]. Deux bases de données d'images CT abdominales sans pathologie ont été utilisées : *CT Pancreas* avec 82 images ceCT d'abdomens, et *CT Colonography* qui contient des images CT sans contraste dont 82 sujets sains ont été conservés pour la cohérence avec le premier jeu de données. Pour les deux ensembles de données, 72 patients sont utilisés pour l'apprentissage et 10 pour les tests qualitatifs. Toutes les coupes axiales ont été redimensionnées de 512×512 à 128×128 pixels pour rendre l'apprentissage moins coûteux en temps de calcul et en mémoire. Pour les tests quantitatifs, nous avons utilisé une base privée de données pédiatriques d'images abdominales ceCT-CT appariées de 10 patients atteints de tumeurs rénales. Il est important de souligner que pour ce jeu de données le nombre restreint de patients ne permet pas son utilisation pour les modèles génératifs.

Apprentissage Les hyperparamètres sont les mêmes que ceux dans [11]. Tous les tests ont été effectués à l'aide d'un GPU

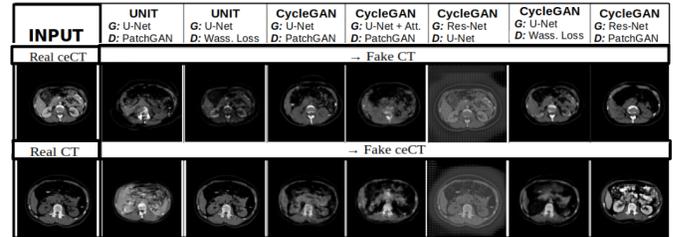


FIGURE 2 – Comparaison de quelques méthodes sur des coupes de l'ensemble de test non apparié. 1e ligne : de ceCT à CT. 2e ligne : de CT à ceCT. Les coupes dans tous les tests sont sélectionnées avec PBS [9]. L'entrée dans l'autre sens donne une idée de ce à quoi devrait ressembler le résultat attendu.

avec 16 Go de VRAM et une taille de *mini-batch* de 8. De plus, l'utilisation des seules coupes de la région rénale a été jugée essentielle pour nos expériences afin d'obtenir des images anatomiquement cohérentes, et notre détection automatique de la région abdominale s'est avérée efficace, supprimant la nécessité d'une sélection manuelle. Pour le SSBR, nous avons opéré comme dans [8], en utilisant ResNet-34 comme réseau *backbone*. Empiriquement, la meilleure combinaison de poids dans l'équation 4 a été trouvée comme $\alpha = 0.005, \beta = 1, \gamma = 10$.

Résultats qualitatifs sur les données non appariées Tout d'abord, nous avons testé plusieurs des méthodes existantes [3]. L'évaluation a permis d'identifier la meilleure méthode pour la synthèse ceCT-CT dans le CycleGAN original [11] avec ResNet comme réseau générateur et PatchGAN comme mécanisme discriminant. Dans tous ces tests, nous avons utilisé la méthode PBS [9] pour la sélection des coupes à la même position relative, entre sujets des domaines médicaux non appariés, qui peuvent présenter des ROI de longueurs différentes. Les résultats sont présentés dans la figure 2.

Malgré les bons résultats du CycleGAN identifié comme le meilleur, en matière de forme générale et d'intensité du contraste, la sélection PBS n'est pas suffisante et plusieurs artefacts anatomiques apparaissent. Puis nous avons testé l'utilisation du recalage affine. Étant donné la grande variabilité entre les deux domaines et la difficulté d'identifier l'image fixe pour l'ensemble de données, nous avons décidé d'effectuer le recalage à chaque itération entre les deux patients sélectionnés. La cohérence anatomique est améliorée au prix d'un temps de calcul de 15h au lieu de 3h, mais quelques artefacts importants apparaissent encore. Enfin, la sélection avec SSBR a été testée, ce qui a permis de diminuer la sévérité des artefacts.

Nous avons ensuite combiné toutes les améliorations proposées. La méthode complète avec l'utilisation de la fonction de coût L_{ACL} , l'addition de l'entrée à la sortie et le masque binaire a produit des images synthétiques de haute qualité, sans artefact visuel et avec intensité de contraste réaliste. Quelques résultats qualitatifs sont montrés dans la figure 3.

Étude quantitative sur les données appariées Une étude quantitative d'ablation et comparative a été faite sur notre base

TABLE 1 – Étude quantitative sur les 10 patients avec images appariées. L’erreur quadratique moyenne (MSE), la similarité de structures (SSIM) et le rapport signal sur bruit maximal (PSNR) sont indiqués en moyenne et en écart-type. TIME est le temps d’apprentissage. Tests reposant sur CycleGAN. Les quatre dernières lignes présentent l’étude d’ablation.

CycleGAN Method	MSE [10^{-2}] (\downarrow)	SSIM [10^{-1}] (\uparrow)	PSNR (\uparrow)	TIME (\downarrow)
real CT \rightarrow fake ceCT vs real ceCT				
PBS	10,05 (2,89)	5,76 (0,65)	16,14 (1,15)	3h 2m
AFFINE REG.	8,16 (1,80)	6,36 (0,57)	16,99 (0,87)	16h 33m
SSBR selection	9,07 (2,39)	5,99 (0,71)	16,56 (1,07)	7h 5m
+ L_{ACL}	8,55 (2,28)	6,19 (0,69)	16,82 (1,07)	7h 49m
+ input addition	6,41 (1,97)	6,67 (0,63)	18,11 (1,22)	7h 55m
+ binary mask	6,37 (2,01)	6,81 (0,62)	18,14 (1,23)	7h 55m
real ceCT \rightarrow fake CT vs real CT				
PBS	8,26 (1,97)	5,36 (0,28)	16,96 (1,04)	3h 2m
AFFINE REG.	4,72 (0,95)	6,77 (0,37)	19,36 (0,93)	16h 33m
SSBR selection	7,15 (2,16)	5,68 (0,52)	17,64 (1,26)	7h 5m
+ L_{ACL}	5,87 (1,73)	6,08 (0,22)	18,47 (1,12)	7h 49m
+ input addition	4,24 (0,86)	6,80 (0,37)	19,83 (0,92)	7h 55m
+ binary mask	4,05 (0,83)	7,23 (0,53)	20,03 (0,92)	7h 55m

de données appariées de 10 patients en utilisant les méthodes présentées, pré-entraînées sur les ensembles de données non appariées. Les résultats sont présentés dans la table 1 en utilisant l’erreur quadratique moyenne (MSE), la similarité de structures (SSIM) et le rapport signal sur bruit maximal (PSNR) entre les images *real* et *fake*, et le temps d’apprentissage (TIME). Toutes nos contributions améliorent le CycleGAN original avec PBS [9], au prix de quelques heures d’apprentissage supplémentaires. L’utilisation du recalage affine semble être quantitativement comparable à l’utilisation du SSBR pour la sélection et la fonction de coût L_{ACL} , mais le réseau nécessite un temps de calcul très élevé et conduit de plus à la création d’artefacts, comme illustré dans la figure 3.

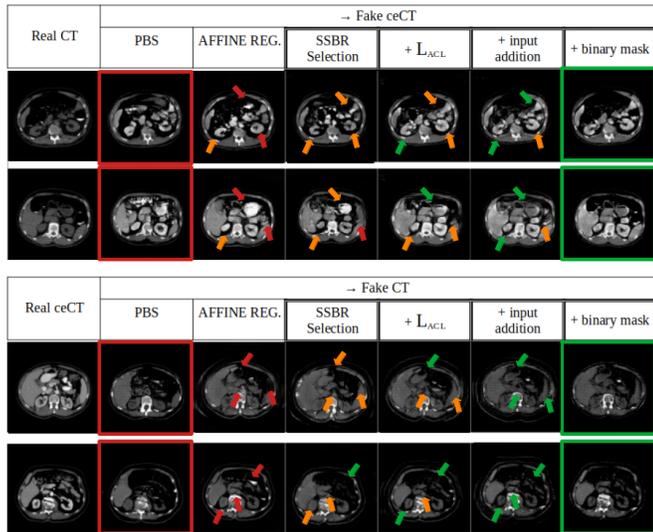


FIGURE 3 – Résultats qualitatifs sur des coupes non appariées. Tableau en haut : du CT au ceCT. En bas : du ceCT au CT. Tests reposant sur CycleGAN. Quatre dernières colonnes : résultats avec nos méthodes (détaillées dans le texte). Flèches : artefacts élevés (rouge), légers (orange) et absents (vert).

4 Conclusion

Nous avons présenté une extension de CycleGAN pour contraindre anatomiquement la synthèse non supervisée d’images ceCT-CT. Nous avons montré des améliorations significatives dans les images générées par rapport aux méthodes existantes. Pour valider davantage la méthode, les travaux futurs visent à prouver que les images synthétisées peuvent être utilisées pour guider une méthode de segmentation en compensant, sans perdre en performance, l’absence de certaines acquisitions.

Références

- [1] H Chang, J Lu, F Yu, et al. PairedCycleGAN : Asymmetric style transfer for applying and removing makeup. In *IEEE CVPR*, page 40–48, 2018.
- [2] K Clark, B Vendt, K Smith, et al. The Cancer Imaging Archive (TCIA) : maintaining and operating a public information repository. *Journal of Digital Imaging*, 26(6) :1045–1057, 2013.
- [3] S Karthika and M Durgadevi. Generative adversarial network (gan) : a general review on different variants of gan and applications. In *ICCES*, pages 1–8, 2021.
- [4] M Y Liu, T Breuel, and J Kautz. Unsupervised image-to-image translation networks. In *NeurIPS*, pages 700–708, 2017.
- [5] A Odena, C Olah, and J Shlens. Conditional Image Synthesis With Auxiliary Classifier GANs. In *ICLR*, 2017.
- [6] V Sandfort, K Yan, P Pickhardt, et al. Data augmentation using generative adversarial networks (CycleGAN) to improve generalizability in CT segmentation tasks. *Scientific Reports*, 9(1) :1–9, 2019.
- [7] C Song, B He, H Chen, et al. Non-contrast CT Liver Segmentation Using CycleGAN Data Augmentation from Contrast Enhanced CT. In *iMIMIC Workshop (MICCAI)*, pages 122–129, 2020.
- [8] K Yan, L Lu, and R Summers. Unsupervised body part regression via spatially self-ordering convolutional neural networks. In *IEEE ISBI*, pages 1022–1025, 2018.
- [9] H Yang, J Sun, An Carass, et al. Unsupervised MR-to-CT synthesis using structure-constrained cyclegan. *IEEE TMI*, 39(12) :4249–4261, 2020.
- [10] X Yi, E Walia, and P Babyn. Generative adversarial network in medical imaging : A review. *Medical Image Analysis*, 58 :101552, 2019.
- [11] J Y Zhu, T Park, P Isola, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *IEEE ICCV*, pages 2223–2232, 2017.
- [12] Y Zhu, Y Tang, Y Tang, et al. Cross-domain medical image translation by shared latent Gaussian mixture model. In *MICCAI*, pages 379–389, 2020.