



HAL
open science

Synthèse de mouvement respiratoire par réseau antagoniste génératif doublement conditionnel en imagerie tomodensitométrie 4D

Yi-Heng Cao, Vincent Jaouen, Vincent Bourbonne, Nicolas Boussion, Ulrike Schick, Julien Bert, Dimitris Visvikis

► **To cite this version:**

Yi-Heng Cao, Vincent Jaouen, Vincent Bourbonne, Nicolas Boussion, Ulrike Schick, et al.. Synthèse de mouvement respiratoire par réseau antagoniste génératif doublement conditionnel en imagerie tomodensitométrie 4D. ORASIS 2021, Centre National de la Recherche Scientifique [CNRS], Sep 2021, Saint Ferréol, France. hal-03339736

HAL Id: hal-03339736

<https://hal.science/hal-03339736>

Submitted on 9 Sep 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Synthèse de mouvement respiratoire par réseau antagoniste génératif doublement conditionnel en imagerie tomodensitométrie 4D

Yi-heng Cao¹
Ulrike Schick^{1,2}

Vincent Jaouen¹
Julien Bert^{1,2}

Vincent Bourbonne^{1,2}
Dimitris Visvikis¹

Nicolas Boussion^{1,2}

¹ UMR 1101 Inserm LaTIM, Université de Bretagne Occidentale, IMT Atlantique, Brest, France

² CHRU Brest University Hospital, Brest, France

yi-heng.cao@univ-brest.fr

Résumé

Les réseaux image-à-image tels que pix2pix et CycleGAN ont permis des progrès importants en synthèse d'image ces dernières années. Ils sont désormais appliqués régulièrement en imagerie médicale 3D pour un grand nombre d'applications. Ces méthodes ont toutefois été peu étudiées dans le cadre de l'imagerie médicale dynamique 4D, qui tient compte de la composante temporelle présente dans certaines acquisitions.

Dans ce travail, nous proposons une nouvelle architecture de réseaux antagonistes génératifs (GAN) doublement conditionnels pour l'imagerie respiratoire 4D scanner (4DCT) à partir d'images statiques (3DCT) et d'une mesure externe de mouvement respiratoire. Une mesure scalaire d'amplitude de respiration du patient agit comme un style au travers d'une couche Adaptive Instance Normalization (AdaIN) permettant une régression des différentes phases respiratoires. Nous présentons des résultats sur des images synthétiques à divers degrés de perturbation et des images réelles d'acquisitions pulmonaires 4DCT.

Mots Clef

Synthèse d'images, GAN, imagerie dynamique, mouvement respiratoire, radiothérapie

Abstract

Image-to-image translation GANs such as pix2pix and CycleGAN have made significant progress in recent years. They are now regularly applied in 3D medical imaging for a wide range of applications. However, these methods have been relatively unexplored in the context of 4D dynamic medical imaging, which takes into account the temporal component present in some acquisitions.

In this work, we propose a new architecture based on a doubly conditional generative adversarial networks (GANs) for 4D respiratory computed tomography (4DCT) imaging using static images (3DCT) and an external respiratory motion measure. A scalar measure of the patient's breathing amplitude acts as a style through an Adaptive Instance Normalization (AdaIN) layer allowing regression

of the different breathing phases. We present results on synthetic images with various degrees of perturbation and real images of 4DCT lung acquisitions.

Keywords

Image synthesis, GAN, dynamic imaging, respiratory motion, radiotherapy

1 Introduction

La radiothérapie (RT) est indiquée chez environ 50% des patients atteints de cancer [6], et dans plus de 75% des cas pour le cancer du poumon [25]. Dans ce contexte, la tomodensitométrie 4D (scanner 4D ou 4DCT) est une modalité d'acquisition dynamique permettant d'acquérir les différentes phases du mouvement respiratoire du patient, et ainsi d'identifier de manière plus précise les volumes-cibles cliniques tenant compte de ce mouvement. L'imagerie 4DCT s'est ainsi imposée en routine clinique dans la planification du traitement du cancer du poumon par RT [7].

Cependant, ces acquisitions entraînent une exposition plus élevée aux rayonnements, jusqu'à six fois supérieure à celle de l'imagerie scanner conventionnelle (3DCT), en raison des temps d'acquisition plus longs [5]. Avec les progrès rapides de l'apprentissage profond en imagerie médicale, il existe un intérêt croissant pour le développement de nouvelles méthodes logicielles basées sur l'intelligence artificielle qui pourraient réduire la dose induite par de telles acquisitions [8].

Les réseaux de synthèse d'image-à-image (*Image-to-Image translation* - I2I) basés sur des réseaux antagonistes génératifs (GANs) tels que *pix2pix* [14] et ses dérivés [28, 13, 27, 20] ont permis des progrès spectaculaires en synthèse d'images ces dernières années et ont été appliqués avec succès en imagerie médicale pour des applications telles que la segmentation, la reconstruction ou la réduction du bruit [16] du fait de leur haut niveau de réalisme et de leur relative facilité d'entraînement [1]. Actuellement, la plupart des GANs I2I en imagerie médicale sont toutefois appliqués à l'imagerie conventionnelle volumétrique 3D, avec peu de méthodes dédiées à l'imagerie dynamique.

Dans le domaine de la vision par ordinateur, des modèles I2I récents permettent de modifier les attributs de l'image synthétisée avec notamment la possibilité d'exercer un contrôle explicite du style au travers d'apprentissage d'attributs style-spécifiques [3], ou encore de contrôler le degré d'interpolation inter-domaine dans l'espace latent [26]. Récemment, Pizzati et al. ont proposé une architecture permettant de nouveaux types de transitions continues contrôlées par des modèles physiques naïfs tels que des transitions basées sur des mouvements simples [21]. Toutefois, les transitions obtenues, bien que contraintes par une fonction coût antagoniste, ne permettent pas un contrôle explicite du degré de régularité de la déformation, ce qui est une limitation en imagerie médicale. Dans l'objectif de permettre un meilleur contrôle des déformations spatiales, Guo et al. [9] ont récemment proposé un réseau d'interpolation volumétrique spatio-temporel conçu pour les images médicales cardiaques 4D, où un réseau de recalage et un réseau de déformation font l'objet d'un apprentissage simultané pour permettre l'interpolation entre deux phases dynamiques.

Parallèlement, nous avons proposé pour la première fois en 2019 de considérer des modèles I2I basé sur des réseaux antagonistes génératifs conditionnels (cGANs) en vue d'une synthèse réaliste de dynamique respiratoire en imagerie 4DCT à partir d'images 3DCT [15]. Cette preuve de concept se base sur un ensemble de cGANs bidimensionnels synthétisant chacun une phase respiratoire de l'acquisition 4DCT. Mori et al. ont par la suite proposé une méthode basée sur des idées similaires pour synthétiser des champs de vecteurs de déformation obtenus par une méthode de recalage conventionnelle utilisant un critère d'information mutuelle et une contrainte de régularité de type B-spline [18]. Le recalage non-rigide d'images médicales résiste en effet jusqu'à présent mieux que d'autres domaines aux progrès en apprentissage profond. Le contrôle explicite de la régularité de la déformation paramétrique produite par des méthodes d'optimisation conventionnelles basées sur une fonction de coût impliquant une régularisation spatiale demeure actuellement la stratégie la plus robuste [24, 10]

L'une des principales limites des approches de synthèse respiratoire mentionnées est que les phases obtenues ne correspondent pas à la dynamique respiratoire du patient mais à une moyenne de chaque phase de la base d'entraînement. Dans un objectif de réduction de dose en radiothérapie, il est toutefois naturellement impératif que cette synthèse soit spécifique au patient. À notre connaissance, le potentiel des réseaux I2I pour générer des images dynamiques à partir d'images statiques reproduisant fidèlement l'amplitude respiratoire réelle du patient n'a pas encore été exploré.

Dans ce travail, nous proposons une nouvelle architecture de réseau I2I doublement conditionnel pour la synthèse d'images 4D à partir d'images 3D et de mesure d'amplitude du mouvement respiratoire pour la radiothérapie in-

terventionnelle du cancer du poumon. La valeur injectée durant la phase d'apprentissage correspond à un paramètre scalaire d'amplitude respiratoire, une information typiquement disponible en routine clinique par des ceintures de synchronisation. Nous démontrons qu'un tel conditionnement double par l'image et l'amplitude permet de synthétiser des images 4D fidèles et réalistes à partir d'images 3D statiques. Dans nos expérimentations, nous présentons des résultats à la fois sur des données synthétiques de complexité croissante et en présence de bruit, ainsi que sur des acquisitions pulmonaires 4DCT réelles, où nous montrons que le modèle est en mesure de reproduire fidèlement la phase respiratoire du patient.

Nous notons que, très récemment, Chang et al. [2] ont proposé une méthode basée sur un réseau auto-encodeur variationnel conditionnel combiné à un régresseur latent pour la génération de mouvements respiratoires réalistes et variés dans des fantômes de patient 4D-XCAT, une méthodologie alternative pour un objectif de synthèse différent [22].

La suite de cet article est organisée de la manière suivante. La méthode proposée est décrite dans la section suivante. La section 3 détaille le protocole expérimental de validation. Nous présentons nos résultats en section 4. Enfin, nous terminons par une conclusion en section 5.

2 Méthode proposée

Dans cette section, nous détaillons notre méthodologie de synthèse de mouvement 4DCT à partir d'images 3DCT et d'une information scalaire d'amplitude respiratoire. Un schéma général de principe est présenté sur la Fig. 1.

Les images 4DCT sont obtenues par la technique dite de "gating" qui consiste à synchroniser l'imagerie avec la respiration du patient. On obtient ainsi une image dynamique $\mathcal{J} : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}$, où $\mathcal{J}(\mathbf{x}, g)$ est la valeur de l'image à la position \mathbf{x} pour la phase (gate) $g \in \{1, 2, \dots, N_g\}$.

Soit une image 3DCT de référence $\mathcal{I}(\mathbf{x})$ où $\mathbf{x} \in \mathbb{R}^3$ est le domaine de l'image. À l'aide d'une architecture I2I, nous cherchons à apprendre une application implicite $\varphi_g(\mathbf{x}, \alpha)$ doublement conditionnée par l'image \mathcal{I} et une valeur d'amplitude α , de telle manière que l'image générée par un générateur $\mathcal{G}, \mathcal{I} \circ \varphi_g(\mathbf{x}, \alpha) = \mathcal{G}(\mathbf{x}, \alpha)$ partage la même dynamique respiratoire que l'image réelle $\mathcal{J}(\mathbf{x}, g)$. Par conséquent, nous supposons ainsi qu'il existe une relation de bijection entre une amplitude respiratoire α et une phase g obtenue par gating : toute phase g est uniquement identifiée par une amplitude α .

Nous adaptons une architecture GAN 3D existante simplement conditionnelle (par une image d'entrée) basée sur l'architecture *pix2pix* nommée *vox2vox*, récemment proposée pour la segmentation de tumeurs cérébrales [4]. Nous modifions le générateur G afin qu'il soit conditionné par deux facteurs : à la fois l'image \mathcal{I} et une valeur scalaire d'amplitude α tel que $G : \{\mathcal{I}(\mathbf{x}), \alpha\} \rightarrow \mathcal{G}(\mathbf{x}, \alpha)$. La fonction de coût de notre GAN conditionnel se définit comme :

$$\mathcal{L}_{cGAN}(G, D) = \mathbb{E}_{\mathcal{I}, \mathcal{J}}[\log D(\mathcal{I}, \mathcal{J})] + \mathbb{E}_{\mathcal{I}, \mathcal{G}}[\log(1 - D(\mathcal{I}, \mathcal{G}))], \quad (1)$$

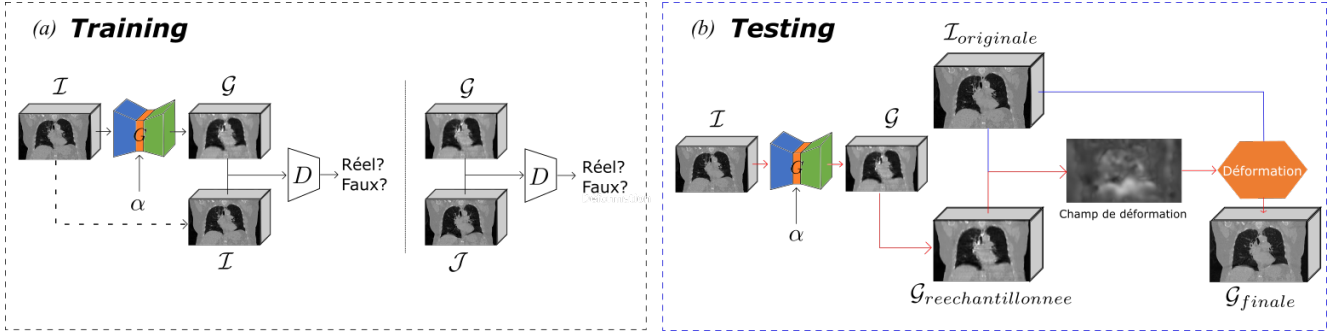


FIGURE 1 – Apprentissage de paires image-amplitude. (a) Un GAN est doublement conditionné par 1) une image \mathcal{I} et 2) l’amplitude respiratoire α correspondant à une phase g . L’image générée \mathcal{G} est ensuite comparée à l’image \mathcal{I} et l’image vérité-terrain \mathcal{J} de phase g . (b) À l’inférence, un champ de déformation est calculé à partir de l’image d’entrée et la sortie du GAN par une méthode de recalage conventionnelle B-spline pour contraindre le réalisme de la déformation. Ce champ de déformation est finalement appliqué à l’image d’entrée.

où G essaie de minimiser cet objectif contre le modèle adverse D qui essaie de le maximiser, c’est-à-dire :

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) \quad (2)$$

Conformément à l’objectif de synthèse, nous remplaçons la Dice-loss proposée dans le cas de $vox2vox$ par une perte voxel-à-voxel L_1 telle que décrite dans l’implémentation originale de $pix2pix$:

$$G^* = \arg \min_G \max_D \mathcal{L}_{cGAN}(G, D) + \lambda \mathcal{L}_{L_1}(G) \quad (3)$$

Pour permettre au modèle d’être doublement conditionnel, nous ajoutons une couche à la fin de l’encodeur du générateur au niveau du goulot d’étranglement afin de le conditionner par la valeur α . Cette couche est composée de l’*Adaptive Instance Normalization* (AdaIN, [12]) habituellement utilisée dans les tâches de transfert de style où elle permet de conditionner un générateur par un style d’une image de référence. Dans le cas présent, nous remplaçons cette image de référence par une valeur scalaire d’amplitude α . Elle est ainsi définie par :

$$\text{AdaIN}(\mathbf{x}, \alpha) = \gamma(\alpha) \left(\frac{\mathbf{x} - \mu(\mathbf{x})}{\sigma(\mathbf{x})} \right) + \beta(\alpha), \quad (4)$$

où γ et β sont déterminées par le scalaire α après être passé par une couche entièrement connectée et une couche de convolution à 1-dimension.

Ainsi, le générateur G reçoit une image d’entrée \mathcal{I} et est entraîné à synthétiser une image cible $\mathcal{G}(\mathbf{x}, \alpha)$, tandis que le discriminateur D est entraîné à distinguer l’image de phase réelle $\mathcal{J}(\mathbf{x}, g)$ de $\mathcal{G}(\mathbf{x}, \alpha)$. Le modèle apprend donc à produire un mouvement respiratoire spécifique à partir d’images 3DCT et d’un scalaire correspondant à l’amplitude respiratoire.

Comme évoqué précédemment, les images ainsi générées sont susceptibles de présenter des déformations peu réalistes du fait de l’absence de contrainte explicite sur la nature de la déformation. Il convient ainsi de contraindre

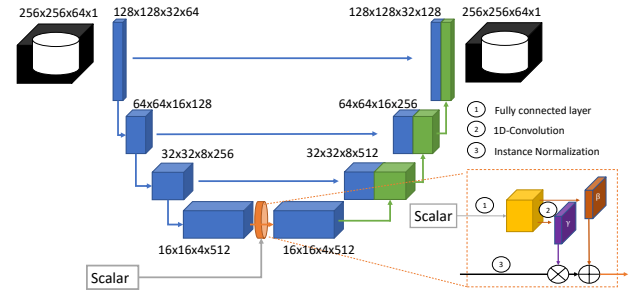


FIGURE 2 – Architecture du générateur proposée avec la couche de conditionnement par injection scalaire.

de manière plus explicite cette déformation. Dans ce but, nous proposons à l’inférence d’employer une méthode de recalage conventionnelle entre l’image 4DCT et la sortie rééchantillonnée dans le domaine d’origine. En pratique, nous utilisons une méthode populaire en imagerie 4DCT basée sur l’algorithme de recalage B-spline de Staring et al. développé pour la librairie *elastix* [23]. On cherche une transformation paramétrique φ_{μ}^{\dagger}

$$\hat{\mu} = \arg \min_{\mu} \mathcal{C}(\varphi_{\mu}^{\dagger}; \mathcal{G}, \mathcal{I}), \quad (5)$$

paramétrée par le vecteur de paramètres μ , où \mathcal{C} est la fonction de coût qui définit la qualité de l’alignement.

La fonction de coût employée repose sur le coefficient de corrélation normalisée (*Zero-mean Normalised Correlation Coefficient* - ZNCC) qui convient pour le recalage d’images mono-modales, et qui peut compenser les différences d’intensité globales dues aux différences de niveau d’inspiration. Elle est définie par :

$$\text{ZNCC}(\mu; \mathcal{G}, \mathcal{I}) = \frac{\sum_{\mathbf{x}_i} (\mathcal{G}(\mathbf{x}_i) - \bar{\mathcal{G}})(\mathcal{I}(\varphi_{\mu}^{\dagger}(\mathbf{x}_i)) - \bar{\mathcal{I}})}{\sqrt{\sum_{\mathbf{x}_i} (\mathcal{G}(\mathbf{x}_i) - \bar{\mathcal{G}})^2 \sum_{\mathbf{x}_i} (\mathcal{I}(\varphi_{\mu}^{\dagger}(\mathbf{x}_i)) - \bar{\mathcal{I}})^2}}, \quad (6)$$

où x_i sont des échantillons tirés des images avec $\bar{\mathcal{I}}$ et $\bar{\mathcal{G}}$ les valeurs d'intensités moyennes pour \mathcal{I} et \mathcal{G} .

Cette hybridation entre une approche conventionnelle de recalage et apprentissage profond permet d'exploiter la capacité de génération de notre architecture doublement conditionnelle tout en préservant les propriétés souhaitables des méthodes de recalage conventionnelles. De cette façon, nous améliorons la qualité de l'image résultante et obtenons des images d'un réalisme supérieur tout en gardant un contrôle explicite de la déformation de l'image.

3 Protocole expérimental

Cette section décrit les jeux de données utilisés et des détails concernant l'implémentation de notre approche.

3.1 Jeux de données

Images synthétiques. Nous avons généré différents jeux de données synthétiques 3D afin de tester notre méthode à des niveaux de complexité variables. Nous avons ainsi synthétisé, du plus simple au plus complexe, des boules, des ellipsoïdes uniques, des paires d'ellipsoïdes et des pseudo-poumons avec différentes intensités de bruit (Fig. 3). Les pseudo-poumons sont des paires d'ellipsoïdes avec une coupe en arc de cercle qui permet d'imiter le diaphragme. Concernant le jeu de données bruitées, nous appliquons un bruit additif gaussien d'intensité élevée et variable (de 50% à 80%). Dans ces expérimentations les images volumétriques ont une dimension de $128 \times 128 \times 128$ voxels. Afin d'obtenir des variations de volumes et générer des dynamiques apparentées à un mouvement, nous synthétisons 200 objets par jeu de données, dont chacun subit plusieurs dilatations morphologiques à l'aide d'éléments structurants de taille variable et aléatoire. Ensuite, chaque cas est dilaté pour un total de 1000 images. Nous utilisons un split de

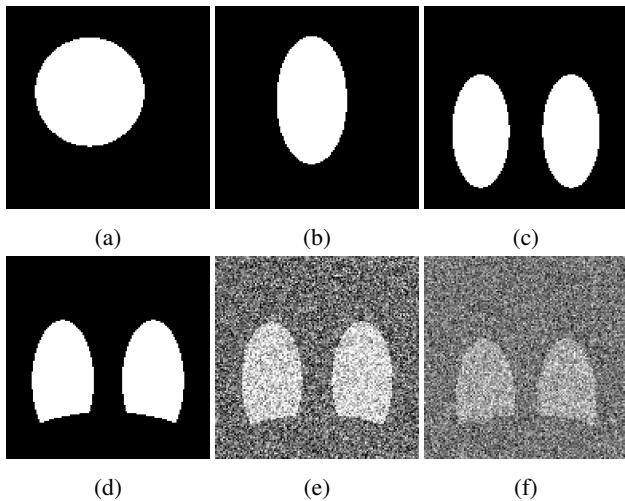


FIGURE 3 – Exemple d'images synthétisées. (a) Boule (b) Ellipsoïde (c) Paire d'ellipsoïdes (d) Pseudo-poumons non-bruités (e) Pseudo-poumons bruités à 50% (f) Pseudo-poumons bruités à 80%

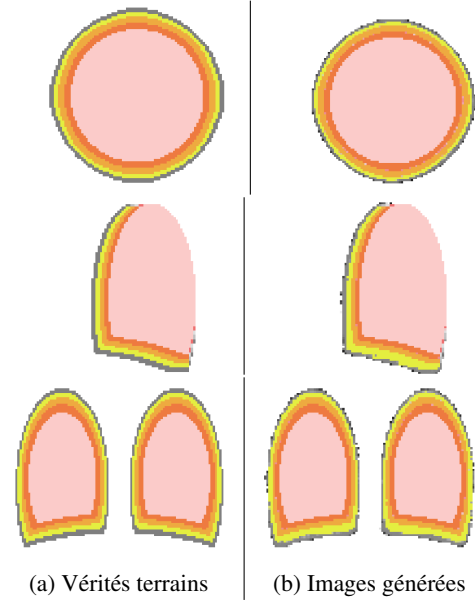


FIGURE 4 – Exemple de résultat de synthèse d'images synthétiques de boules et de pseudo-poumons (plan sagittal et coronal).

4/5 du jeu de données pour l'entraînement du modèle, le reste étant utilisé pour la phase de test.

Données cliniques. Notre base de données clinique est composée de 47 patients dont les acquisitions ont été effectués au service de radiothérapie du CHRU de Brest. Chaque patient possède une à quatre images 4DCT. Toutefois, toutes les images n'ont pas été utilisées dans l'entraînement, certaines étant inutilisables (la moitié de la cage thoracique, artefacts trop importants etc.). Après une sélection manuelle basée sur des critères visuels, nous avons retenu 66 images 4DCT pour nos expérimentations. Les images ont toutes été rééchantillonnées en $256 \times 256 \times 64$ voxels par interpolation linéaire.

Parmi les 66 images 4DCT, le jeu de données pour l'entraînement est composé de 48 images 4DCT de 26 patients qui correspondent chacune à 8 phases respiratoires d'un patient (de 0% à 100% pour la phase inspiratoire et de 75% à 25% pour la phase expiratoire, avec un pas de 25%). Notre jeu de donnée de validation est composé de 18 images 4DCT provenant de 9 patients supplémentaires.

3.2 Implémentation

Nous avons entraîné notre réseau à partir de zéro sur une carte graphique NVIDIA GeForce GTX 1080Ti avec 11Go de mémoire avec une taille de batch de 1. Nos modèles pour les images synthétisées et les données cliniques ont été entraînés durant respectivement 100 et 200 itérations en utilisant l'optimiseur ADAM [17].

Nous utilisons comme conditionnement scalaire en amplitude respiratoire le ratio de volume pulmonaire (*Percentage Volume Difference - PVD*) entre l'image en entrée V_I et la phase voulue V_J , défini par :

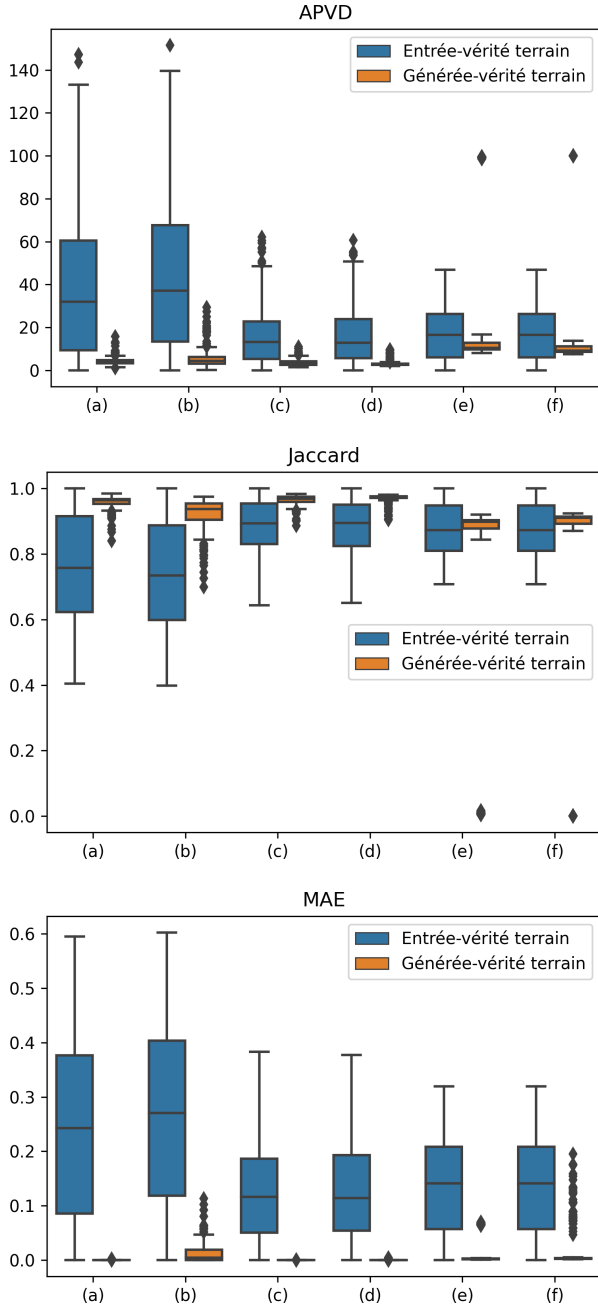


FIGURE 5 – Résultats quantitatifs des jeux de données d'images de synthèse. (a) Boule (b) Ellipsoïde (c) Paire d'ellipsoïdes (d) Pseudo-poumons non-bruités (e) Pseudo-poumons bruités à 50% (f) Pseudo-poumons bruités à 80%

$$PVD = \frac{V_J - V_I}{V_I} \times 100 \quad (7)$$

Dans le cas des pseudo-poumons bruités, nous appliquons un bruit additif gaussien correspondant à une variance σ^2 égale à 0.5^2 et 0.8^2 . Dans le cas des données cliniques, nous avons utilisé une méthode robuste de segmentation des poumons automatique par apprentissage profond [11]

afin d'obtenir les masques des poumons servant ensuite à calculer les volumes correspondant pour le conditionnement scalaire. La méthode de recalage d'image effectuée en sortie de réseau s'appuie sur un fichier de paramètres *elastix* fourni par les auteurs [23].

Pour évaluer quantitativement les résultats, nous avons calculé la similarité entre le volume généré et le volume de la vérité terrain en utilisant 1) le pourcentage absolu de différence de volume (*Absolute Percentage Volume Difference* - APVD), 2) le chevauchement des volumes pulmonaires avec l'indice de Jaccard (intersection sur union) et une métrique pixel par pixel couramment employée en imagerie scanner, l'erreur absolue moyenne (*Mean Absolute Error* - MAE). Idéalement, la valeur de l'indice de Jaccard doit être égale à 1 et à l'inverse les valeurs de MAE et de APVD doivent être égales à 0, signifiant ainsi une similitude parfaite entre les deux échantillons.

4 Résultats

Dans un premier temps, nous évaluons la qualité de la synthèse sur des données synthétiques. Nous validons dans un second temps notre approche sur des données 4DCT réelles.

4.1 Données synthétiques

À l'inférence, nous avons utilisé les cas uniques du jeu de donnée de test afin de générer des exemplaires dilatés par notre méthode qui sont ensuite comparés à leurs vérités terrains respectives (Fig. 4). D'un point de vue qualitatif, nous observons que les images générées ressemblent à celles des vérités terrains en conservant globalement la même forme et la même taille. Les résultats quantitatifs sont présentés sur la Fig. 5 qui montrent les variations entre les images d'entrée et les vérités terrain ainsi que l'erreur calculée entre les images générées et les vérités terrain.

Après évaluation, nous obtenons une erreur relativement faible en moyenne dans l'ensemble des jeux de données avec, par exemple, pour la mesure de APVD, $4.3 \pm 2.0\%$, $5.9 \pm 5.2\%$, $3.7 \pm 1.7\%$ pour les boules, les ellipsoïdes simples et les paires d'ellipsoïdes respectivement. Dans le cas des pseudo-poumons, nous observons une augmentation des erreurs APVD en fonction de l'amplitude du bruit (15.3% pour 50% de bruit et 21.4% pour 80% de bruit) par rapport à la version non-bruités ($3.1 \pm 1.3\%$). Ces comportements sont aussi obtenus par les mesures de Jaccard et de MAE. Des valeurs aberrantes représentant les cas défailants sont d'autant plus nombreux que le bruit est important.

4.2 Données cliniques

Pour les données cliniques 4DCT, nous observons d'un point de vue qualitatif une bonne correspondance entre les phases respiratoires réelles et synthétiques, en restituant globalement la dynamique du mouvement respiratoire avec une précision que nous jugeons satisfaisante (Fig. 6).

Lors de la phase de prédiction, nous appliquons notre mé-

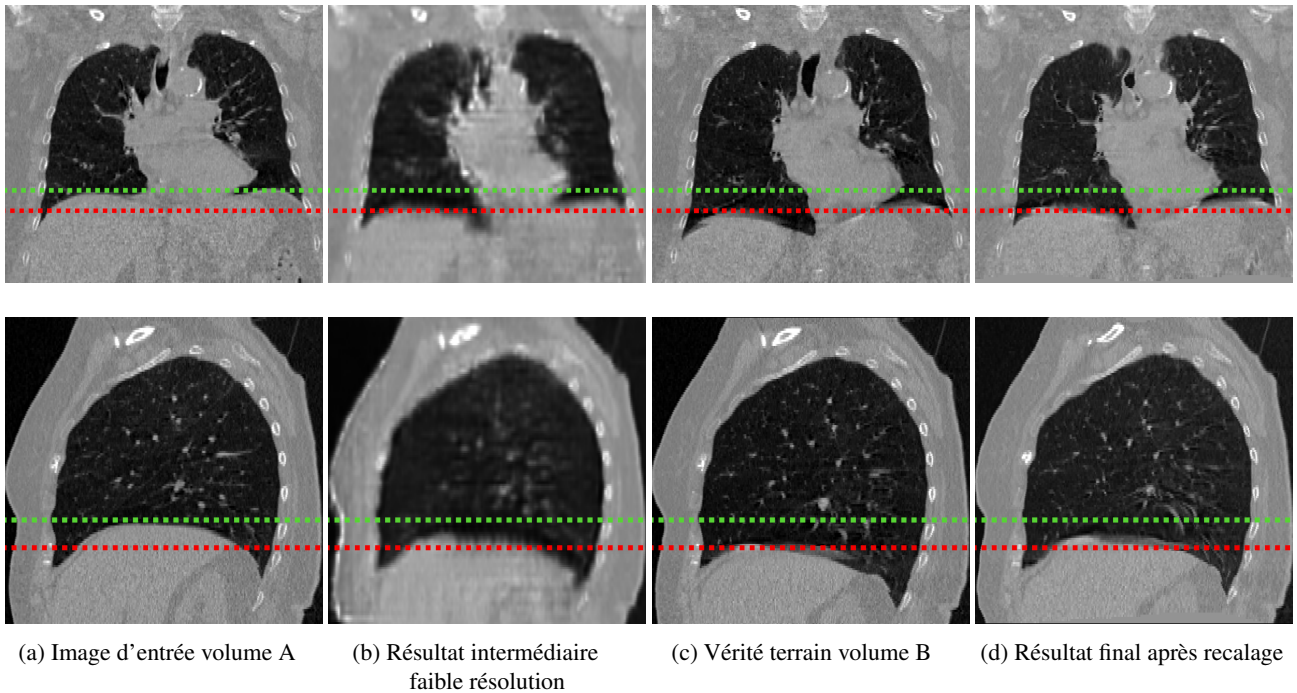


FIGURE 6 – Exemples de résultat de synthèse sur les coupes sagittales et coronales. Les lignes verte et rouge représentent respectivement la position initiale et finale du diaphragme.

thode décrite en section 2. Ainsi, avec une image d'entrée (Fig. 6a), nous générons une image intermédiaire (Fig. 6b) et obtenons après recalage de l'image initiale une image de l'amplitude respiratoire désirée (Fig. 6d). La Fig. 6d montre un exemple représentatif de phase synthétique, où l'on peut observer un déplacement diaphragmatique entre les phases en bon accord avec la vérité terrain (Fig. 6c), visible aussi bien en coupe frontale que sagittale. Les résultats quantitatifs sont présentés dans le Tableau. 1, où nous observons une réduction significative des erreurs de volume pulmonaires inter-phases après synthèse (de 5.7% à 1.2%). Ces résultats suggèrent une bonne capacité du modèle conditionnel proposé à reproduire la phase respiratoire réelle du patient.

5 Conclusion

Dans ce travail, nous montrons pour la première fois qu'une dynamique respiratoire spécifique au patient peut être synthétisée à partir d'images 3DCT statiques en utilisant un GAN I2I doublement conditionné par une image d'entrée et un paramètre scalaire d'amplitude respiratoire. Ces résultats préliminaires constituent une deuxième étape

	APVD
Distance inter-phase	5.57 ± 4.47
Après synthèse	1.23 ± 1.12

Table 1 – Résultats quantitatifs de la situation réelle avec les poumons.

[15] vers une approche complète basée sur l'apprentissage automatique pour la synthèse d'images dynamiques à partir d'images statiques 3DCT. Dans nos travaux futurs, nous nous efforcerons (i). d'améliorer la qualité visuelle des images générées en affinant l'étape de recalage et (ii). d'étendre la méthode de conditionnement pour inclure des informations telles que, par exemple, la phase, l'amplitude d'un signal respiratoire ou des images dynamiques de surface du patient fournies par un système de suivi respiratoire sans contact [19]. Une validation clinique devrait également suivre prochainement.

Références

- [1] Burgos. *Simulation and Synthesis in Medical Imaging : 5th International Workshop, SASHIMI 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 4, 2020, Proceedings*, volume 12417. Springer Nature, 2020.
- [2] Chang et al. A generative adversarial network (gan)-based technique for synthesizing realistic respiratory motion in the extended cardiac-torso (xcat) phantoms. *Physics in Medicine & Biology*, 2021.
- [3] Choi et al. Stargan v2 : Diverse image synthesis for multiple domains. *CoRR*, abs/1912.01865, 2019.
- [4] Cirillo et al. Vox2Vox : 3D-GAN for Brain Tumour Segmentation. *CoRR*, 2020.
- [5] JRS de Koste et al. Renal mobility during uncoached quiet respiration : An analysis of 4DCT

- scans. *International Journal of Radiation Oncology*Biophysics*, 64(3) :799–803, 2006.
- [6] Delaney et al. The role of radiotherapy in cancer treatment : estimating optimal utilization from a review of evidence-based clinical guidelines. *Cancer : Interdisciplinary International Journal of the American Cancer Society*, 104(6) :1129–1137, 2005.
- [7] Fayad et al. Generation of 4-Dimensional CT Images Based on 4-Dimensional PET-Derived Motion Fields. *Journal of Nuclear Medicine*, 54(4) :631–638, 2013.
- [8] Frangi et al. Simulation and synthesis in medical imaging. *IEEE Trans. Med. Imaging*, 37(3) :673–679, 2018.
- [9] Guo et al. A Spatiotemporal Volumetric Interpolation Network for 4D Dynamic Medical Image. In *Proceedings of the IEEE/CVF CVPR*, 2020.
- [10] Hansen et al. Graphregnet : Deep graph regularisation networks on sparse keypoints for dense registration of 3d lung cts. *IEEE Transactions on Medical Imaging*, 2021.
- [11] Hofmanninger et al. Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *European Radiology Experimental*, 4(1), Aug 2020.
- [12] Huang et al. Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization. In *IEEE ICCV*, pages 1510–1519, 2017.
- [13] Huang et al. Multimodal unsupervised image-to-image translation. *CoRR*, abs/1804.04732, 2018.
- [14] Isola et al. Image-to-image translation with conditional adversarial networks. In *IEEE CVPR*, pages 1125–1134, 2017.
- [15] Jaouen et al. 4D respiratory motion synchronized image synthesis from static CT images using GANs. In *IEEE NSS/MIC*, Manchester, United Kingdom, October 2019.
- [16] Kazemini et al. GANs for medical image analysis. *Artificial Intelligence in Medicine*, 109 :101938, 2020.
- [17] Kingma et al. Adam : A method for stochastic optimization, 2017.
- [18] Mori et al. Simulated four-dimensional CT for markerless tumor tracking using a deep learning network with multi-task learning. *Physica Medica*, 80 :151–158, 2020.
- [19] Nazir et al. Kinect v2 surface filtering during gantry motion for radiotherapy applications. *Medical physics*, 45(4) :1400–1407, 2018.
- [20] Park et al. Contrastive learning for unpaired image-to-image translation. In *European Conference on Computer Vision*, 2020.
- [21] Pizzati et al. Comogan : continuous model-guided image-to-image translation, 2021.
- [22] Segars et al. 4d xcat phantom for multimodality imaging research. *Medical physics*, 37(9) :4902–4915, 2010.
- [23] Staring et al. Pulmonary image registration with elastix using a standard intensity-based algorithm. *Medical Image Analysis for the Clinic : A Grand Challenge*, pages 73–79, 2010.
- [24] Viergever et al. A survey of medical image registration—under review, 2016.
- [25] Vinod et al. Radiotherapy treatment for lung cancer : Current status and future directions. *Respirology*, 25 :61–71, 2020.
- [26] Wu et al. Relgan : Multi-domain image-to-image translation via relative attributes, 2019.
- [27] Zhu et al. Toward multimodal image-to-image translation. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 465–476. Curran Associates, Inc., 2017.
- [28] Zhu et al. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020.