



HAL
open science

Lung CT Synthesis Using GANs with Conditional Normalization on Registered Ultrashort Echo-Time MRI

Arthur Longuefosse, Gaël Dournes, Ilyes Benlala, Baudouin Denis de Senneville, François Laurent, Pascal Desbarats, Fabien Baldacci

► To cite this version:

Arthur Longuefosse, Gaël Dournes, Ilyes Benlala, Baudouin Denis de Senneville, François Laurent, et al.. Lung CT Synthesis Using GANs with Conditional Normalization on Registered Ultrashort Echo-Time MRI. ISBI 2023 - 20th IEEE International Symposium on Biomedical Imaging, Apr 2023, Cartagena de Indias, Colombia. pp.1-5, 10.1109/ISBI53787.2023.10230331 . hal-04268682

HAL Id: hal-04268682

<https://hal.science/hal-04268682v1>

Submitted on 11 Nov 2023

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.



Distributed under a Creative Commons Attribution 4.0 International License

LUNG CT SYNTHESIS USING GANS WITH CONDITIONAL NORMALIZATION ON REGISTERED ULTRASHORT ECHO-TIME MRI

Arthur Longuefosse¹ Gaël Dournes² Ilyes Benlala² Baudouin Denis De Senneville³
François Laurent² Pascal Desbarats¹ Fabien Baldacci¹

¹ LaBRI, Université de Bordeaux, Talence, France

² Service d’Imagerie Médicale Radiologie Diagnostique et Thérapeutique, CHU de Bordeaux, France

³ Institut de Mathématiques de Bordeaux, CNRS, Université de Bordeaux, Talence, France

ABSTRACT

In clinical practice, the modality of choice for lung diagnosis is usually computed tomography (CT), which exposes patients to ionizing radiations and could potentially affect patients’ health. Conversely, MR scan is considered safe and non-invasive but seems challenging due to the low proton density of the lungs and respiratory artifacts. Recently, ultrashort echo-time (UTE) MRI has been developed for lung assessment and shows promising results. In this work, we propose generating 2D synthetic CT slices from UTE MR slices, to improve the image quality and interpretability. Lung MR and CT volumes of 110 patients acquired on the same day were registered using an accurate edge-based non-rigid registration method. We trained and compared paired state-of-the-art generative models based on adversarial, feature-matching and perceptual losses, and also evaluated the impact of conditional batch normalization, namely SPADE [17], on image synthesis. Quantitative and qualitative evaluations showed that this approach was able to synthesize CT images that closely approximate ground truth CT images, and also enables the use of algorithms originally designed for real CT.

Index Terms— CT Synthesis, Lung, UTE MRI, Generative Adversarial Networks

1. INTRODUCTION

Lung imaging using MRI is still a major challenge in medical imaging, owing to the low proton content, the rapid decay of lung signal due to susceptibility artifacts and respiratory motions. Still, lung MRI would be a desirable approach to computed tomography (CT) given its lack of radiation exposure, and the possibility to combine both morphological and functional information. Recently, the advent of lung MRI with ultrashort or zero echo-time (UTE/ZTE) appears to have initiated some advances to obtain structural imaging of the lung at high resolution [1, 2]. However, due to differences in lung appearance (cf Figure 1), notably imaging texture, blurring and noise, that would still substantially differ from that of

CT, the UTE/ZTE technique has not been widely adopted for clinical use. Although several statements have been published to encourage its clinical use [3, 4], other studies reported that this MR technique still requires specific MRI training and expertise from radiologists, which would sometimes prevent its clinical adoption [5, 6]. Moreover, several UTE/ZTE sequence schemes have been published, and a current issue with this novel technique remains the standardization of results. We believe that the generation of CT images from MR could solve the main issues of lung imaging, especially by using UTE MR images as input.

Over the recent years, deep learning approaches, notably Generative Adversarial Networks (GAN) [7], have been extensively studied for image synthesis in medical imaging. Previous methods in cross-modality synthesis have used GANs in several regions, such as brain [8, 9], aorta [10], pelvic region [11]. A study has also been performed on the generation of thoracic CTs from Dixon MRI [12], but, although it can be useful for radiation treatment, the results do not allow reconstruction of fine structures inside the lung, such as vessels and bronchi. The vast majority of these methods are based on unpaired training, notably with the development of cycle-consistent adversarial networks (CycleGAN) [13]. This type of network only requires that the backward transformed image is similar to the original source image, which can lead to visually convincing results, at the expense of adding excessive deformations [14]. Methods based on paired data are often neglected, since CT and MR images are acquired on separate scanners, and slight voxel-wise misalignment between MR and CT images may lead to synthesis of blurred images [8, 15]. However, recent methods in multi-modal deformable image registration, based on edge alignment, seems particularly well suited for the generation of paired dataset with images from different modalities [16]. In this work, we propose a complete framework for the generation of thoracic CT, based on ultrashort echo-time MRI. We use an accurate edge-based deformable registration as an input to compare state-of-the-art paired image-translation GANs. We demonstrate the applicability of our method and

the benefits of using the SPADE [17] generator, by evaluating synthetic CTs with quantitative and qualitative metrics. Finally, we illustrate the robustness of the MR-to-CT translation by applying bronchial segmentation algorithms, originally designed for real CT, on synthesized CT.

2. MATERIALS

This study is performed on MR and CT images of 110 patients in the routine follow-up of cystic fibrosis, from 2018 to 2022. Both modalities were acquired on the same day. The CT images were obtained using a Siemens SOMATOM Force and a GE Medical Systems Revolution CT, in end-expiration, with sharp filters. The parameters used were a DLP of 10 mGy.cm and a SAFIRE iterative reconstruction. UTE MR images were acquired using the SpiralVibe sequence on a SIEMENS Aera scanner, with the following parameters : TR/TE/flip angle=4.1ms/0.07ms/5°. Since the slice plane is encoded in Cartesian mode, native acquisition was performed in the coronal plane with field-of-view outside the anterior and posterior chest edges to prevent aliasing. To avoid artifacts due to chest motion, prospective respiratory gating was used. To this end, the user had to check which coil element was the closest to the area of maximum respiratory motions, typically the diaphragm.

3. METHOD

We present here the state-of-the-art deep learning models used in paired image synthesis.

The pix2pix model [18], uses a conditional GAN to learn a mapping from input to output images. It consists of a generator G that tries to minimize adversarial and pixel-wise loss functions, and a discriminator D that tries to maximize adversarial loss function. The adversarial loss is defined as

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

The pix2pix model can produce realistic translations, but is still dependent on the L1 distance, which tends to yield blurry target images [18, 19].

In [20], the authors proposed an improvement of the pix2pix method, namely pix2pixHD, by using a multi-scale discriminator and a robust adversarial learning objective function. This method is no longer dependent on the pixel-wise loss but on a new feature matching loss, defined as

$$\mathcal{L}_{FM}(G, D) = \mathbb{E}_{x,y} \sum_{i=1}^T \frac{1}{N_i} [||D^{(i)}(x, y) - D^{(i)}(x, G(x))||_1] \quad (2)$$

T being the total number of layers, N the number of elements in each layer, and $D^{(i)}$ the i -th layer of D. This loss helps to stabilize the training, since the generator has to produce natural features at multiple scales.

The authors of pix2pixHD also experimented adding a perceptual loss, based on the pre-trained VGG network :

$$\mathcal{L}_{VGG}(x, y) = \sum_{i=1}^T \frac{1}{N_i} [||F^{(i)}(y) - F^{(i)}(G(x))||_1] \quad (3)$$

$F^{(i)}$ being the i -th layer with N_i elements of the VGG network. This loss is closer to perceptual similarity than traditional pixel-wise losses and slightly improves the results. Finally, SPADE [17] focuses on normalization layers, which tend to discard semantic information due to their non-conditional approach. For that purpose, this method introduces a spatially adaptive normalization based on the inputs, that improves the performance and reliability of the generator. The SPADE architecture can be integrated in other models, such as pix2pixHD, to apply additional constraints on the inputs and guide training.

4. EXPERIMENTS AND RESULTS

The aim of this study is to compute and compare synthesized CT volumes based on the state-of-the-art paired methods, and determine if the conditional normalization presented in SPADE improves the performance of the reconstruction. Although 3D GANs allow perception of volumetric and neighborhood spatial information, they involve an excessive computational cost and a reduction of the number of samples in the dataset. Therefore, we choose to train the models on the 2D axial slices of the CT and MR volumes. Unlike most studies in the literature, we keep a voxel size of $0.6 \times 0.6 \times 0.6 \text{ mm}^3$ for both modalities, which allows us to obtain high resolution 2D input and output slices of size 512x512. After inference, the acquired 2D synthetic CT slices can be stacked to reconstruct the synthetic CT volume.

We present here the preprocessing applied to generate the 2D slices, the training architecture and parameters, as well as a quantitative evaluation to compare the methods.

4.1. Preprocessing

To generate paired sets, all volumes are resampled on a common grid with a voxel size of $0.6 \times 0.6 \times 0.6 \text{ mm}^3$. CT images are aligned on MR images using a rigid translation based on a gradient-driven optimization scheme [21], and a deformable registration is applied using the EVOlution algorithm [16, 22]. This method is based on a similarity term that favors edge alignment, and on a diffeomorphic transformation that ensures the preservation of the CT volume topology. We perform a manual verification to detect images acquired at different times of the respiratory activity, that lead to poorly registered volumes, and remove them from the training set. To speed up calculations and guide training, an automatic lung segmentation of the registered CT images is computed using

the U-net R-231 convolutional network proposed in [23]. This segmentation allows to extract the Volume of Interest (VOI) of lungs CT, which also corresponds to the VOI of lungs MR, given that CT is registered on MR. CT intensities are then cropped to $[-1000; 2000]$ Hounsfield unit and rescaled to $[-1; 1]$. Since MR intensity values strongly depend on acquisition parameters, we normalize them using zero mean and unit variance, crop the values to $[-4\sigma; 4\sigma]$ to remove outliers, σ being the standard deviation, and rescale them to $[-1; 1]$ based on minimum and maximum intensities. All axial slices are then either cropped or zero-padded to 512×512 , depending on the CT lung mask size, and saved as 16-bit NumPy arrays. The training set is composed of 82 patients (33002 slices), and the testing set is composed of 28 patients (10725 slices).

4.2. Training

All models are trained using the same procedure and architecture defined in SPADE [17], apart from the dataloader and inference parts, which have been adapted to support 16-bit input and output arrays. We use the Adam optimizer [24], a batch size of 1, a learning rate of 0.0001 and ReLU activation for the generator, and a learning rate of 0.0004 and LeakyReLU activation for the discriminator. The loss function is defined as a combination of GAN, feature-matching and VGG losses. The training process includes 100 epochs, and lasted around 400 hours on a 12GB NVIDIA RTX 2080 Ti. MR to CT inference with a trained model requires around thirty seconds on one single GPU.

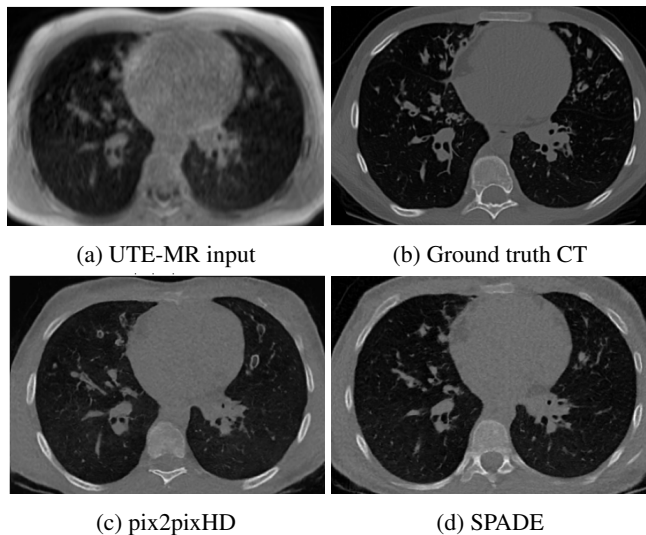


Fig. 1: Comparison between MR axial slice, synthetic CT from pix2pixHD and SPADE models and ground truth CT.

4.3. Quantitative Validation

Fig. 1 shows an example UTE MR input slice, synthesized slices obtained with pix2pixHD and SPADE methods, and the reference CT slice. Both models achieve to produce a high quality CT slice, but differs on some aspects ; the pix2pixHD output presents a noise-free image, an accurate reconstruction of the bones and body, but seems to deviate from the CT at the level of the lung parenchyma, in particular with the appearance of bronchi not visible on the ground-truth CT. Conversely, the SPADE output presents a slightly less sharp image, a less accurate reconstruction of the bones, but a very faithful synthesis of the lung parenchyma, in particular at the level of the vessels and bronchi.

To validate these assumptions, we reconstruct the 3D CT volumes by stacking each 2D slices, and compute standard metrics, namely MSE, cross correlation and SSIM between ground truth CT volumes and synthesized CT volumes. In order to compare only the intensities inside the lung parenchyma, these metrics are constrained on the intersection between the CT lung mask and the generated CT lung masks, such that the results reflect the quality of vessels and bronchi reconstruction (Table 1).

Table 1: Mean squared error (MSE), cross correlation (CC) and structural similarity index (SSIM) between real CT and synthesized CT, based on intensities inside the intersection of lung masks [23].

	MSE	CC	SSIM
pix2pixHD	81.6 ± 22.6	0.860 ± 0.05	0.907 ± 0.02
SPADE	51.3 ± 15.7	0.893 ± 0.04	0.922 ± 0.01

Results show that SPADE, i.e. pix2pixHD method combined with a conditional normalization layer, provides better performances than the pix2pixHD method alone. We can therefore argue that the use of the spatially adaptive normalization helps to ensure content preservation by adding constraints on the input MRI, and avoids the hallucination of anatomic structures, a typical trait of GANs [25].

Regarding bone reconstruction, we assert that since bones in MRI have low intensities and bones in CT have high intensities, using a spatial normalization such as SPADE will add constraints on the MRI input and will eventually lead to a less accurate reconstruction of the synthesized CT bones. From a clinical point of view, the importance seems to remain in the precise reconstruction of structures in the lung parenchyma, such as vessels and bronchi, notably to allow the detection and evaluation of severity of diseases such as bronchiectasis.

5. DISCUSSION

In this study, we demonstrate that paired image-to-image translation GANs can be trained to synthesize a lung CT image from UTE MRI. Unpaired methods, such as Cycle-GAN

[13], have often been employed in the literature. From our experiments, this kind of models can lead to an alteration of fine structures, and artifacts not present in the input image may be added. These observations are in correlation with statements in state-of-the-art [15, 14], and validated our initial assumption of using paired methods.

The quantitative evaluation (Table 1) shows that the correspondence between synthesized and ground truth CT images using the pix2pixHD model combined with SPADE normalization is superior to the pix2pixHD model alone, that can lead to false positives in the generation of fine structures in lung parenchyma. Regarding the assessment of results, we found it more relevant to use standard metrics, such as MSE, SSIM, and cross correlation, constrained within the lung masks, rather than GAN metrics such as FID [26] and KID [27], that have been proven to be insensitive to the global structure of the data distribution, since they are based on extrinsic properties [28].

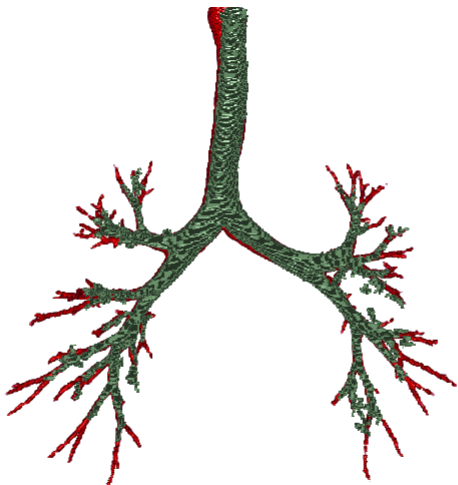


Fig. 2: Airways segmentations from real CT (red) and SPADE (green), using the NaviAirway pipeline [29].

To illustrate the accurate reconstruction of the SPADE model, we compute the airways segmentations of synthesized and ground truth CT using NaviAirway [29], a bronchiole-sensitive airway segmentation pipeline designed for CT data. Results, presented in Figure 2, show that the synthetic CT volume achieves to generate high quality airways ; the generation levels of the bronchial tree are nearly as deep as the ground truth, and present very few false positives.

A qualitative validation based on the correlation coefficient of the bronchial dilatation subscores [30] has also been conducted by a radiologist (Table 2), and illustrates the diagnostic improvement of synthetic CT over UTE MRI. These results are most likely due to the better overview of the lung from the synthetic CT, along with denoising of the initial MR image and enhanced visualization of structures.

Table 2: Correlation of bronchial dilatation subscores between UTE MRI and SPADE synthetic CT with real CT, based on a radiologist with 4 years of experience in thoracic imaging.

	UTE MRI	SPADE
Correlation with real CT	0.70	0.91

Despite these positive results, there is still room for improvement, such as the use of 3D information for training instead of 2D slices, which could lead to a better reconstruction of the CT volume. Future works will aim at a pseudo-3D training, which can alleviate the GPU constraints of traditional 3D training, by using each of the 2D axes (axial, coronal, sagittal) and reconstructing the 3D volume using the information of each axis.

6. CONCLUSION

A complete framework using paired generative adversarial networks is presented to synthesize lung CT volumes with UTE MR volumes as inputs. From a cohort of 110 patients, with both modalities acquired on the same day, we generated paired dataset using an accurate edge-based deformable registration, to compare state-of-the-art image-to-image translation methods. Based on obtained results, we have seen that the pix2pixHD model can provide high quality images at the expense of false positives in fine structures. We demonstrated that the use of a spatially conditioned normalization, namely SPADE [17], can overcome these false positives by adding constraints on the inputs, and delivers a promising high quality generation that could have implications for MR-only radiotherapy treatment. The reconstructed synthetic CT volume enables the use of algorithms originally designed for real CT, such as lung mask segmentation [23] and airways extraction [29]. Future works will aim at adding 3D information in the training process, and also developing a model specific to the constraints of lung synthesis (i.e. preserving morphological and functional information during the CT translation) while extending evaluation by using a segmentation of the bronchi and vessels, and a comparison of quantitative 3D measures.

7. REFERENCES

- [1] G. Dournes, D. Grodzki, J. Macey et al. : Quiet Submillimeter MR Imaging of the Lung Is Feasible with a PETRA Sequence at 1.5 T. *Radiology*. 276. 258-65. (2015)
- [2] G. Dournes, J. Yazbek, W. Benhassen et al. : 3D ultra-short echo-time MRI of the lung using stack-of-spirals and spherical k -Space coverages: Evaluation in healthy volunteers and parenchymal diseases: Lung MRI With 3D UTE Spiral VIBE Sequence. *JMRI*, 48. (2018)

- [3] H. Hatabu, Y. Ohno, W. Gefter et al. : Expanding Applications of Pulmonary MRI in the Clinical Evaluation of Lung Disorders: Fleischner Society Position Paper. *Radiology*, 201138 (2020)
- [4] G. Dournes, L. Walkup, I. Benlala, et al. : The Clinical Use of Lung MRI in Cystic Fibrosis: What, Now, How?. *Chest*. 159. (2020)
- [5] D. Gräfe, R. Anders, F.Prenzel et al. : Pediatric MR lung imaging with 3D ultrashort-TE in free breathing: Are we past the conventional T2 sequence?. *Pediatric pulmonology*, 56(12) (2021)
- [6] K.S. Sodhi, P. Ciet, S. Vasanawala et al. : Practical protocol for lung magnetic resonance imaging and common clinical indications. *Pediatr Radiol* 52, 295–311 (2022)
- [7] I. Goodfellow, J. Pouget-Abadie, Mirza, B. Xu, et al. : Generative adversarial nets. In: *Advances in neural information processing systems*. pp. 2672–2680 (2014)
- [8] J. Wolterink, A. Dinkla, M. Savenije et al. : Deep MR to CT Synthesis using Unpaired Data. (2017)
- [9] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang : Medical Image Synthesis with Context-Aware Generative Adversarial Networks. 10435. 417-425. (2017)
- [10] C. Wang, G. Macnaught, G. Papanastasiou, et al. : Un-supervised Learning for Cross-Domain Medical Image Synthesis Using Deformation Invariant Cycle Consistency Networks: MICCAI (2018)
- [11] Y. Lei, J. Harms, T. Wang et al. : MRI-Only Based Synthetic CT Generation Using Dense Cycle Consistent Generative Adversarial Networks. *Medical Physics*. (2019)
- [12] A. Baydoun, K. Xu, H. Yang, et al. : Dixon-based thorax synthetic CT generation using Generative Adversarial Network. *Intelligence-Based Medicine*. 3-4. (2020)
- [13] J. Zhu, T. Park, P. Isola and A. Efros : Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. 2017 IEEE (ICCV) pp. 2242-2251, (2017)
- [14] B. Cao, Z. Han, W. Nannan et al. : Auto-GAN: Self-Supervised Collaborative Learning for Medical Image Synthesis. *Proceedings of the AAAI Conference on Artificial Intelligence*. 34. (2020)
- [15] S. Kaji, S. Kida : Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging. *Radiol Phys Technol*. 12. (2019)
- [16] B. Denis de Senneville, C. Zachiu, M. Ries, and C. T. W. Moonen : Evolution: an edge-based variational method for non-rigid multi-modal image registration. *Physics in Medicine and Biology*, 61(20):7377 (2016)
- [17] T. Park, M.-Y. Liu, T.-C. Wang, et al : Semantic Image Synthesis with Spatially-Adaptive Normalization. *CVPR*, 2337-2346 (2019)
- [18] P. Isola, J. Zhu, T. Zhou, A. Efros : Image-to-Image Translation with Conditional Adversarial Networks. *CVPR*, 5967-5976. (2017)
- [19] A. Larsen, S. Sønderby, O. Winther : Autoencoding beyond pixels using a learned similarity metric. (2015)
- [20] T. Wang, M. Liu, J. Zhu, et al : High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs. *CVPR*, 8798-8807. (2018)
- [21] M. Irani and P. Anandan : Robust multi-sensor image alignment. In *IEEE Computer Vision, Sixth International Conference*. 959-966 (1998)
- [22] L. Lafitte, C. Zachiu, L. G. W. Kerkmeijer, et al : Accelerating multi-modal image registration using a supervoxel-based variational framework. *Physics in Medicine and Biology* , 63(23):235009, (2018)
- [23] J. Hofmanninger, F. Prayer, J. Pan, et al. : Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem. *European Radiology Experimental*. (2020)
- [24] D. Kingma, J. Ba : Adam: A Method for Stochastic Optimization. *International Conference on Learning Representations*. (2014)
- [25] J. P. Cohen, M. Luck, S. Honari : Distribution Matching Losses Can Hallucinate Features in Medical Image Translation. *MICCAI* (2018)
- [26] M. Heusel, H. Ramsauer, T. Unterthiner, et al. : GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Proceedings of the 31st International Conference on NIPS* 6629–6640 (2017)
- [27] M. Bińkowski, D.J. Sutherland, M. Arbel, A. Gretton : Demystifying MMD GANs. *CLR* 2018 (2018)
- [28] A. Tsitsulin, M. Munkhoeva, D. Mottin, et al : The Shape of Data: Intrinsic Distance for Data Distributions, *ICLR'2020* (2020)
- [29] A. Wang, T. Chi Chun Tam, H. Ming Poon et al. : Navi-Airway: a bronchiole-sensitive deep learning-based airway segmentation pipeline for planning of navigation bronchoscopy. *Transactions on Medical Imaging* (2022)
- [30] M. Bhalla, et al. "Cystic fibrosis: scoring system with thin-section CT". *Radiology* 179.3 : 783-788 (1991).