



HAL
open science

Image and volume conditioning for respiratory motion synthesis using GANs

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

► **To cite this version:**

Yi-Heng Cao, Vincent Jaouen, Vincent Bourbonne, Nicolas Bousson, Ulrike Schick, et al.. Image and volume conditioning for respiratory motion synthesis using GANs. NSS/MIC 2021: IEEE Nuclear science symposium and medical imaging conference, Oct 2021, Virtual, Japan. hal-03658055

HAL Id: hal-03658055

<https://hal.science/hal-03658055v1>

Submitted on 3 May 2022

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.

Image and volume conditioning for respiratory motion synthesis using GANs

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

Abstract—Four-dimensional computed tomography (4DCT) acquisitions are used routinely in lung cancer radiotherapy treatment planning to identify target volumes and safety margins. However, they expose the patient to higher radiation dose compared to static 3DCT acquisitions. In this work, we demonstrate the possibility of generating synthetic 4DCT acquisitions from a 3DCT image following the actual patient’s respiratory amplitude. To this end, we propose a new image-to-image generative adversarial network (GAN) architecture. More specifically, we propose a new scalar injection mechanism based on Adaptive Instance Normalization to condition the generator on the breathing amplitude. Such information can be obtained in practice using external respiratory tracking devices. We show preliminary results on a series of 4DCT images where we compare our synthesized 4DCT to real respiratory phase-gated acquisitions, paving the way for 4DCT-free treatment planning.

I. INTRODUCTION

DYNAMIC four-dimensional computed tomography (4DCT) imaging is used routinely for respiratory motion synchronised image acquisition in radiotherapy treatment planning [1]. However, such acquisitions lead to higher radiation exposure up to six times a standard 3DCT imaging due to longer acquisition times [2]. With the fast progress of deep learning in medical imaging, there is a growing interest for innovative methods based on artificial intelligence that could reduce image-induced radiation [3].

Generative image-to-image (I2I) translation networks such as *pix2pix* [4] have enabled dramatic progress in image synthesis in recent years and have been applied with success in medical imaging for applications such as segmentation, reconstruction or noise reduction [5]. Currently, only a few deep learning methods have however been applied to dynamic medical imaging. Guo et al. [6] recently proposed a spatiotemporal volumetric interpolation network designed for 4D dynamic medical images where a registration and warping networks are learned simultaneously between two dynamic phases. In 2019, we proposed for the first time to consider I2I models to synthesize realistic breathing dynamics in 4DCT imaging [7] using an ensemble of two-dimensional (2D) I2I networks, each synthesizing one respiratory phase-gated frame of the 4DCT acquisition. Mori et al. [8] later used similar concepts to synthesize deformation vector fields representing different typical breathing phases. A major limitation of the former approaches is that the synthesized breathing phases

The authors are with LaTIM, UMR Inserm 1101, Université de Bretagne Occidentale, IMT Atlantique, Brest, France. V. Bourbonne, N. Bousson, U. Schick and J. Bert are also with CHRU Brest University Hospital, Brest, France. Corresponding e-mail: yi-heng.cao@univ-brest.fr.

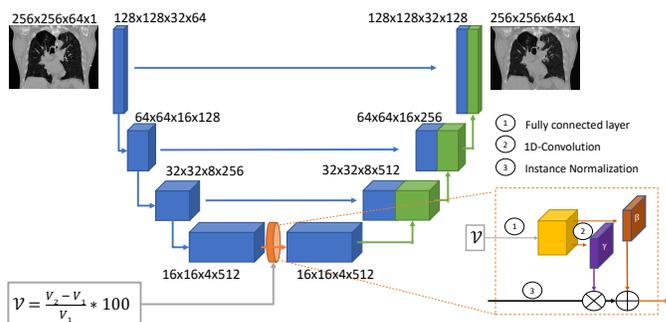


Fig. 1: Generator architecture with proposed scalar conditioning injection layer on respiratory amplitude \mathcal{V} .

do not correspond to the actual patient motion but are rather typical average learnt phases. To the best of our knowledge, the potential of I2I networks for generating dynamic images with realistic conditioning based on the actual respiratory amplitude has not been explored yet, which is the aim of the present work.

In this paper, we show that realistic respiratory motion dynamics can be synthesized from static 3DCT images using a new I2I network architecture doubly conditioned by the image and a scalar value corresponding to the actual respiratory amplitude, thereby mimicking the setup where such information is available from respiratory tracking devices. In our experiments, we present preliminary synthesis results on real 4DCT clinical data where we demonstrate accurate recovery of the actual breathing phase.

II. METHOD

Let $\mathcal{I}(x)$ be a reference breath-hold single helical CT image, where $x \in \mathbb{R}^3$ is the image domain. Our objective is to synthesize a vector-valued phase-gated 4DCT acquisition $\mathcal{J}^* : \mathbb{R}^3 \times \mathbb{R} \rightarrow \mathbb{R}$, where $\mathcal{J}^*(x, g)$ is the image value at location x and gate $g \in \{1, 2, \dots, N_g\}$. To this end, using a deep I2I architecture, we seek to learn an implicit mapping $\varphi_g(x, a)$ doubly conditioned on an input image \mathcal{I} and a (scalar) respiratory amplitude a such that a generated image $\mathcal{I} \circ \varphi_g(x, a) = \mathcal{J}^*(x, a)$ shares similar dynamics with a real phase-gated 4DCT acquisition $\mathcal{J}(x, g)$. We therefore assume there exists a bijection (i.e. one-to-one correspondence) between amplitude a and gate g . In the clinic, the amplitude a can be obtained from e.g. a respiratory tracking belt.

To this end, we adapt an existing single condition GAN architecture based on *pix2pix* [4] in 3D (*vox2vox* [9]). To allow for the model to be doubly conditional (i.e. also in amplitude a), we add a layer at the end of the encoder to condition the

generator by a scalar value (Fig. 1). This layer is a modified Adaptive Instance Normalization (AdaIN, [10]) usually used to condition the generator to a style image in style transfer. Here, AdaIN uses a single scalar value which is the relative change \mathcal{V} between the volume of the input image V_1 and the expected volume V_2 .

The generator G is given an input \mathcal{I} and trained to synthesize a target image $\mathcal{J}^*(x, a)$, while the discriminator D is trained to distinguish real 4DCT image $\mathcal{J}(x, g)$ from $\mathcal{J}^*(x, a)$. The model hence learns to produce specific respiratory motion from 3DCT images and a scalar corresponding to the actual respiratory amplitude.

III. EXPERIMENTS AND RESULTS

We performed experiments by training our model on 48 real 4DCT images divided into 8 breathing phases (0%-to-100% inspiration following by 75%-to-25% expiration by steps of 25%) from 26 patients. Our validation set consisted of 10 additional patients. We used the automatic lung segmentation method in [11] to compute lung volumes as a scalar conditioner.

During testing, original CT images were warped to the network output (Fig. 2b) using a two-stage affine-then-B-spline registration algorithm [12] to explicitly control the image deformation and achieve superior image realism. Fig. 2d shows a representative example synthetic phase, where a diaphragmatic displacement between phases in good agreement with ground truth can be observed (Fig. 2c). To quantitatively assess the results, we calculated similarity between generated and ground truth respiratory phase using 1) the Absolute Percentage Volume Difference (APVD) in the lungs, 2) lung volume overlap with the Jaccard index and a pixel-wise metric in the patient body (Mean Absolute Error - MAE). Corresponding results are shown in Table. 1. For all metrics, a significant improvement was achieved after conditional synthesis. For instance, the APVD between generated and ground truth was reduced from 5.6% to 1.3% on average compared to initial differences between phases, suggesting a good ability of the proposed conditional model in reproducing the actual patient phase.

IV. CONCLUSION

In this work, we showed for the first time that patient-specific respiratory dynamics can be synthesized from static 3DCT images using I2I GANs doubly conditioned on an input image and a scalar parameter (i.e. the lung volume). These preliminary results are a second step [7] towards a complete machine learning-based approach for the synthesis of dynamic images from 3DCT images.

In future works, we will work towards (i). improving the visual quality of the generated images eliminating the registration step and (ii). extending the conditioning method to include

	APVD	MAE	Jaccard index
phase distance	5.57 ± 4.47	59.24 ± 26.30	0.93 ± 0.05
after synthesis	1.25 ± 1.20	56.93 ± 14.72	0.94 ± 0.03

Table 1: Quantitative scores including Absolute Percentage Volume Difference, Mean Absolute Errors and Jaccard index.

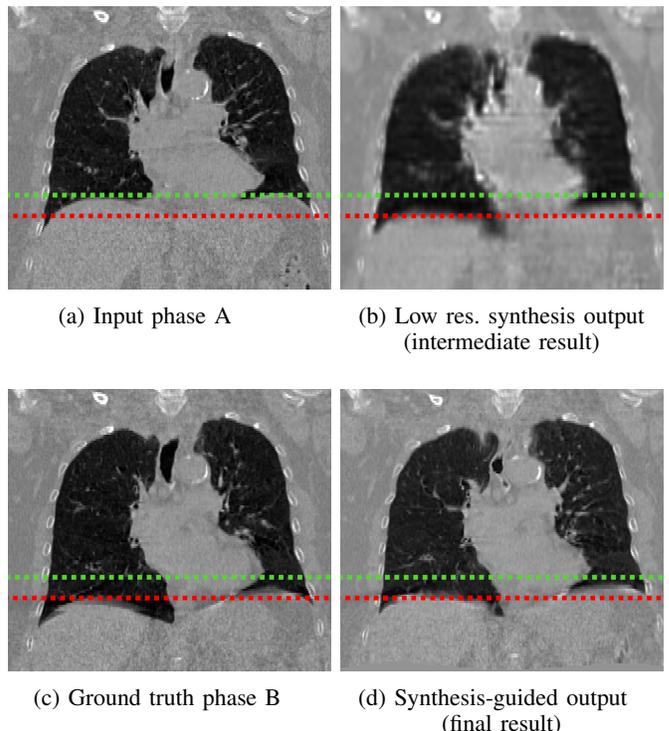


Fig. 2: Example breathing synthesis result. The green and red lines represent the initial and final diaphragm position.

information such as e.g. phase, amplitude of a respiratory signal or patient dynamic surface images provided by contactless respiratory tracking system [13].

REFERENCES

- [1] Fayad *et al.*, “Generation of 4-Dimensional CT Images Based on 4-Dimensional PET-Derived Motion Fields,” *Journal of Nuclear Medicine*, vol. 54, no. 4, pp. 631–638, 2013.
- [2] J. de Koste *et al.*, “Renal mobility during uncoached quiet respiration: An analysis of 4DCT scans,” *International Journal of Radiation Oncology*Biophysics*Physics*, vol. 64, no. 3, pp. 799–803, 2006.
- [3] Frangi *et al.*, “Simulation and synthesis in medical imaging,” *IEEE Trans. Med. Imaging*, vol. 37, no. 3, pp. 673–679, 2018.
- [4] Isola *et al.*, “Image-to-image translation with conditional adversarial networks,” in *IEEE CVPR*, 2017, pp. 1125–1134.
- [5] Kazemina *et al.*, “GANs for medical image analysis,” *Artificial Intelligence in Medicine*, vol. 109, p. 101938, 2020.
- [6] Guo *et al.*, “A Spatiotemporal Volumetric Interpolation Network for 4D Dynamic Medical Image,” in *Proceedings of the IEEE/CVF CVPR*, 2020.
- [7] Jaouen *et al.*, “4D respiratory motion synchronized image synthesis from static CT images using GANs,” in *IEEE NSS/MIC*, Manchester, United Kingdom, Oct. 2019.
- [8] Mori *et al.*, “Simulated four-dimensional CT for markerless tumor tracking using a deep learning network with multi-task learning,” *Physica Medica*, vol. 80, pp. 151–158, 2020.
- [9] Cirillo *et al.*, “Vox2Vox: 3D-GAN for Brain Tumour Segmentation,” *CoRR*, 2020.
- [10] Huang *et al.*, “Arbitrary Style Transfer in Real-Time with Adaptive Instance Normalization,” in *IEEE ICCV*, 2017, pp. 1510–1519.
- [11] Hofmanninger *et al.*, “Automatic lung segmentation in routine imaging is primarily a data diversity problem, not a methodology problem,” *European Radiology Experimental*, vol. 4, no. 1, Aug 2020.
- [12] Staring *et al.*, “Pulmonary image registration with elastix using a standard intensity-based algorithm,” *Medical Image Analysis for the Clinic: A Grand Challenge*, pp. 73–79, 2010.
- [13] Nazir *et al.*, “Kinect v2 surface filtering during gantry motion for radiotherapy applications,” *Medical physics*, vol. 45, no. 4, pp. 1400–1407, 2018.