Original Research Article

Patient-specific three-dimensional image reconstruction from a single X-ray projection using a convolutional neural network for on-line radiotherapy applications

Estelle Loÿen [*], Damien Dasnoy-Sumell, Benoit Macq

Institute of Information and Communication Technologies, Electronics and Applied Mathematics (ICTEAM), UCLouvain, Place de l'Université 1, 1348 Louvain-la-Neuve, Belgium

A B S T R A C T

Background and purpose: Radiotherapy is commonly chosen to treat thoracic and abdominal cancers. However, irradiating mobile tumors accurately is extremely complex due to the organs' breathing-related movements. Different methods have been studied and developed to treat mobile tumors properly. The combination of X-ray projection acquisition and implanted markers is used to locate the tumor in two dimensions (2D) but does not provide three-dimensional (3D) information. The aim of this work is to reconstruct a high-quality 3D computed tomography (3D-CT) image based on a single X-ray projection to locate the tumor in 3D without the need for implanted markers.

Materials and Methods: Nine patients treated for a lung or liver cancer in radiotherapy were studied. For each patient, a data augmentation tool was used to create 500 new 3D-CT images from the planning four-dimensional computed tomography (4D-CT). For each 3D-CT, the corresponding digitally reconstructed radiograph was generated, and the 500 2D images were input into a convolutional neural network that then learned to reconstruct the 3D-CT. The dice score coefficient, normalized root mean squared error and difference between the ground-truth and the predicted 3D-CT images were computed and used as metrics.

Results: Metrics' averages across all patients were 85.5% and 96.2% for the gross target volume, 0.04 and 0.45 Hounsfield unit (HU), respectively.

Conclusions: The proposed method allows reconstruction of a 3D-CT image from a single digitally reconstructed radiograph that could be used in real-time for better tumor localization and improved treatment of mobile tumors without the need for implanted markers.

1. Introduction

Radiotherapy is one of the most widely used treatments in oncology and is prescribed for more than half of all cancer patients, either alone or in combination with surgery and chemotherapy [1]. In radiotherapy, ionizing radiation is used to kill cancer cells. A trade-off must be made between delivering the prescribed dose to the target and not delivering large doses to healthy tissues, which could lead to undesirable effects and induce secondary cancer [2]. Applying radiotherapy to lung and liver cancers is even more challenging as the treatment must consider the respiratory motion. This requires specific strategies in the radiotherapy workflow to ensure adequate target coverage through successive treatment fractions. These strategies are generally classified in two categories.

The first category consists in acquiring a four-dimensional computed tomography (4D-CT) scan prior to the treatment and defining security margins. Safety margins ensure target coverage regardless of the breathing phase, but this method irradiates more the surrounding healthy organs [3]. The breathing motion in the treatment room may also differ significantly from the motion captured in the 4D-CT from time to time [4].

The second category encompasses breathing-synchronized methods that aim to minimize the contribution of the tumor's motion in the computation of the safety margins by monitoring the tumor's position or reducing/regularizing its motion amplitude during breathing. These methods gather abdominal compression [5], audio coaching [6], mechanically assisted ventilation [7] and respiratory gating [8]. Tumor

monitoring in these techniques is based on external surrogates of the internal motion to avoid the use of invasive procedures (the placement of markers pinpoints the tumor position with greater accuracy but involves surgery before the treatment [9]). This approach requires a stable correlation between the internal tumor motion and its external surrogate, which is usually not the case when changes occur in the patient's breathing movement.

Image-guided radiation therapy (IGRT) incorporates imaging techniques during each treatment session. By adding detailed images, it ensures that the radiation is narrowly focused on the target. A broad range of IGRT is now available [10]. X-ray projections are commonly acquired to estimate the tumor's position, but their use often requires implanted markers to identify the tumor volume correctly and make it visible on the X-ray projection [11]. Another disadvantage of this method is that it does not provide 3D information.

All these methods result in a small reduction in the safety margins, while adapting the treatment in 3D and in real-time will lead to a big reduction in the motion margins thanks to precise tracking of the 3D anatomical structures. To achieve this, the real-time positions of the target and surrounding organs must be known throughout treatment delivery. Most of the radiotherapy treatment rooms are equipped with 2D fluoroscopy to validate the patient positioning before treatment, we propose to rely on this equipment to estimate the related 3D information.

Many studies that reconstruct a 3D volume from a 2D X-ray projection have already been performed. Different fields of application in the biomedical sector have been explored: Henzler et al. investigated how to reconstruct 3D volumes from 2D cranial x-rays by applying deep learning [12], while Liang et al. developed a new model architecture to reconstruct a tooth in 3D from a single panoramic radiograph [13]. Montaya et al. in [14], as well as Ying et al. in [15], demonstrated that it was possible to reconstruct a 3D-CT image from biplanar X-ray projections using a neural network, and Shen et al. used a neural network to reconstruct a 3D image from a single projection view [16].

In this context, the aim of the work described in this article was to use the 2D information available in the treatment rooms to obtain 3D information. To that end, we use a convolutional neural network that reconstructs a high-quality 3D-CT image based on a single X-ray

projection. This image, predicted in real-time, can then be used by a real-time segmentation method [17] in order to know the tumor and surrounding organs' positions at the moment of acquisition. This process would make it possible to locate the tumor and neighboring structures accurately in 3D during the treatment without requiring implanted markers.

2. Materials and methods

Fig. 1 summarizes the proposed method's workflow. The different steps of the process are detailed in the following sub-sections.

2.1. Dataset generation

The data used in this work come from nine patients who were treated for lung or liver cancer at *Cliniques universitaires Saint-Luc* in Brussels between 2010 and 2015. This retrospective study was approved by the Hospital Research Ethics Committee (B403201628906). Table 1 shows patients information (tumor size and location, and its motion in the different sets). A planning 4D-CT composed of 10 breathing phases evenly spread over the respiratory cycle was acquired for each patient prior to treatment delivery. The dimensions of each 3D-CT image were $512 \times 512 \times 173$, and the voxel size was 1 mm^2 in plane with a slice thickness of 2 mm. The Mid-Position (MidP)-CT image, defined as the local mean position in the respiratory cycle, was computed using the average of all velocity fields obtained by non-rigid registration between the 4D-CT phases [18]. On the MidP-CT image, the gross target volume (GTV) and surrounding organs at risk were delineated manually by an experienced radiation oncologist.

As training a neural network requires a lot of data, it was necessary to generate new 3D-CT images. To do so, we consider a polar coordinate system (r, n) related to a breathing cycle, whose origin is the MidP-CT image and where n are the periodic phases. In this system, we know the deformation fields associated to the 10 breathing phases of the 4D-CT which are $F(1, N)$, with $N \in \{0, 0.1 \ldots, 0.9\}$. Then, to generate the breathing phase n at a normalized distance r of the MidP-CT, we compute the deformation field $F(r, n)$ using a linear interpolation between the two closest discrete breathing phases plus a scaling:
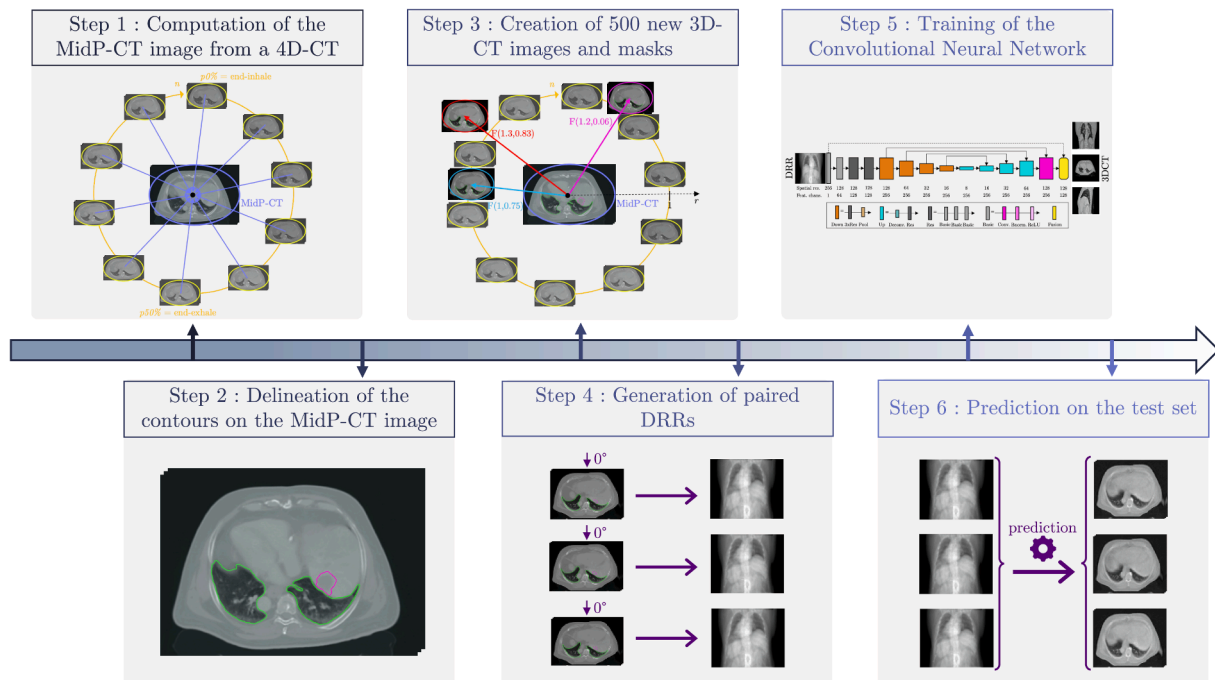

Fig. 1. Overview of the proposed method's workflow.

Table 1

Patient characteristics. MR_{4D-CT}, $MR_{TrainSet}$ and $MR_{TestSet}$ stand for the motion range in 3D of the GTV's centroid in the 4D-CT, training set and test set, respectively. The motion range is defined as the Euclidean distance between the two most distant positions.

Patient ID	Tumor location	GTV size [cm³]	MR_{4D-CT} [mm]	$MR_{TrainSet}$ [mm]	$MR_{TestSet}$ [mm]
Patient 1	Right upper lobe of lung	137.1	11.1	17.2	17.9
Patient 2	Right upper lobe of lung	17.2	9.9	9.7	12.6
Patient 3	Right middle lobe of lung	153.8	24.4	32.4	34.7
Patient 4	Left upper lobe of lung	13.8	14.5	15.2	18.5
Patient 5	Left upper lobe of lung	315.1	9.7	10.1	11.4
Patient 6	Left upper lobe of lung	67.2	11.6	15.2	16.1
Patient 7	Right lobe of liver	28.6	15.1	18.7	26.4
Patient 8	Right lobe of liver	80.4	27.1	29.9	30.8
Patient 9	Left lobe of liver	22.5	24.1	32.3	34.8

$$F(r,n) = [F(1,N) + (F(1,N+0.1) - F(1,N)) \cdot 10 \cdot (n-N)] \cdot r \qquad (1)$$

where $N \leqslant n \leqslant N + 0.1$. Using this method, based on a previous work of our team [19] and developed in [20], we can generate slightly different 3D-CT images, spread around the ten original phases of the 4D-CT, for every patient. The training set was composed of 500 images where n was a uniform random draw between 0 and 1, and r a random sample from a normal distribution $\mathcal{N}(1, 0.25)$ truncated between 0.4 and 1.1. A digitally reconstructed radiograph (DRR) was generated from each of these images using the Beer–Lambert absorption-only model (implemented in the TomoPy Python library [21]) and a projection angle of $0°$ along the anterior-posterior axis. The projection geometry was a 1440×1440 image with a pixel size of 0.296×0.296 mm². The source-to-origin and source-to-detector distances were 1000 mm and 1500 mm. Each patient's training dataset was made up of 500 pairs containing the created 3D-CT image and the associated DRR. An independent test set composed of 100 3D-CT/DRR pairs was also created for each patient. For each image of the test set, the masks of the GTV, lungs and heart were also generated by deforming the MidP-CT image's 3D binary masks. The difference between the test and training sets comes from the normalized distance r used to generate the 3D-CT image. In the case of the training set, r was a random sample from a normal distribution $\mathcal{N}(1, 0.25)$ truncated between 0.4 and 1.1, while r was a random sample from a normal distribution $\mathcal{N}(1, 0.5)$ truncated between 0.8 and 1.5 for the test set. This means that deeper breathing situations were present in the test set than in the training set. All breathing phases were used in both cases.

2.2. Patient-specific deep learning model for 3D-CT reconstruction

The network used for the 3D-CT reconstruction process is a convolutional neural network (CNN) that learns the mapping between a 2D image and a 3D volume. This network was proposed by Henzler et al. in [12] and we have tuned the different hyper-parameters for our challenge. The overall structure of this network is an encoder-decoder with skip connections. The goal of the encoder is to condense the information contained in the training data into a low-dimensional representation, which the decoder then takes as input to predict the output [22]. The input of the network is a DRR of size 256×256, while the output consists of a $128 \times 128 \times 128$ 3D-CT image. The details of the training dataset, namely 3D-CT/DRR pairs, are explained in Section 2.1. The network training was patient-specific, a new network is trained independently for each patient. The same training strategy and hyper-

parameters were used for all patients. The Adam optimizer was used to train the network with an initial learning rate of 10^{-3} and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The model was trained for a total of 300 epochs using a mini-batch size of 16 on a NVIDIA RTX 6000, which brought the training time down to roughly 8 h. Then, it takes about $50\,\mu s$ to predict the output from a new input.

2.3. Performance evaluation

In order to evaluate the performance of the proposed method, 100 3D-CT images independent of the training set were created for each patient. These 3D-CT images are called the *ground truth (GT) 3D-CT images* in the rest of the paper. 100 DRRs were generated from these images to form the test set. The trained network was used on these radiographs to predict the corresponding 3D-CT images, called the *predicted (P) 3D-CT images*. The predicted 3D-CT images were compared with the ground truth 3D-CT images to evaluate the performance of the model using several metrics.

Dice similarity coefficient (DSC) is a common overlap-based metric used to measure the performance of a segmentation algorithm, and is defined by:

$$DSC = \frac{2|A \cap B|}{|A| + |B|} \cdot 100 \quad [\%] \qquad (2)$$

where A and B are the sets containing the matrix indices of both binary masks A and B. In this work, the DSC was computed between a 3D binary mask in the ground-truth 3D-CT image and the corresponding mask in the predicted 3D-CT image to evaluate the quality of the predicted 3D-CT image in terms of anatomical structure positions. The 3D binary masks of a predicted 3D-CT image were obtained by computing the Morphons non-rigid registration [23], then applying the resulting deformation fields to deform the masks on the predicted image. This was done between this predicted image and either the ground-truth 3D-CT image (*GT-based*), or the MidP-CT image (*MidP-based*). Using the ground-truth 3D-CT image for this part serves as a post-training quality evaluation, to evaluate if a state-of-the-art registration algorithm sees a difference between the ground-truth and the predicted images. Using the MidP-CT image simulates how it could be used to evaluate the quality of the predicted images after each treatment fraction as the ground-truth 3D-CT images are not available during a treatment. For both versions, the DSC was computed for the same 50 images of the 100 items constituting the test set, for each organ and each patient. In either case, this metric was an evaluation tool and not part of the real-time process as the computation time of the Morphons is about 150 s. As a complement to this analysis, the Euclidean distance was computed (further details in Appendix A. Supplementary data).

Normalized root mean squared error (NRMSE) was computed between two images A and B, and is defined by:

$$NRMSE = \frac{\sqrt{\frac{\sum_{a=1}^{n}(A_a - B_a)^2}{n}}}{A_{max} - A_{min}} \qquad (3)$$

where X_a is the voxel a in the image X. A_{max} and A_{min} stand for the maximum and minimum in image A, the ground-truth 3D-CT image. The NRMSE was computed between the latter and the corresponding predicted 3D-CT image. This was repeated for all images in the test set.

Difference was computed between a ground-truth 3D-CT image and the corresponding predicted 3D-CT image, and the mean and median of the difference were studied, as well as quantifying the percentage of the absolute value of the difference below a certain threshold to evaluate the proportion of the image that was correctly reconstructed.

3. Results

3.1. Dice similarity coefficient

The results of the DSC analysis for both *GT-based* and *MidP-based* versions are summarized in Table 2. For the *GT-based* version, the mean, the median and the 95th percentile of the DSC vary respectively from 93.2% to 99.8%, from 93.2% to 99.9%, and from 95.1% to 99.9% for the GTV; from 96.3% to 99.8%, from 96.5% to 99.9%, and from 96.8% to 99.9% for both lungs; from 93.5% to 99.8%, from 94.3% to 99.8%, and from 95.1% to 99.9% for the heart. While, for the *MidP-based* version, the mean, the median and the 95th percentile of this metric vary respectively from 76.7% to 90.6%, from 77.6% to 90.8%, and from 82.7% to 93.4% for the GTV; from 90.9% to 97.3%, from 93.4% to 97.1%, and from 96.1% to 98.3% for both lungs; from 78.1% to 90.1%, from 79.2% to 89.9%, and from 81.5% to 91.7% for the heart.

The DSC results of the *MidP-based* version are lower than those of *GT-based*, but still over 75%. As the same 50 images were used for both, the difference might be due to the approximations in the deformations and re-binarization of the masks, that probably have a higher impact with deformations over multiple voxels, but this was not quantified.

3.2. Normalized root mean squared error

The results of the NRMSE analysis are displayed in Fig. 2. The mean of this metric is lower for Patients 5, 2, 6 and 1 who have smaller motions in the test set (from 0.032 to 0.039) than the mean obtained for Patients 7, 8, 3 and 9 (from 0.047 to 0.051) who have larger motions. This is also observed for the median and the 95th percentile, which range respectively from 0.032 to 0.038, and from 0.039 to 0.045 for the first batch of patients, while they are respectively between 0.045 and 0.052, and between 0.051 and 0.059 for the second group of patients. This analysis also shows that the breathing phases have no impact on the reconstruction process as there are uniformly distributed along the NRMSE values range.
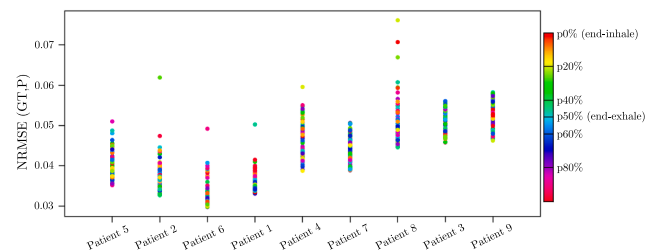
3.3. Difference

The results of the difference analysis are summarized in Table 3. The mean of the difference between a ground-truth 3D-CT image and the corresponding predicted 3D-CT image ranges from −1.32 Hounsfield unit (HU) to 2.24 HU, with an average over all patients of 0.45 HU. The median of this metric is between −0.26 HU and 1.93 HU, with an average over all patients of 0.24 HU. Depending on the patient, 25.1% to 39.8% of the image volume has an absolute value of the difference lower than 5 HU, 69.9% to 81.9% below 25 HU, and 88.6% to 94.6% less than 50 HU. In summary, the difference between the ground-truth and the predicted images is very small, with about 91% of the image volume having an absolute value of the difference smaller than 50 HU, which represents 1.25% of the range of possible values, since the scale of a 3D-

Fig. 2. Results of the NRMSE analysis. The NRMSE was computed between the ground-truth 3D-CT image and the corresponding predicted 3D-CT image for each test set data. The color of a dot represents the breathing phase at which the ground-truth 3D-CT image was created. Patients are sorted by increasing motion range in the test set.

Table 3
Results of the difference analysis. $V_{<5HU}$, $V_{<25HU}$ and $V_{<50HU}$ stand for the percentage of the 3D-CT image's volume having an absolute value of the difference below 5 HU, 25 HU and 50 HU.

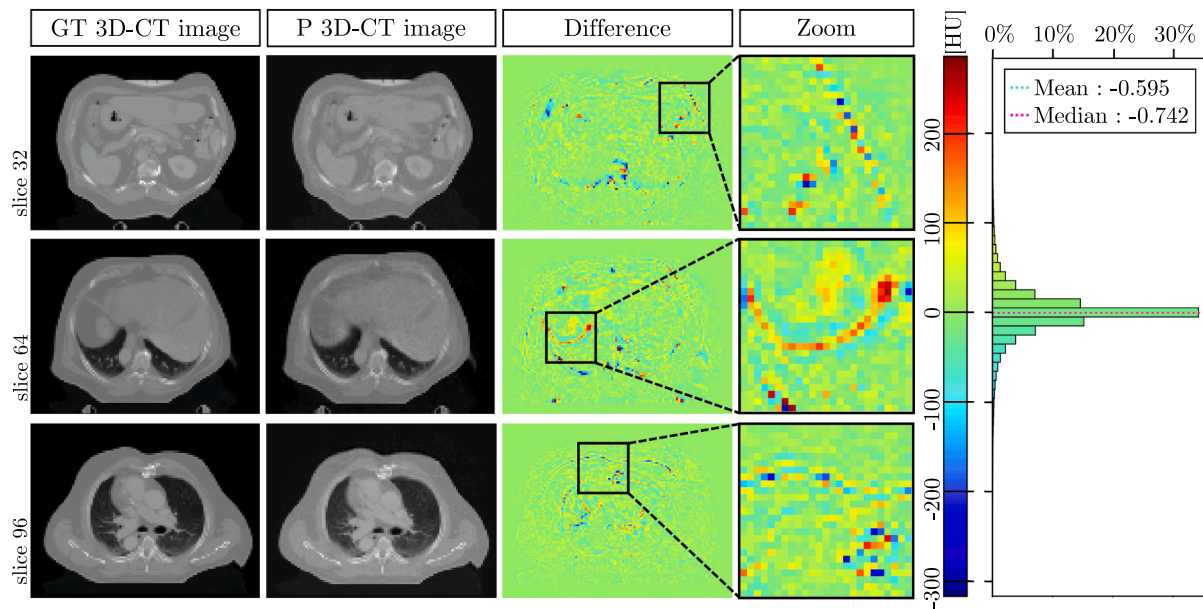| Patient ID	Mean [HU]	Median [HU]	$V_{<5HU}$ [%]	$V_{<25HU}$ [%]	$V_{<50HU}$ [%]
Patient 1	0.36	−0.02	25.4	74.1	91.1
Patient 2	0.31	−0.13	34.5	80.1	93.7
Patient 3	0.46	−0.26	31.8	80.7	94.6
Patient 4	0.51	0.04	39.8	81.9	94.2
Patient 5	0.65	0.08	29.9	75.1	91.5
Patient 6	0.53	−0.16	29.7	76.8	91.9
Patient 7	0.37	0.56	32.4	75.9	88.8
Patient 8	−1.32	0.09	27.1	74.4	89.9
Patient 9	2.24	1.93	25.1	69.9	88.6

CT image typically runs from −1000 HU for air to 3000 HU for dense bone [24].

A representative example (whose results are: DSC_{GT} (GTV) = 98.5%, DSC_{MidP} (GTV) = 88.6%, NRMSE = 0.053, mean of the difference = −1.73 HU and $V_{<25HU}$ = 80.3%) of the results obtained using the proposed method can be seen in Fig. 3. For a human eye, the predicted 3D-CT image looks pretty close in terms of anatomical structures. The zoom shows that a red pixel (difference ≈ 200 HU) is commonly adjacent to a blue pixel (difference ≈ −200 HU) or surrounded by two turquoise pixel (difference ≈ −100 HU). This phenomenon is usually observed at tissue borders. Looking at the histogram, one sees that there are few voxels with a significant difference and over 30% of the voxels have a difference between −5 HU and 5 HU.

4. Discussion

In this paper, it has been showed that the proposed CNN-based methodology (which requires a patient-specific training) allows to reconstruct a high-quality 3D-CT image from a single digitally reconstructed radiograph.

Table 2
Results of the DSC analysis for both *GT-based* and *MidP-based* versions. DSC_{GT} and DSC_{MidP} stand for the mean of the DSC over the 50 images taken from the test set for the *GT-based* version and *MidP-based* version, respectively. Patient 5's lungs and heart were not delineated.

Patient ID	GTV		Lung R		Lung L		Heart	
	DSC_{GT} [%]	DSC_{MidP} [%]	DSC_{GT} [%]	DSC_{MidP} [%]	DSC_{GT} [%]	DSC_{MidP} [%]	DSC_{GT} [%]	DSC_{MidP} [%]
Patient 1	94.1	89.4	98.4	94.9	97.5	93.3	99.5	83.1
Patient 2	93.2	89.1	99.2	96.4	97.4	97.3	99.8	85.8
Patient 3	99.8	81.3	99.3	96.7	98.9	95.6	99.2	80.8
Patient 4	92.5	87.9	98.8	95.6	98.7	93.2	98.9	90.1
Patient 5	96.4	90.4	NA	NA	NA	NA	NA	NA
Patient 6	97.7	90.6	99.8	95.6	99.7	93.4	99.8	89.9
Patient 7	93.3	78.2	97.2	92.8	96.3	90.9	93.5	78.1
Patient 8	99.3	86.3	98.8	94.6	98.7	95.1	99.4	83.8
Patient 9	99.2	76.7	99.4	93.3	99.1	94.5	96.3	80.3

Fig. 3. Visualization of three slices of the ground-truth 3D-CT image of one patient compared with the corresponding slices of the predicted 3D-CT image, as well as the results of the difference analysis and a zoom of the boxed area. On the right of the color bar is the histogram of the difference concatenated for all patients and the 100 images of the nine test sets.

The dice values computed between the masks of the predicted 3D-CT image and the corresponding ground-truth 3D-CT are all greater than 75%, which is reliable. If we compare our results of the *MidP-based* version (Table 2) for lungs and heart (94.6% and 83.9%) to previous works [25–27], whose goal was to segment organs at risk in lung cancer utilizing deep learning algorithms, (best in [27]: 97.5% and 92.5%), lungs have similar results to the literature and the heart has a higher difference. However, our results should be taken in hindsight, given that the masks in the predicted image are defined as the manually segmented masks on the MidP-CT image deformed using the deformation fields obtained by the Morphons registration between both images.

The mean of the difference between the ground truth image and the predicted image is small for each patient, with an average value of 0.45 HU over all patients. Comparing these results (Fig. 3) with those obtained by [16] when they use only 1 view, the quality of our reconstructed image is similar to their own. Their method also performs less at tissue borders. However, there is no scale or numerical value in their difference analysis, so it is not clear that the difference values are similar.

One limitation of this study is that the CNN was trained using training sets composed of 3D-CT images created from deformations of a planning 4D-CT acquired prior to the treatment and paired DRRs generated using the Beer–Lambert absorption-only model. This method supposes that inter-fraction variations such as tumor shrinking, tumor baseline shift and stomach and bladder fillings are not included in the training set. A next step of this work is to evaluate whether the network must be retrained for each fraction or whether these variations are negligible in the reconstruction process. Another possibility to counteract this limitation is to improve the data augmentation tool and incorporate inter-fraction changes in the training set.

An additional potential purpose of the predicted 3D-CT image would be to use it to compute the dose delivered during the treatment (either on-line or inter-fraction). To this end, the voxel value representing tissue density is a crucial piece of information to have the dose delivered at the right place. This paper shows that, for the human eye, the predicted 3D-CT image is really close to the ground-truth 3D-CT image but the results of the difference should be discussed further and it will be necessary to assess whether the maximum of the difference is located on the beam's

path or the difference, no matter how small, has too great an impact on the computed dose. Furthermore, in order to get a clinically usable dose, the standard resolution of a 3D-CT scan would be needed. Therefore, the predicted 3D-CT image should be oversampled to get the desired resolution.

In conclusion, this study presents a method that allows reconstruction of a 3D-CT image from a single DRR. This method relies on a data augmentation algorithm and on a patient-specific training of a CNN. However, the study still needs to integrate inter-fractions changes and adjust the image resolution to confirm the potential clinical use of the method.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.phro.2023.100444.

References

[1] Baskar R, Lee K, Yeo R, Yeoh K. Cancer and Radiation Therapy: Current Advances and Future Directions. Int J Med Sci 2012;9:193–9. https://doi.org/10.7150/ijms.3635.

[2] Warkentin B, Stavrev P, Stavreva N, Field C, Fallone B. A TCP-NTCP estimation module using DVHs and known radiobiological models and parameter sets. J Appl Clin Med Phys 2004;5:50–63. https://doi.org/10.1120/jacmp.v5i1.1970.

[3] Rietzel E, Bert C. Respiratory motion management in particle therapy. Med Phys 2010;37:449–60. https://doi.org/10.1118/1.3250856.

[4] Dhont J, Vandemeulebroucke J, Burghelea M, Poels K, Depuydt T, Van Den Begin R, et al. The long- and short-term variability of breathing induced tumor motion in lung and liver over the course of a radiotherapy treatment. Radiother Oncol 2018;126:339–46. https://doi.org/10.1016/j.radonc.2017.09.001.

[5] Piippo-Huotari O, Norrman E, Anderzén-Carlsson A, Geijer H. New patient-controlled abdominal compression method in radiography: radiation dose and image quality. Acta Radiol Open 2018;7:1–7. https://doi.org/10.1177/2058460118772863.

[6] Nakamura M, Narita Y, Matsuo Y, Narabayashi M, Nakata M, Sawada A, et al. Effect of audio coaching on correlation of abdominal displacement with lung tumor motion. Int J Radiat Oncol Biol Phys 2009;75:558–63. https://doi.org/10.1016/j.ijrobp.2008.11.070.

[7] Van Ooteghem G, Dasnoy-Sumell D, Lee JA, Geets X. Mechanically-assisted and non-invasive ventilation for radiation therapy: A safe technique to regularize and modulate internal tumour motion. Radiother Oncol 2019;141:283–91. https://doi.org/10.1016/j.radonc.2019.09.021.

[8] Muirhead R, Featherstone C, Duton A, Moore K, McNee S. The potential benefit of respiratory gated radiotherapy (RGRT) in non-small cell lung cancer. Radiother Oncol 2010;95:172–7. https://doi.org/10.1016/j.radonc.2010.02.002.

[9] Hirai R, Watanabe W, Sakata Y, Tanizawa A. Real-time linear fiducial marker tracking in respiratory-gated radiotherapy for hepatocellular carcinoma. Int J Radiat Oncol Biol Phys 2019;105:E750–1. https://doi.org/10.1016/j.ijrobp.2019.06.769.

[10] Ren XC, Liu YE, Li J, Lin Q. Progress in image-guided radiotherapy for the treatment of non-small cell lung cancer. World J Radiol 2019;11:46–54. https://doi.org/10.4329/wjr.v11.i3.46.

[11] Soete G, Verellen D, Michielsen D, Vinh-Hung V, Van de Steene J, Van den Berge D, et al. Clinical use of stereoscopic X-ray positioning of patients treated with conformal radiotherapy for prostate cancer. Int J Radiat Oncol Biol Phys 2002;54:948–52. https://doi.org/10.1016/S0360-3016(02)03027-4.

[12] Henzler P, Rasche V, Ropinski T, Ritschel T. Single-image Tomography: 3D Volumes from 2D Cranial X-Rays. arXiv 2017. https://doi.org/10.48550/ARXIV.1710.04867.

[13] Liang Y, Song W, Yang J, Qiu L, Wang K, He L. X2Teeth: 3D Teeth Reconstruction from a Single Panoramic Radiograph. arXiv 2021. https://doi.org/10.48550/arXiv.2108.13004.

[14] Montoya JC, Zhang C, Li Y, Li K, Chen GH. Reconstruction of three-dimensional tomographic patient models for radiation dose modulation in CT from two scout views using deep learning. Med Phys 2021;49:1–16. https://doi.org/10.1002/mp.15414.

[15] Ying X, Guo H, Ma K, Wu J, Weng Z, Zheng Y. X2CT-GAN: Reconstructing CT From Biplanar X-Rays With Generative Adversarial Networks. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019. p. 10611–20. doi: 10.1109/CVPR.2019.01087.

[16] Shen L, Zhao W, Xing L. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. Nat Biomed Eng 2019;3:880–8. https://doi.org/10.1038/s41551-019-0466-4.

[17] Zhou S, Xu X, Bai J, Bragin M. Combining multi-view ensemble and surrogate lagrangian relaxation for real-time 3D biomedical image segmentation on the edge. Neurocomputing 2022;512:466–81. https://doi.org/10.1016/j.neucom.2022.09.039.

[18] Wolthaus J, Sonke J, van Herk M, Damen E. Reconstruction of a time-averaged midposition CT scan for radiotherapy planning of lung cancer patients using deformable registration. J Appl Clin Med Phys 2008;35:3998–4011. https://doi.org/10.1118/1.2966347.

[19] Dasnoy-Sumell D, Aspeel A, Souris K, Macq B. Locally tuned deformation fields combination for 2D cine-MRI-based driving of 3D motion models. Phys Med 2022;94:8–16. https://doi.org/10.1016/j.ejmp.2021.12.010.

[20] Wuyckens S, Dasnoy D, Janssens G, Hamaide V, Huet M, Loÿen E, et al. OpenTPS – Open-source treatment planning system for research in proton therapy. arXiv 2023. https://doi.org/10.48550/arXiv.2303.00365.

[21] Gürsoy D, De Carlo F, Xiao X, Jacobsen C. TomoPy: a framework for the analysis of synchrotron tomographic data. J Synchrotron Radiat 2014;21:1188–93. https://doi.org/10.1107/S1600577514013939.

[22] Minaee S, Boykov Y, Porikli F, Plaza A, Kehtarnavaz N, Terzopoulos D. Image Segmentation Using Deep Learning: A Survey. IEEE Trans Pattern Anal Mach Intell 2022;44:3523–42. https://doi.org/10.1109/TPAMI.2021.3059968.

[23] Janssens G, Jacques L, Orban de Xivry J, Geets X, Macq B. Diffeomorphic registration of images with variable contrast enhancement. Int J Biomed Imaging 2011, 2011,:1–16. https://doi.org/10.1155/2011/891585.

[24] Bibb R, Eggbeer D, Paterson A. 2 - Medical imaging. In: Medical Modelling (Second Edition) Woodhead Publishing; 2015. p. 7-34. https://doi.org/10.1016/B978-1-78242-300-3.00002-0.

[25] Zhu J, Zhang J, Qiu B, Liu Y, Liu X, Chen L. Comparison of the automatic segmentation of multiple organs at risk in CT images of lung cancer between deep convolutional neural network-based and atlas-based techniques. Acta Oncol 2019;58:257–64. https://doi.org/10.1080/0284186X.2018.1529421.

[26] Dong X, Lei Y, Wang T, Thomas M, Tang L, Curran WJ, et al. Automatic multiorgan segmentation in thorax CT images using U-net-GAN. Med Phys 2019;46:2157–68. https://doi.org/10.1002/mp.13458.

[27] Feng X, Qing K, Tustison NJ, Meyer CH, Chen Q. Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3D images. Med Phys 2019;46:2169–80. https://doi.org/10.1002/mp.13466.