



Noise2Aliasing: Unsupervised Deep Learning for View Aliasing and Noise Reduction in 4DCBCT

Samuele Papa^{1,2}✉, Efstratios Gavves¹, and Jan-Jakob Sonke²

¹ University of Amsterdam, Amsterdam, The Netherlands
s.papa@nki.nl

² The Netherlands Cancer Institute, Amsterdam, The Netherlands

Abstract. Respiratory Correlated Cone Beam Computed Tomography (4DCBCT) is a technique used to address respiratory motion artifacts that affect reconstruction quality, especially for the thorax and upper-abdomen. 4DCBCT sorts the acquired projection images in multiple respiratory correlated bins. This technique results in the emergence of aliasing artifacts caused by the low number of projection images per bin, which severely impacts the image quality and limits downstream use. Previous attempts to address this problem relied on traditional algorithms, while only recently deep learning techniques are being employed.

In this work, we propose Noise2Aliasing, which reduces both view-aliasing and statistical noise present in 4DCBCT scans. Using a fundamental property of the FDK reconstruction algorithm, and prior results from the literature, we prove mathematically the ability of the method to work and specify the underlying assumptions.

We apply the method to a public dataset and to an in-house dataset and show that it matches the performance of a supervised approach and outperforms it when measurement noise is present in the data.

Keywords: Medical Imaging · Adaptive Radiotherapy · 4DCBCT · Deep Learning · Unsupervised Learning

1 Introduction

Radiotherapy (RT) is one of the cornerstones of cancer patients. It utilizes ionizing radiation to eradicate all cells of a tumor. The total radiation dose is typically divided over 3–30 daily fractions to optimize its effect. As the surrounding normal tissue is also sensitive to radiation, highly accurate delivery is vital. Image guided RT (IGRT) is a technique to capture the anatomy of the day using in room imaging in order to align the treatment beam with the tumor location [1]. Cone Beam CT (CBCT) is the most widely used imaging modality for IGRT.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/978-3-031-43999-5_46.

A major challenge especially for CBCT imaging of the thorax and upper-abdomen is the respiratory motion that introduces blurring of the anatomy, reducing the localization accuracy and the sharpness of the image.

A technique used to alleviate motion artifacts is Respiratory Correlated CBCT (4DCBCT) [16]. From the projections, it is possible to extract a respiratory signal [12], which indicates the position of the organs within the patient during breathing. With this, subsets of the projections can be defined to create reconstructions that resolve the motion. However, only 20 to 60 respiratory periods are imaged. This limits the number of projections available and results in view-aliasing [16]. Additionally, the projections are affected by stochastic measurement noise caused by the finite imaging dose used, which further degrades the quality of the reconstruction even when all projections are used.

Several traditional methods based on iterative reconstruction algorithms and motion compensation techniques are used to reduce view-aliasing in 4DCBCTs [7, 10, 11, 14, 15]. Although effective, these methods suffer from motion modeling uncertainty and prolonged reconstruction times.

Deep learning has been proposed as a way to address view-aliasing with accelerated reconstruction [6]. However, the method cannot reduce measurement noise because it is still present in the images used as targets during training.

A different method, called Noise2Inverse, uses an unsupervised approach to reduce measurement noise in the traditional CT setting [4]. There are two ways to apply it to 4DCBCT and both fail to reduce stochastic noise effectively. The first is to apply Noise2Inverse to each respiratory-correlated reconstruction. In this case, the method will struggle because of the very low number of projections that are available. The second is to apply Noise2Inverse directly to all the projections. In this case, the motion artifacts that blur the image will appear again, as Noise2Inverse requires averaging the sub-reconstructions to obtain a clean reconstruction.

We propose **Noise2Aliasing** to address these limitations. The method can be used to *provably* train models to reduce both *view-aliasing artifacts* and *stochastic noise* from 4DCBCTs in an *unsupervised* way. Training deep learning models for medical applications often needs new data. This was not the case for Noise2Aliasing, and historical clinical data sufficed for training.

We validated our method on publicly available data [15] against a supervised approach [6] and applied it to an internal clinical dataset of 30 lung cancer patients. We explore different dataset sizes to understand their effects on the reconstructed images.

2 Theoretical Background

In this section, we will introduce the concepts and the notation necessary to understand the method and the choices made during implementation.

Unsupervised noise reduction with Noise2Noise. Given input-target pairs $x, y \in \mathbb{R}$ we can define the regression problem in the one-dimensional setting as finding

$f^* : \mathbb{R} \rightarrow \mathbb{R}$ which satisfies the following:

$$f^* = \arg \min_f \mathbb{E}_{x,y} \left[\|f(x) - y\|_2^2 \right], \quad (1)$$

which can be minimized point-wise [3], yielding:

$$f^*(x) = \mathbb{E}_{y|x} [y|x]. \quad (2)$$

In Noise2Noise [5], input-target pairs are two samples of the same image that only differ because of some independent mean-zero noise $(x + \delta_1, x + \delta_2)$ with $\mathbb{E}_{\delta_2} [x + \delta_2|x + \delta_1] = x$. Then f^* will recover the input image without any noise:

$$f^*(x + \delta_1) = \mathbb{E}_{\delta_2} [x + \delta_2|x + \delta_1] = x. \quad (3)$$

Denoising for Tomography with Noise2Inverse. During a CT scan, a volume \mathbf{x} is imaged by acquiring projections $\mathbf{y} = A\mathbf{x}$ using an x-ray source and a detector placed on the opposite side of the volume. The projections can then be used by an algorithm that computes a linear operator R to obtain an approximation of the original distribution of x-ray attenuation coefficients $\hat{\mathbf{x}} = Ry$. The algorithm can also operate on a subset of the projections. Let $\mathcal{J} = \{1, 2, \dots\}$ be the set of all projections and $J \subset \mathcal{J}$, then $\hat{\mathbf{x}}_J = R_J y_J$ is the reconstruction obtained using only projections y_J . Let us now assume that the projections have some mean-zero noise $\tilde{y}_i = y_i + \epsilon$ with $\mathbb{E}_\epsilon (\tilde{y}_i) = y_i$. Then, in Noise2Inverse [4] the results from Noise2Noise are extended to find a function f^* which removes projection noise when trained using noisy reconstructions $\tilde{\mathbf{x}}_J = R_J \tilde{y}_J = R_J y_J + R_J \epsilon = \hat{\mathbf{x}}_J + R_J \epsilon$ and the expected MSE as loss function. In particular, they find that the loss function can be decomposed in the following way:

$$\mathcal{L} = \mathbb{E} \|f(\tilde{\mathbf{x}}_{J'}) - \tilde{\mathbf{x}}_J\|_2^2 = \mathbb{E} \|f(\tilde{\mathbf{x}}_{J'}) - \hat{\mathbf{x}}_J\|_2^2 + \mathbb{E} \|\hat{\mathbf{x}}_J - \tilde{\mathbf{x}}_J\|_2^2, \quad (4)$$

where J is a random variable that picks subsets of projections at random and J' is its complementary.

Given Eq. 2, we observe that function f^* which minimizes \mathcal{L} is:

$$f^*(\tilde{\mathbf{x}}_{J'}) = \mathbb{E}_J (\hat{\mathbf{x}}_J | \tilde{\mathbf{x}}_{J'}). \quad (5)$$

When using reconstructions from a subset of noisy projections as input and reconstructions from their complementary as its output, a neural network will learn to predict the expected reconstruction without the noise.

Property of Expectation over Subsets of Projections Using FDK. Now let J be a random variable that selects subsets of projections $J \subset \mathcal{J}$ at random such that each projection is selected at least once. Define $R_J : \mathbb{R}^{D_d \times |J|} \rightarrow \mathbb{R}^{D_v}$ to be the FDK reconstruction algorithm [2] that reconstructs a volume of dimensionality D_v from projections J each with dimensionality D_d (geometrical details on the exact setup are not relevant). The FDK uses, as its fundamental step, the dual

Radon transform [9], which is a weighted summation that can be written as an expectation. Then, the following holds:

$$\hat{x} = R_J y = \mathbb{E}_{J \sim \mathcal{J}} [R_J y_J] = \mathbb{E}_{J \sim \mathcal{J}} [\hat{x}_J]. \quad (6)$$

3 Noise2Aliasing

Here, we propose Noise2Aliasing, an unsupervised method capable of reducing both view-aliasing and projection noise in 4DCBCTs. At the core of this method is the following proposition.

Proposition. Given the projection set $\mathcal{J} = \{1, 2, \dots\}$, the FDK reconstruction algorithm R , and the noisy projections $\tilde{\mathbf{y}} = A\mathbf{x} + \epsilon$ with ϵ mean-zero element-wise independent noise. Let J_1, J_2 be two random variables that pick different subsets at random belonging to a partition of \mathcal{J} , and $(\tilde{x}_{J_1} = R_{J_1} \tilde{y}_{J_1}, \tilde{x}_{J_2} = R_{J_2} \tilde{y}_{J_2}) \in \mathcal{D}$ be the input-target pairs in dataset \mathcal{D} of reconstructions using disjoint subsets of noisy projections. Let \mathcal{L} be the expected MSE over \mathcal{D} with respect to a function $f : \mathbb{R}^{D_v} \rightarrow \mathbb{R}^{D_v}$ and the previously-described input-target pairs. Then, we find that the function f^* that minimizes \mathcal{L} for any given $J \in \mathcal{J}$ will reconstruct the volume using all the projections and remove the noise ϵ :

$$f^*(\tilde{x}_J) = \hat{x}. \quad (7)$$

Proof. The loss function \mathcal{L} is defined in the following way:

$$\mathcal{L} = \mathbb{E}_{\mathcal{D}} \|f(\tilde{x}_{J_2}) - \tilde{x}_{J_1}\|_2^2. \quad (8)$$

Additionally, J_1, J_2 are disjoint, the noise is mean-zero element-wise, and we are using the FDK reconstruction algorithm which defines a linear operator R . These allow us to use Eq. 5 to find that the function f^* that minimizes \mathcal{L} is the following:

$$f^*(\tilde{x}_J) = \mathbb{E}_{J_1, J_2} (\hat{x}_{J_1} | \tilde{x}_{J_2} = \tilde{x}_J). \quad (9)$$

This is sufficient to reduce stochastic noise but we need to further manipulate this expression to address view aliasing. Simplifying notation and using the properties of conditional expectations, we can write:

$$f^*(z) = \mathbb{E}_{j_1 \sim J_1} [\mathbb{E}_{j_2 \sim J_2} (\hat{x}_{j_1} | \tilde{x}_{j_2} = z)], \quad (10)$$

now assume that \hat{x}_{j_1} is the clean reconstruction that is consistent with the observed noisy reconstruction z obtained from each disjoint subset j_2 , then:

$$f^*(z) = \mathbb{E}_{j_1 \sim J_1} (\hat{x}_{j_1}). \quad (11)$$

Finally, we use the property of the FDK from Eq. 6:

$$f^*(z) = \mathbb{E}_{j_1 \sim J_1} (\hat{x}_{j_1}) = \hat{x}. \quad (12)$$

□

3.1 Design Choices Based on the Proposition

The proposition guided the choice of reconstruction method to be FDK and the design of the subset selection method from considerations that are now explained.

Equation 12 holds true only when the *same* underlying *clean* reconstruction \hat{x} can be determined from the noisy reconstruction using any subset from a partition of the projections \mathcal{J} . This means that, in our dataset, we should have at our disposal reconstructions of the same underlying volume x using disjoint subsets of projections. In 4DCBCTs this is not the case, as separate respiratory phases are being reconstructed, where the organs are in different positions. We can address this problem by carefully choosing subsets of projections that result in *respiratory-uncorrelated* reconstructions. The reconstructions will display organs in their average position and, therefore, have the same underlying structure. When the projections are selected with the same sampling pattern as the one used in respiratory-correlated reconstructions, then the view-aliasing artifacts display will have the same pattern as the ones present in the 4DCBCTs.

Compared to previous work, to obtain the additional effect of reducing projection noise, the *respiratory-uncorrelated* reconstructions must use non-overlapping subsets of projections. Coincidentally, a previously proposed subset selection method utilized for supervised aliasing reduction fits all these requirements and will, therefore, be used in this work [4].

4 Experiments

First, we used the SPARE Varian dataset to study whether Noise2Aliasing can *match* the performance of the supervised baseline and if it can *outperform* it when adding noise to the projections. Then, we use the internal dataset to explore the *requirements* for the method to be applied to an existing clinical dataset. These required around 64 GPU days on NVIDIA A100 GPUs.

Training of the model is done on 2D slices. The projections obtained during a scan are sub-sampled according to the pseudo-average subset selection method described in [6] and then used to obtain 3D reconstructions. In Noise2Aliasing these are used for both input and target during training. Given two volumes (x, y) , the training pairs $(x_{i(k)}, y_{i(k)})$ are the same i -th slice along the k -th dimension of each volume chosen to be the axial plane.

The Datasets used in this study are two:

1. The SPARE Varian dataset was used to provide performance results on publicly available patient data. To more closely resemble normal respiratory motion per projection image, the 8 min scan has been used from each patient (five such scans are available in the dataset). Training is performed over 4 patients while 1 patient is used as a test set. The hyperparameters are optimized over the training dataset.

2. An internal dataset (IRB approved) of 30 lung cancer patients' 4DCBCTs from 2020 to 2022, originally used for IGRT, with 25 patients for training and 5 patients for testing. The scans are 4 min 205° scans with 120keV source and 512 × 512 sized detector, using Elekta LINACs. The data were anonymized prior to analysis.

Projection Noise was added using the Poisson distribution to the SPARE Varian dataset to evaluate the ability of the unsupervised method to reduce it. Given a projected value of p and a photon count π (chosen to be 2500), the rate of the Poisson distribution is defined as πe^{-p} and given a sample q from this distribution, then the new projected value is $\tilde{p} = -\log(\frac{q}{\pi})$.

The Architecture used in this work is the Mixed Scale Dense CNN (MSD) [8], the most successful architecture from Noise2Inverse [4]. The MSD makes use of dilated convolutions to process features at all scales of the image. We use the MSD with depth 200 and width 1, Adam optimizer, MSE loss, a batch size of 16, and a learning rate of 0.0001.

The Baselines we compare against are two. The first is the traditional FDK obtained using RTK [13]. The second is the supervised approach proposed by [6], where we replace the model with the MSD, for a fair comparison. In the supervised approach, the model is trained by using as input reconstructions obtained from subsets defined with pseudo-average subset selection while the targets use all of the projections available.

The Metrics used in this work are the Root Mean Squared Error (RMSE), Peak Signal-to-Noise Ratio (PSNR), and Structural Similarity Index Measure (SSIM) [17]. All the metrics are defined between the output of the neural network and a 3D (CB)CT scan. For the SPARE Varian dataset, we use the ROIs defined provided [15] and used the 3D reconstruction using all the projections available as a ground truth. For the internal dataset, we deformed the planning CT to each of the phases reconstructed using the FDK algorithm and evaluate the metric over only the 4DCBCT volume boundaries.

5 Results and Discussion

SPARE Varian. Inference speed with the NVIDIA A100 GPU averages 600ms per volume made of 220 slices. From the qualitative evaluation of the methods in Fig. 1, Noise2Aliasing matches the visual quality of the supervised approach on the low-noise dataset on both soft tissue and bones. The metrics in Table 1 show mean and standard deviation across all phases for a single patient. In the low-noise setting, both supervised and Noise2Aliasing outperform FDK with very similar results, often within a single standard deviation.

Noise2Aliasing successfully matches the performance of the supervised baseline.

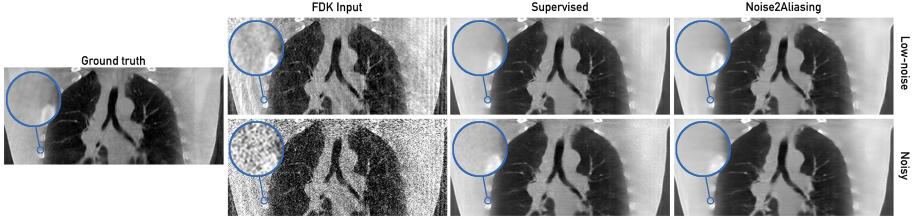


Fig. 1. Qualitative comparison between methods using coronal view of the patient in the test set. Noise2Aliasing and the Supervised method produce very similar images in the low-noise case. With noisy data, the supervised method tends to re-create the noise seen during training.

Table 1. Metrics for the comparison between FDK, Supervised method, and Noise2Aliasing (N2A). Values are *mean* and *std* computed across all phases of patient 1 of the SPARE Varian dataset. The Planning Target Volume (PTV) ROI is less affected by noise compared to the whole *Body*, which is what causes the supervised model to outperform N2A in terms of PSNR and RMSE.

	SSIM $\uparrow (\times 10^{-2})$		PSNR \uparrow		RMSE $\downarrow (\times 10^{-3})$	
Noisy	<i>Body</i>	<i>PTV</i>	<i>Body</i>	<i>PTV</i>	<i>Body</i>	<i>PTV</i>
FDK	12.99 ± 2.1	25.31 ± 4.2	14.66 ± 1.0	13.83 ± 1.0	6.70 ± 0.8	5.51 ± 0.6
Supervised	59.76 ± 2.1	72.61 ± 2.7	22.20 ± 0.2	20.59 ± 0.4	2.80 ± 0.1	2.52 ± 0.1
N2A	64.90 ± 0.8	76.33 ± 1.4	22.31 ± 0.2	20.41 ± 0.4	2.76 ± 0.1	2.57 ± 0.1

Low-Noise						
	<i>Body</i>	<i>PTV</i>	<i>Body</i>	<i>PTV</i>	<i>Body</i>	<i>PTV</i>
FDK	41.75 ± 2.1	56.77 ± 4.2	20.86 ± 1.0	19.09 ± 1.0	3.27 ± 0.8	2.99 ± 0.6
Supervised	67.49 ± 0.8	79.54 ± 2.3	22.68 ± 0.2	20.92 ± 0.5	2.65 ± 0.1	2.43 ± 0.1
N2A	67.13 ± 0.7	79.52 ± 2.0	22.50 ± 0.2	20.79 ± 0.5	2.70 ± 0.1	2.46 ± 0.1

Noisy SPARE Varian. From Fig. 1 and Table 1, the supervised approach reproduces the noise that was seen during training, while Noise2Aliasing manages to remove it consistently, outperforming the supervised approach, especially in the soft tissue area around the lungs, where the noise affects attenuation coefficients the most.

Noise2Aliasing is capable of reducing the artifacts present in reconstructions caused by stochastic noise in the projections used, outperforming the supervised baseline.

Internal Dataset. Noise2Aliasing trained on 25 patients and tested on 5 achieved mean PSNR of 35.24 and SSIM of 0.91, while the clinical method achieved mean PSNR of 29.97 and 0.74 SSIM with p-value of 0.048 for the PSNR and 0.0015 for the SSIM, so Noise2Aliasing was significantly better according to both metrics. Additionally, from Fig. 3 we can see how the breathing extent is matched with sharp reconstruction of the diaphragm. Overall, using more patients results in better noise reduction and sharper reconstructions (see Fig. 2),

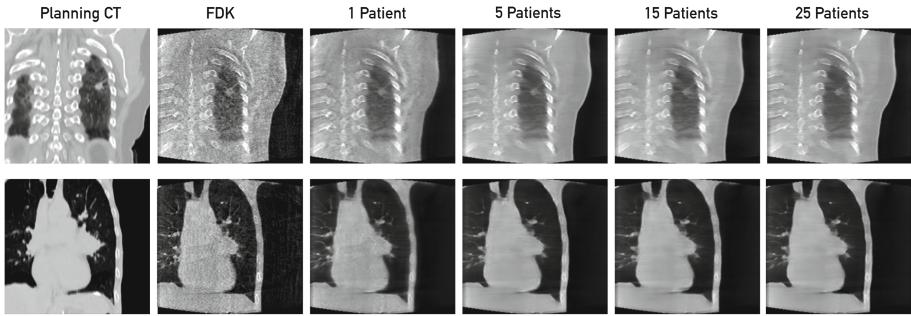


Fig. 2. Reconstruction using Noise2Aliasing with different-sized datasets. With fewer patients, the model is more conservative and tends to keep more noise, but also smudges the interface between tissues and bones. With more patients, more of the view-aliasing is addressed, and the reconstruction is sharper, however, a few small anatomical structures tend to be suppressed by the model.

especially between fat tissue and skin and around the bones. However, the model also tends to remove small anatomical structures as high-frequency objects that cannot be distinguished from the noise.

When applied to a clinical dataset, Noise2Aliasing benefits from more patients being included in the dataset, however, qualitatively good performance is already achieved with 5 patients. No additional data collection was required and the method can be applied without major changes to the current clinical practice.

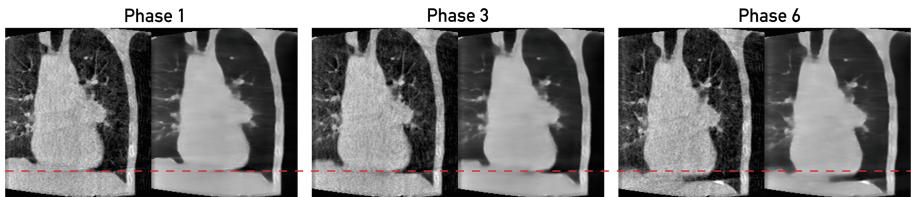


Fig. 3. Motion extent is accurately resolved by Noise2Aliasing when using 25 patients. On the left is the FDK, while on the right is the output of the model.

6 Conclusion

We have presented Noise2Aliasing, a method to provably remove both view-aliasing and stochastic projection noise from 4DCBCTs using an unsupervised deep learning method. We have empirically demonstrated its performance on a publicly available dataset and on an internal clinical dataset. Noise2Aliasing

outperforms a supervised approach when stochastic noise is present in the projections and matches its performance on a popular benchmark. Noise2Aliasing can be trained on existing historical datasets and does not require changing current clinical practices. The method removes noise more reliably when the dataset size is increased, however further analysis is required to establish a good quantitative measurement of this phenomenon. As future work, we plan to study Noise2Aliasing in the presence of changes in the breathing frequency and amplitude between patients and during a scan.

Acknowledgements and Disclosures. We thank Celia Juan de la Cruz, Nikita Moriakov, Xander Staal, and Jonathan Mason for helping during the development of this work.

This work was funded by ROV with grant number PPS2102 and Elekta Oncology AB and was supported by an institutional grant of the Dutch Cancer Society and the Dutch Ministry of Health.

References

1. Dawson, L.A., Jaffray, D.A.: Advances in image-guided radiation therapy. *J. Clin. Oncol.* **25**(8), 938–946 (2007)
2. Feldkamp, L.A., Davis, L.C., Kress, J.W.: Practical cone-beam algorithm. *Josa a* **1**(6), 612–619 (1984)
3. Hastie, T., Tibshirani, R., Friedman, J.H., Friedman, J.H.: *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, vol. 2. Springer, New York (2009)
4. Hendriksen, A.A., Pelt, D.M., Batenburg, K.J.: Noise2Inverse: self-supervised deep convolutional denoising for tomography. *IEEE Trans. Comput. Imaging* **6**, 1320–1335 (2020). <https://doi.org/10.1109/TCI.2020.3019647>
5. Lehtinen, J., et al.: Noise2Noise: learning image restoration without clean data. In: *Proceedings of the 35th International Conference on Machine Learning*, pp. 2965–2974. PMLR, July 2018
6. Madesta, F., Sentker, T., Gauer, T., Werner, R.: Self-contained deep learning-based boosting of 4D cone-beam CT reconstruction. *Med. Phys.* **47**(11), 5619–5631 (2020). <https://doi.org/10.1002/mp.14441>
7. Mory, C., Janssens, G., Rit, S.: Motion-aware temporal regularization for improved 4d cone-beam computed tomography. *Phys. Med. Biol.* **61**(18), 6856 (2016)
8. Pelt, D.M., Sethian, J.A.: A mixed-scale dense convolutional neural network for image analysis. *Proc. Natl. Acad. Sci.* **115**(2), 254–259 (2018)
9. Quinto, E.T.: An introduction to x-ray tomography and radon transforms. In: *Proceedings of Symposia in Applied Mathematics*, vol. 63, p. 1 (2006)
10. Ren, L., et al.: A novel digital tomosynthesis (DTS) reconstruction method using a deformation field map. *Med. Phys.* **35**(7Part1), 3110–3115 (2008)
11. Riblett, M.J., Christensen, G.E., Weiss, E., Hugo, G.D.: Data-driven respiratory motion compensation for four-dimensional cone-beam computed tomography (4D-CBCT) using GroupWise deformable registration. *Med. Phys.* **45**(10), 4471–4482 (2018)
12. Rit, S., van Herk, M., Zijp, L., Sonke, J.J.: Quantification of the variability of diaphragm motion and implications for treatment margin construction. *Int. J. Radiat. Oncol. * Biol.* Phys.* **82**(3), e399–e407 (2012)

13. Rit, S., Oliva, M.V., Brousmiche, S., Labarbe, R., Sarrut, D., Sharp, G.C.: The reconstruction toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the insight toolkit (ITK). *J. Phys. Conf. Ser.* **489**, 012079 (2014)
14. Rit, S., Wolthaus, J.W., van Herk, M., Sonke, J.J.: On-the-fly motion-compensated cone-beam CT using an a priori model of the respiratory motion. *Med. Phys.* **36**(6Part1), 2283–2296 (2009)
15. Shieh, C.C., et al.: Spare: sparse-view reconstruction challenge for 4d cone-beam CT from a 1-min scan. *Med. Phys.* **46**(9), 3799–3811 (2019)
16. Sonke, J.J., Zijp, L., Remeijer, P., van Herk, M.: Respiratory correlated cone beam CT: respiratory correlated cone beam CT. *Med. Phys.* **32**(4), 1176–1186 (2005). <https://doi.org/10.1118/1.1869074>
17. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. *IEEE Trans. Image Process.* **13**(4), 600–612 (2004)