

Deep Learning based Respiratory Surrogate Signal Extraction

Jean Radig, Pascal Paysan, and Stefan Scheib

Varian, a Siemens Healthineers company, Taefernstrasse 7, 5405 Daettwil, Switzerland

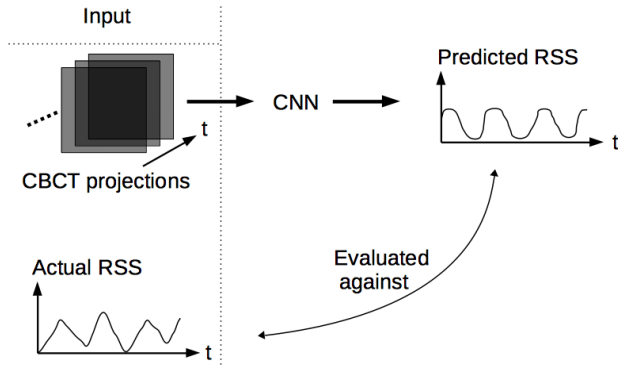
ABSTRACT

We present a feasibility study on extracting the respiratory surrogate signal (RSS) from cone-beam computed tomography (CBCT) projections using a supervised convolutional neural network (CNN) model. Determining the intrinsic RSS instead of using an external surrogate signal, provided by optical tracing hardware such as the Real-time Position Management (RPM) system, has the advantage that it spares patient setup time and hence permits faster 4D CBCT acquisition. Another convenience of such an approach is that it can be applied retrospectively without special preparation or equipment. For the implementation, we made use of the MONAI open source library. Our model is based on a modified version of the MONAI regressor class. We trained the model using CBCT projections with the corresponding RSS as recorded by an external marker block using the RPM system. The model is to deduce the RSS given the CBCT. Using a specific dataset with CBCT from anesthetized animals breathing with the help of a mechanical ventilator, results show a good correlation between the actual and predicted RSS. Unlike the Amsterdam Shroud algorithm, our method shows promising results to predict the normalized amplitude of the breathing signal. Further work could extend the model to permit RSS prediction for its online use in radiation therapy or detection of sudden motion deteriorating CBCT image quality. To conclude, we have made a first step towards proving the concept which consists in using a deep learning model to extract the RSS out of acquired CBCT projection images. The approach is promising but requires more work for robustness, i.e. for sufficient accuracy both in the frequency and normalized amplitude extraction.

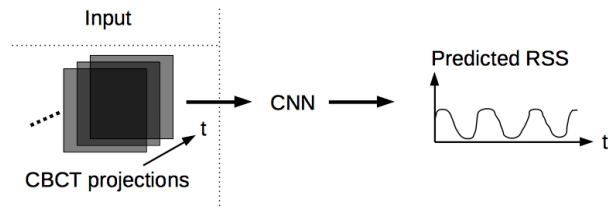
Keywords: CBCT, respiratory phase, respiratory amplitude, convolutional neural network, MONAI

1. INTRODUCTION

In medical imaging as well as in radiation therapy, patient motion is a challenge that needs to be dealt with. Knowing the internal and external motions of the patient permits taking them into consideration for image reconstruction and treatment planning. Of interest is for example the patient's chest motion, as it gives a proxy for the respiratory motion. Obtaining the patient's chest motion can be done by recording the position of a marker block, a device positioned onto the patient's chest. The position is recorded by optical tracing such as the Real-Time Position Management (RPM) system. The technique requires additional installation time, which we would like to avoid. To this end, we will focus on extracting the patient's chest motion, called for generality the respiratory surrogate signal (RSS), intrinsically from the patient's x-ray projections. In the following, we will focus on CBCT projection images. Selecting projections for the same bin of the RSS periodically at different angles around the patient permits their use for 4D CBCT reconstruction. The 4D CBCT can then be used e.g. for radiotherapy planning, patient setup, motion analyses, and treatment beam gating. Hence the reliable intrinsic determination of the RSS would permit efficient 4D CBCT acquisition and clinical use. Various methods to obtain the RSS directly from the CBCT projections have been proposed such as the Amsterdam Shroud (AS) algorithm,¹ a Fourier transform-phase based method,² intensity-based determination,³ a center-of-mass based⁴ method, and the LPCA method.^{5,6} A comparison of the methods has been presented in.⁷ Even methods using AI to segment the diaphragm⁸ for tumor tracking and breathing phase extraction have been proposed. We compare the presently proposed method against our in-house Amsterdam Shroud implementation which shows good accuracy against recorded RSS from the RPM system but lacks the ability to recover any kind of motion amplitude and makes certain assumptions about expected frequencies. In this work, we would like to investigate if intrinsic RSS extraction could be obtained by using a 3D-CNN model. Such a method would be both fast and easy to use. As a first approach, we modified the MONAI regressor class to serve our purpose. MONAI⁹ is an open-source framework for artificial intelligence in medical applications. We use supervised learning by giving both the complete CBCT projection images and the RSS as recorded by the RPM system to the training loop. The model learns to minimize the difference between the predicted RSS and the one as recorded by the RPM system, which we will denote as the "actual" RSS. In the evaluation loop, we would like to get the predicted RSS



(a) Scheme of the training loop



(b) Scheme of the evaluation loop

Figure 1: Scheme of both the training and the evaluation loops

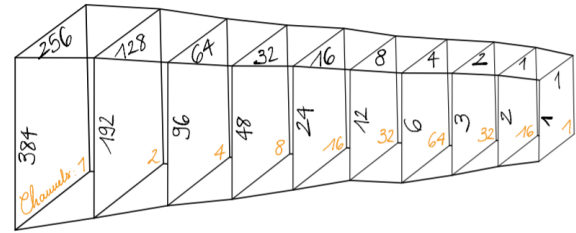


Figure 2: Successive dimension reductions via 9 convolutions are shown for the projection dimensions. The number of projections stays the same throughout the process, only the dimensions in y and x are altered. The channels per convolution are evolving as (2, 4, 8, 16, 32, 64, 32, 16, 1). The final dimension (number of projections) is reflecting the dimension of the recorded RSS.

solely from the CBCT (Fig. 1). In the following, we present the framework and the architecture of our model. We then show some preliminary results and finally conclude by discussing further development and use of the work.

2. FRAMEWORK, ARCHITECTURE, AND DATA

Our CNN model is based on a modified version of the open-source MONAI regressor class implemented in the PyTorch Lightning framework. Given the CBCT raw data, we define the input of dimensions (batch, channels, number of projections, projection dimension along y , projection dimension along x) which in our case, using (by factor 2,4) down-sampled projections, read (1, 1, number of projections [variable in function of the dataset], 384, 256). We operate successive convolutions to collapse the x, y dimensions and preserve the dimension corresponding to the number of projections. Accordingly, the actual RSS for each projection is recorded during CBCT acquisition and given as input. The network consists of 9 convolutions with kernel size (3, 3, 3), padding (1, 0, 0), and strides (1, 2, 2). We set the number of channels for the 9 convolutions to (2, 4, 8, 16, 32, 64, 32, 16, 1).

We replaced the fully connected layer of the MONAI regressor model with a convolution layer to preserve the temporal relation between projections and the corresponding breathing amplitude. Additionally, this leads to a favorable reduction of the number of trainable parameters to 41,9 million for our current model.

To train the model we used as a first instance the CBCT Animal Motion Imaging Study (CAMIS) dataset. In this set, CBCT were taken from animals under general anesthesia using mechanical ventilation and under enforced breath-hold. This results in very regular RSS for the ventilated scans. The amount of data at our disposal was: 21 data sets comprising about 2000 projections and 20 breath-hold scans with about 900 projections each. In order to increase the variation in the training set, we divided the full CBCT scans into batches of subsets of projections. The data points (both RSS and projections) were normalized between 0 and 1 according to the 2% percentile of the complete set. To compensate for systematic baseline drifts in the actual RSS, we applied linear regression to correct for the estimated slope and offset. Doing so we lose information about the absolute breathing amplitude, which is anyhow strongly dependent on the placement of the marker block but preserves the normalized amplitude.

Patient	Free Breathing Model		Free Breathing / Breath-Hold Model	
	Pearson correlation coefficient	Validation loss	Pearson correlation coefficient	Validation loss
A04	0.967	0.004	0.977	0.018
A05	0.943	0.013	0.950	0.012
A06	0.967	0.009	0.973	0.012
A07	0.948	0.013	0.953	0.007
A08	0.977	0.005	0.976	0.009
A09	0.944	0.011	0.959	0.018
A10	0.981	0.009	0.978	0.010
Average	0.961	0.009	0.966	0.012

Table 1: Comparison of our model trained on two different sets. One set only contains "free breathing" animal data while the other also contains "breath-hold" animal data.

3. RESULTS

We trained our model using the animal dataset. To train the model, we subdivided some of the CBCT into batches containing 128 projections along with the corresponding 128 points from the recorded RSS (Figure 3a). In this way, we increased the variance with respect to the angular range of the projections and phase shift of the RSS. After having trained the model, we evaluate it on complete sets of CBCT projection images separated from the training data (Figure 3b, Figure 3c). Note that the model can be applied to CBCT projection image data sets containing any arbitrary number of projections. In addition to the validation loss, we were also interested in the correlation between actual and predicted data points. To visualize the correlation we used scatter plots of the actual versus the predicted RSS. The closer the points to the identity line, the greater the correlation, (Figure 3d, Figure 3e). The Pearson correlation coefficient

$$p = \frac{\text{cov}(y, \hat{y})}{\text{std}(y) \text{std}(\hat{y})}$$

between the actual y and the predicted \hat{y} RSS was applied as quantitative metric to compare results.

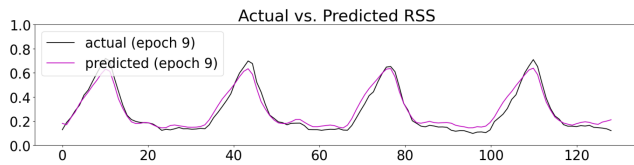
In a second training, we included batches of the enforced breath-hold CBCTs in the training set. We evaluated against the same data, i.e. on complete sets of 4D CBCT acquisitions of mechanically ventilated animals. The idea was to bring more variance into the training set and see in which way it would affect the evaluation performance (Table 1).

Additionally, the model trained using breath-hold data was tested on enforced breath-hold CBCTs (Figure 4).

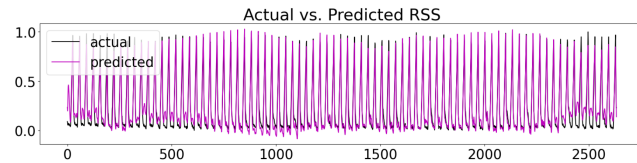
In order to perform a fair comparison with our in-house implementation of the Amsterdam Shroud (AS) algorithm, we base the comparison on the retrospectively calculated phase of the signals. The phase calculation is an in-house algorithm that finds the local maxima (end inhale) of the full signal and assigns the phases accordingly. The reason to compare phases is that the AS algorithm is not able to recover the amplitude of the signal (Figure 5a) but provides sufficient accurate peaks for the phase (Figure 5b) calculation. The retrospective calculated phase of the RPM signal serves as a ground truth referred to as the "actual" phase. As presented in (Table 2) the proposed method slightly outperforms the AS method in terms of Pearson Correlation for most cases. We examined the cases and found a slight phase shift of the extracted signal that needs further attention.

4. DISCUSSION

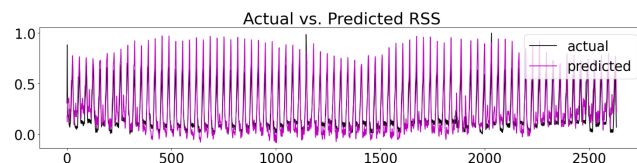
From Figure 3b and 3c we readily see that evaluating data from the animal dataset, i.e. on data with similar variance in phase, frequency, and amplitude yields acceptable results. In particular, the phase and frequency are matched accurately. We still observe variance concerning the matching of the amplitude. A good fitting on the normalized data will provide a good absolute amplitude fitting given we apply the inverse of the normalization process back. In both the best and worst cases the correlation between the actual and predicted RSS is linear, which comes to underline the good accuracy with respect to the frequency matching.



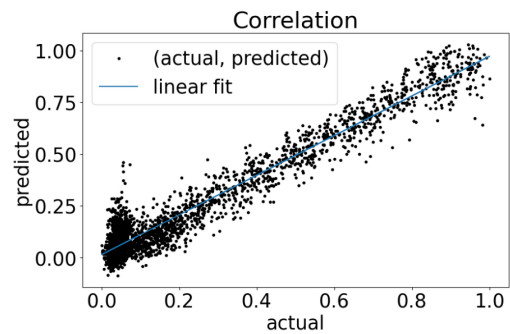
(a) Example of training data. The actual RSS in black and the predicted RSS in violet, over which we optimize.



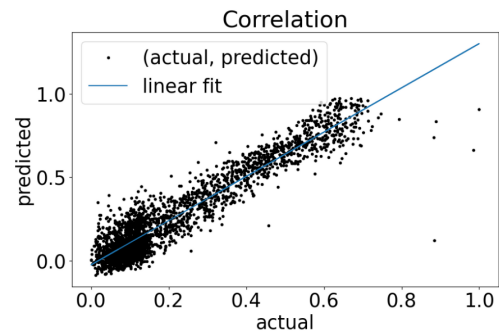
(b) Best evaluation data (w.r.t. the absolute difference between actual and predicted RSS) using a full CBCT projection image data set from the animal dataset.



(c) Worst evaluation data (w.r.t. the absolute difference between actual and predicted RSS) using a full CBCT projection image from the animal dataset.



(d) Correlation between actual and predicted RSS value as derived from the best evaluation data above.



(e) Correlation between actual and predicted RSS value as derived from the worst evaluation data above.

Figure 3: Comparison of the Amsterdam Shroud algorithm and our proposed model. The actual is given by the recorded RSS.

Looking in more detail at the performance of the model trained on free-breathing data versus the model trained on free-breathing plus breath-hold data (Table 1), we remark that both give very similar results, the former having the advantage in the evaluation loss but the latter the one in the correlation between actual and predicted RSS. From these few data, it is not clear whether bringing variance in the training set has consistently a negative impact on evaluation, but this should nonetheless be looked after.

An advantage of our model as trained on breath-hold data (and free-breathing data), is that it systematically manages to correctly discern between breath-hold and free-breathing data (Figure 4). Here the proposed method is able to predict breath-hold signals, unlike the AS method which shows a strong bias towards finding a breathing rate in the expected range. The motivation is that this opens new applications such as monitoring of breath-hold compliance, automatic image quality estimation, or reconstruction method determination.

In Figure 5 and Table 2 are shown comparisons of results obtained using our model against results obtained from the Amsterdam shroud algorithm. We remark that the amplitude (Figure 5a), the phase shift (Figure 5b), and the correlation between actual and predicted RSS (Table 5) are better evaluated by our model.

The results need to be put in contrast with the goal we want to achieve. Evaluating the model trained on the animal dataset on human patient data would not yield concluding results with respect to frequency or amplitude fit. We would need to train our model on a large amount of patient data with a large variance in the RSS to be able to conclude the robustness of the model on patients.

Along with training with a more diverse and human patient-oriented dataset, we could also modify the network. Augmentation techniques (e.g. noise or virtual frame rate changes by skipping projections) can be applied to further increase the variance in the training set.

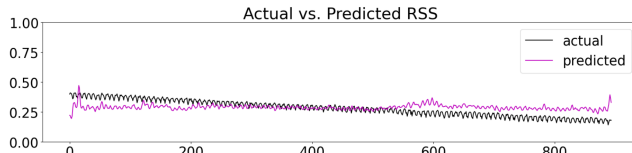
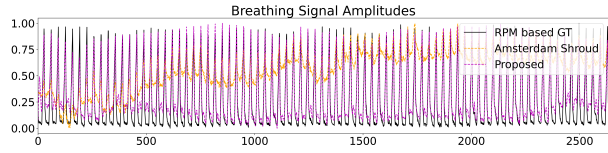
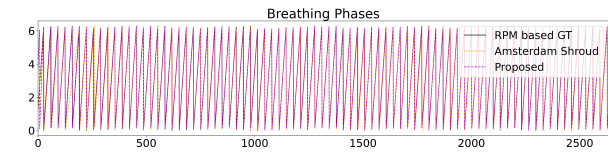


Figure 4: Model evaluated on breath-hold data.



(a) Comparison of RPM with the extracted breathing signals applying the AS and the proposed method. It can be seen that AS method fails to recover the amplitude.



(b) Retrospectively calculated phases show good agreement for all methods. In this case, only the AS method misses one breathing cycle around projection number 200. Figure 5: Comparison of the Amsterdam Shroud algorithm and our proposed model. The actual is given by the recorded RSS.

Table 2: Pearson correlation coefficient of retrospectively calculated phases. Shown is the actual phase calculated based on the RPM signal versus the Amsterdam Shroud (AS) result and the result of the proposed method.

Patient	Pearson	Correla-	Pearson	Correla-
	Actual vs. Phase	tion AS	Actual vs. Proposed Phase	tion Pro-
A04	0.899		0.953	
A05	0.929		0.935	
A06	0.948		0.882	
A07	0.886		0.944	
A08	0.867		0.967	
A09	0.948		0.875	
A10	0.896		0.960	
Average	0.910		0.931	

5. CONCLUSION

We presented our preliminary work concerning the automated extraction of the RSS given a full CBCT projection image data set via the use of a CNN model. We trained and tested our model on an animal (dog) dataset and obtained encouraging results w.r.t. phase, frequency, and normalized amplitude extraction. Concluding on the robustness of the presented method would require training and evaluation of a large amount of patient data. Along selecting a more diverse dataset to train our model we could change its architecture as well to allow for more precision in the amplitude extraction. The following steps would include the use of the predicted RSS for the reconstruction of 4D CBCT.

6. ACKNOWLEDGMENT

The authors thank Prof. Michael Kent from the University of California, Davis, School of Veterinary Medicine for providing the in vivo canine CBCT data.

REFERENCES

- [1] Zijp, L. et al., "Extraction of the respiratory signal from sequential thorax cone-beam x-ray images," in [International conference on the use of computers in radiation therapy], 507–509 (2004).
- [2] Vergalasova, I. et al., "A novel technique for markerless, self-sorted 4D-CBCT: Feasibility study," *Medical Physics* **39**(3), 1442–1451 (2012).
- [3] Kavanagh, A. et al., "Obtaining breathing patterns from any sequential thoracic x-ray image set," *Physics in Medicine and Biology* **54**(16), 4879–4888 (2009).
- [4] Bartling, S. H. et al., "Intrinsic respiratory gating in small-animal CT," *European Radiology* **18**(7), 1375 (2008).
- [5] Yan, H. et al., "Extracting respiratory signals from thoracic cone beam CT projections," *Physics in Medicine and Biology* **58**(5), 1447 (2013).
- [6] Tsai, P. et al., "Tumor phase recognition using cone-beam computed tomography projections and external surrogate information," *Medical Physics* **47**(10), 5077–5089 (2020).
- [7] Martin, R. et al., "Evaluation of intrinsic respiratory signal determination methods for 4d cbct adapted for mice," *Medical Physics* **42**(1), 154–164 (2015).
- [8] Edmunds, D. et al., "Automatic diaphragm segmentation for real-time lung tumor tracking on cone-beam CT projections: a convolutional neural network approach," *Biomedical Physics & Engineering Express* **5**(3), 035005 (2019).
- [9] Consortium, M., "Monai: Medical open network for ai," (Mar. 2020).