



Respiratory Motion Modelling Using cGANs

Alina Giger¹, Robin Sandkühler¹, Christoph Jud¹(✉), Grzegorz Bauman^{1,2},
Oliver Bieri^{1,2}, Rares Salomir³, and Philippe C. Cattin¹

¹ Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland
{alina.giger, christoph.jud}@unibas.ch

² Department of Radiology, Division of Radiological Physics,
University Hospital Basel, Basel, Switzerland

³ Image Guided Interventions Laboratory, University of Geneva,
Geneva, Switzerland

Abstract. Respiratory motion models in radiotherapy are considered as one possible approach for tracking mobile tumours in the thorax and abdomen with the goal to ensure target coverage and dose conformation. We present a patient-specific motion modelling approach which combines navigator-based 4D MRI with recent developments in deformable image registration and deep neural networks. The proposed regression model based on conditional generative adversarial nets (cGANs) is trained to learn the relation between temporally related US and MR navigator images. Prior to treatment, simultaneous ultrasound (US) and 4D MRI data is acquired. During dose delivery, online US imaging is used as surrogate to predict complete 3D MR volumes of different respiration states ahead of time. Experimental validations on three volunteer lung datasets demonstrate the potential of the proposed model both in terms of qualitative and quantitative results, and computational time required.

Keywords: Respiratory motion model · 4D MRI · cGAN

1 Introduction

Respiratory organ motion causes serious difficulties in image acquisition and image-guided interventions in abdominal or thoracic organs, such as liver or lungs. In the field of radiotherapy, respiration induced tumour motion has to be taken into account in order to precisely deliver the radiation dose to the target volume while sparing the surrounding healthy tissue and organs at risk. With the introduction of increasingly precise radiation delivery systems, such as pencil beam scanned (PBS) proton therapy, suitable motion mitigation techniques are required to fully exploit the advantages which come with conformal dose delivery [2]. Tumour tracking based on respiratory motion modelling provides a potential solution to these problems, and as a result a large variety of motion models and surrogate data have been proposed in recent years [7].

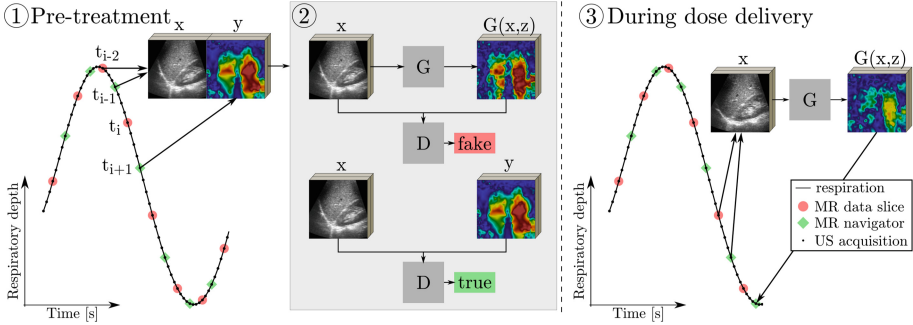


Fig. 1. Schematics of the motion modelling pipeline. See Sect. 2 for details.

In this work we present an image-driven and patient-specific motion modelling approach relying on 2D ultrasound (US) images as surrogate data. The proposed approach is targeted primarily but not exclusively at PBS proton therapy of lung tumours. We combine hybrid US and magnetic resonance imaging (MRI), navigator-based 4D MRI methods [12] and recent developments in deep neural networks [4, 6] into a motion modelling pipeline as illustrated in Fig. 1. In a pre-treatment phase, a regression model between abdominal US images and 2D deformation fields of MR navigator scans is learned using the conditional adversarial network presented in [6]. During dose delivery, US images are used as inputs to the trained model in order to generate the corresponding navigator deformation field, and hence to predict a 3D MR volume of the patient.

Artificial neural networks (ANN) have previously been investigated for time-series prediction in image-guided radiotherapy in order to cope with system latencies [3, 5]. While these approaches rely on relatively simple network architectures, such as multilayer perceptrons with one hidden layer only, a more recent work combines fuzzy logic and an ANN with four hidden layers to predict intra- and inter-fractional variations of lung tumours [8]. Common to the aforementioned methods is that the respiratory motion was retrieved from external markers attached to the patients’ chest, either measured with fluoroscopy [5] or LED sensors and cameras [8]. However, external surrogate data might suffer from a lack of correlation between the measured respiratory motion and the actual internal organ motion [12]. To overcome these limitations, the use of US surrogates for motion modelling offers a potential solution. In [9], anatomical landmarks extracted from US images in combination with a population-based statistical shape model are used for spatial and temporal prediction of the liver. Our work has several distinct advantages over [9]: we are able to build patient-specific and dense volume motion models without the need for manual landmark annotation. Moreover, hybrid US/MR imaging has been investigated for out-of-bore synthesis of MR images [10]. A single-element US transducer was used for generating two orthogonal MR slices.

The proposed image-driven motion modelling approach has only become feasible with recent advances in deep learning, in particular with the introduction

of generative adversarial nets (GANs) [4]. In this framework, two models are trained simultaneously while competing with each other: a generative model G aims to fool an adversarially trained discriminator D , while the latter learns to distinguish between real and generated images. Conditional GANs (cGANs) have shown to be suitable for a multitude of image-to-image translation tasks due to their generic formulation of the loss function [6]. We exploit the properties of cGANs in order to synthesize deformation fields of MR images given 2D US images as inputs.

While all components used within the proposed motion modelling framework have been presented previously, to the best of our knowledge, this is the first approach which suggests to integrate deep neural networks into the field of respiratory motion modelling and 4D MR imaging. We believe the strength of this work lies in the novelty of the motion modelling pipeline and underline two contributions: First, we investigate the practicability of cGANs for medical images where only relatively small training sets are available. Second, we present a patient-specific motion model which is capable of predicting complete MR volumes within reasonable time for image-guided radiotherapy. Moreover, thanks to the properties of the applied 4D MRI method and the availability of ground truth MR scans, we are able to quantitatively validate the prediction accuracy of the proposed approach within a proof-of-concept study.

2 Method

Although MR navigators have been proved to be suitable surrogate data for 4D MR imaging and motion modelling [7, 12], this imaging modality is often not available during dose delivery in radiotherapy. Inspired by image-to-image translation, one could think of a two step process to overcome this limitation: first, a cGAN is trained to learn the relation between surrogate images available during treatment and 2D MR images. Second, following the 4D MRI approach of [12], an MR volume is stacked after registering the generated MR navigator to a master image. The main idea of the approach proposed here is to join these two steps into one by learning the relation between abdominal US images and the corresponding deformation fields of 2D MR navigator slices. Directly predicting navigator deformation fields has the major benefit that image registration during treatment is rendered obsolete as it is inherently learned by the neural network. Since this method is sensitive to the US imaging plane, we assume that the patient remains in supine position and does not stand up between the pre-treatment data acquisition and the dose delivery. The motion modelling pipeline consists of three main steps as illustrated in Fig. 1 and explained below.

2.1 Data Acquisition and Image Registration

Simultaneous US/MR imaging and the interleaved MR acquisition scheme for 4D MR imaging [12] constitute the first key component as shown in step ① of Fig. 1. For 4D MRI, free respiration acquisition of the target volume is performed using

dynamic 2D MR images in sequential order. Interleaved to these so-called data slices, a 2D navigator scan at fixed slice position is acquired. All MR navigator slices are registered to an arbitrary master navigator image in order to obtain 2D deformation fields. Following the slice stacking approach, the data slices representing the organ of interest in the most similar respiration state are grouped to form a 3D MR volume. The respiratory state of the data slices is determined by comparing the deformation fields of the embracing navigator slices. For further details on 4D MRI, we refer to [12]. Unlike [12], deformable image registration of the navigator slices is performed using the approach proposed in [11], which was specifically developed for mask-free lung image registration.

We combine the 4D MRI approach with simultaneous acquisition of US images in order to establish temporal correspondence between the MR navigators and the US surrogate data. For the US image to capture the respiratory motion, an MR-compatible US probe is placed on the subject’s abdominal wall such that the diaphragm’s motion is clearly visible. The US probe is fastened tightly by means of a strap passed around the subject’s chest.

2.2 Training of the Neural Network

We apply image-to-image translation as proposed in [6] in order to learn the regression model between navigator deformation fields and US images. The cGAN is illustrated in step ② of Fig. 1: the generator G learns the mapping from the recorded US images x and a random noise vector z to the deformation field y , i.e. $G : \{x, z\} \mapsto y$. The discriminator D learns to classify between real and synthesised image pairs. For the network to be able to distinguish between mid-cycle states during inhalation and exhalation, respectively, we introduce gradient information by feeding two consecutive US images as input to the cGAN. Since the deformation field has two components, one in x and one y direction, the network is trained for two input and two output channels. Moreover, instead of learning the relation between temporally corresponding data of the two imaging modalities, we introduce a time shift: given the US images at times t_{i-2} and t_{i-1} , we aim to predict the deformation field at time t_{i+1} . Together with the previously generated deformation field at time t_{i-1} , we are then able to reconstruct an MR volume at t_i as the estimates of the embracing navigators are known. In real-time applications, this time shift allows for system latency compensation.

2.3 Real-Time Prediction of Deformation Fields and Stacking

During dose delivery, US images are continuously acquired and fed to the trained cGAN (see step ③ in Fig. 1). The generated deformation fields at times t_{i-1} and t_{i+1} are used to generate a complete MR volume at time t_i by stacking the MR data slices acquired in step ①, analogous to [12].

3 Experiments and Results

Data Acquisition. The data used in this work was tailored to develop a motion model of the lungs with abdominal US images of the liver and the diaphragm

as surrogates. Three hybrid US/MR datasets of two healthy volunteers were acquired on a 1.5 Tesla MR-scanner (MAGNETOM Aera, Siemens Healthineers, Erlangen, Germany) using an ultra-fast balanced steady-state free precession (uf-bSSFP) pulse sequence [1] with the following parameters: flip angle $\alpha = 35^\circ$, TE = 0.86 ms, TR = 1.91 ms, pixel spacing 2.08 mm, slice thickness 8 mm, spacing between slices 5.36 mm, image dimensions $192 \times 190 \times 32$ (rows \times columns \times slice positions). Coronal multi-slice MR scans were acquired in sequential order at a temporal resolution of $f_{MR} = 2.5$ Hz which drops to $f_{MR}/2 = 1.25$ Hz for data slices and navigators considered separately. Simultaneous US imaging was performed at $f_{US} = 20$ Hz using a specifically developed MR-compatible US probe and an Acuson clinical scanner (Antares, Siemens Healthineers, Mountain View, CA). Although the time sampling points of the MR and the US scans did not exactly coincide, we assumed that corresponding image pairs represent the lungs at sufficiently similar respiration states since f_{US} was considerably higher than f_{MR} . The time horizon for motion prediction was $t_h = 1/f_{MR} = 400$ ms.

For each dataset, MR images were acquired for a duration of 9.5 min resulting in 22 dynamics or complete scans of the target volume. Two datasets of the same volunteer were acquired after the volunteer had been sitting for a couple of minutes and the US probe was removed and repositioned. We treat these datasets separately since the US imaging plane and the position of the volunteer within the MR bore changed. The number of data slices and navigators per dataset was $N = 704$ each. Volunteer 2 was advised to breath irregularly for the last couple of breathing cycles. However, we excluded these data for the quantitative analysis below. The datasets were split into $N_{train} = 480$ training images and $N_{test} = \{224, 100, 110\}$ test images for datasets $\{1, 2, 3\}$, respectively. We assumed that the training data represents the pretreatment data as described in Sect. 2.1. It comprised the first 6.4 min or 15 dynamics of the dataset.

Training Details. We adapted the PyTorch implementation for paired image-to-image translation [6] in order for the network to cope with medical images and data with two input and two output channels. The US and MR images were cropped and resized to 256×256 pixels. We used the U-Net based generator architecture, the convolutional PatchGAN classifier as discriminator and default training parameters as proposed in [6]. For each dataset, the network was trained from scratch using the training sets described above and training was stopped after 20 epochs or roughly 7 min.

Validation. For each consecutive navigator pair of the test set a complete MR volume was stacked using the data slices of the training set as possible candidates. In the following, we compare our approach with a reference method and introduce the following notation: RDF is referred to as the reference stacking method using the deformation fields computed on the actually recorded MR navigator slices, and GDF denotes the proposed approach based on the generated deformation fields obtained as a result of the cGAN.

The 2D histogram in Fig. 2 shows the correlation of the slices selected either by RDF or GDF. The bins represent the dynamics of the acquisition and a strong

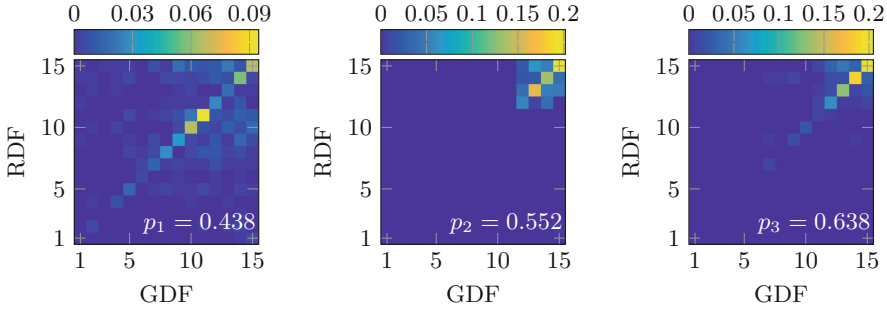


Fig. 2. Slice selection illustrated as joint histogram for reference and generated deformation fields, respectively. From left to right: datasets 1 to 3.

diagonal line is to be expected if the two methods select the same data slices for stacking. The sum over the diagonals, that is the percentage of equally selected slices, is indicated as p_k for dataset $k \in \{1, 2, 3\}$. For all datasets the diagonal is clearly visible and the matching rates are in the range of 43.8% to 63.8%. While these numbers give a first indication of whether the generated deformation fields are able to stack reasonable volumes, they are not a quantitative measure of quality: two different but very similar data slices could be picked by the two methods which would lead to off-diagonal entries but without affecting the image quality of the generated MR volumes.

The histograms for datasets 2 and 3 suggest a further conclusion: the data slices used for stacking are predominantly chosen from the last four dynamics of the training sets (96.5% and 81.7%). Visual inspection of the US images in dataset 2 revealed that one dominant vessel structure appeared more clearly starting from dynamic 11 onwards. This might have been caused by a change in the characteristics of the organ motion, such as organ drift, or a shift of the US probe and emphasises the need for internal surrogate data.

Qualitative comparison of a sample deformation field is shown in Fig. 3a where the reference and the predicted deformations are overlaid. Satisfactory alignment can be observed with the exception of minor deviations in the region of the intestine and the heart. Visual inspection of the stacked volumes by either of the two methods RDF and GDF revealed only minor discontinuities in organ boundaries and vessel structures.

Quantitative results were computed on the basis of image comparison: Each navigator pair of the test set embraces a data slice acquired at a specific slice position. We computed the difference between the training data slice selected for stacking and the actually acquired MR image representing the ground truth. The error was quantified as mean deformation field after 2D registration was performed using the same registration method as in Sect. 2.1 [11]. The median error lies below 1 mm and the maximum error below 3 mm for all datasets and both methods. The average prediction accuracy can compete with previously reported

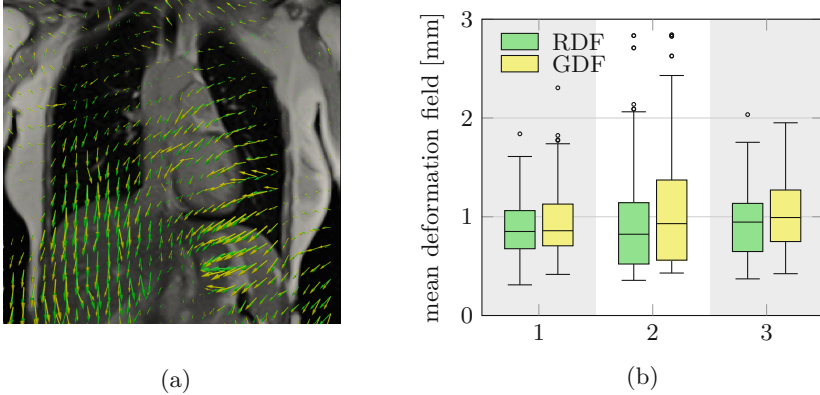


Fig. 3. Qualitative and quantitative results. (a) Sample motion field of dataset 2 with reference (green) and predicted (yellow) deformations, and (b) error distribution quantified as mean deformation field.

values [9]. Comparing RDF and GDF, slightly better results were achieved for the reference method which is, however, not available during treatment.

The proposed method required a mean computation time of 20 ms for predicting the deformation field on a NVIDIA Tesla V100 GPU, and 100 ms for slice selection and stacking on a standard CPU. With a prediction horizon of $t_h = 400$ ms, the motion model is real-time applicable and allows for online tracking of the target volume.

4 Discussion and Conclusion

We presented a novel motion modelling framework which is persuasive in several perspectives: the motion model relies on internal surrogate data, it is patient-specific and capable of predicting dense volume information within reasonable computation time for real-time applications, while training of the regression model can be performed within 7 min only.

We are aware, though, that the proposed approach demands further investigation: It shares the limitation with most motion models that respiration states which have not been observed during pretreatment imaging cannot be reconstructed during dose delivery. This includes in particular, extreme respiration depth or baseline shifts due to organ drift. Also, the motion model is sensitive to the US imaging plane, and a small shift of the US probe may have adverse effects on the outcome. Therefore, the proposed framework requires the patients to remain in supine position with the probe attached to their chests. Future work will aim to alleviate this constraint by, for example, investigating the use of skin tattoos for a precise repositioning of the US probe. Furthermore, the motion model relies on a relatively small amount of training data which bears the danger of overfitting. The current implementation of the cGAN includes dropout but

one could consider to additionally apply data augmentation on the input images. Further effort will be devoted towards the development of effective data augmentation strategies and must include a thorough investigation of the robustness of cGANs within the context of motion modelling. Moreover, the formulation of a control criterion which is capable of detecting defective deformation fields or MR volumes is considered an additional necessity in future works.

Acknowledgement. We thank Tony Lomax, Miriam Krieger, and Ye Zhang from the Centre for Proton Therapy at the Paul Scherrer Institute (PSI), Switzerland, and Pauline Guillemin from the University of Geneva, Switzerland, for fruitful discussions and their assistance with the data acquisition. This work was supported by the Swiss National Science Foundation, SNSF (320030_163330/1).

References

1. Bieri, O.: Ultra-fast steady state free precession and its application to in vivo 1H morphological and functional lung imaging at 1.5 tesla. *Magn. Reson. Med.* **70**(3), 657–663 (2013)
2. De Ruysscher, D., Sterpin, E., Haustermans, K., Depuydt, T.: Tumour movement in proton therapy: solutions and remaining questions: a review. *Cancers* **7**(3), 1143–1153 (2015)
3. Goodband, J., Haas, O., Mills, J.: A comparison of neural network approaches for on-line prediction in IGRT. *Med. Phys.* **35**(3), 1113–1122 (2008)
4. Goodfellow, I., et al.: Generative adversarial nets. In: *Advances in neural information processing systems*, pp. 2672–2680 (2014)
5. Isaksson, M., Jalden, J., Murphy, M.J.: On using an adaptive neural network to predict lung tumor motion during respiration for radiotherapy applications. *Med. Phys.* **32**(12), 3801–3809 (2005)
6. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-Image Translation with Conditional Adversarial Networks. In: *CVPR* (2017)
7. McClelland, J.R., Hawkes, D.J., Schaeffter, T., King, A.P.: Respiratory motion models: a review. *Med. Image Anal.* **17**(1), 19–42 (2013)
8. Park, S., Lee, S.J., Weiss, E., Motai, Y.: Intra-and inter-fractional variation prediction of lung tumors using fuzzy deep learning. *IEEE J. Transl. Eng. Health Med.* **4**, 1–12 (2016)
9. Preiswerk, F., et al.: Model-guided respiratory organ motion prediction of the liver from 2D ultrasound. *Med. Image Anal.* **18**(5), 740–751 (2014)
10. Preiswerk, F.: Hybrid MRI-ultrasound acquisitions, and scannerless real-time imaging. *Magn. Reson. Med.* **78**(3), 897–908 (2017)
11. Sandkühler, R., Jud, C., Pezold, S., Cattin, P.C.: Adaptive graph diffusion regularisation for discontinuity preserving image registration. In: Klein, S., Staring, M., Durrleman, S., Sommer, S. (eds.) *WBIR 2018*. LNCS, vol. 10883, pp. 24–34. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-92258-4_3
12. von Siebenthal, M., Szekely, G., Gamper, U., Boesiger, P., Lomax, A., Cattin, P.: 4D MR imaging of respiratory organ motion and its variability. *Phys. Med. Biol.* **52**(6), 1547 (2007)