Deep correction of breathing-related artifacts in real-time MR-thermometry

Baudouin Denis de Senneville, Pierrick Coupé, Mario Ries, Laurent Facq, Chrit Moonen

HAL Id: hal-03066089
https://hal.science/hal-03066089

Submitted on 15 Dec 2020

Deep correction of breathing-related artifacts in real-time MR-thermometry

B Denis de Senneville^{a,b,c,*}, P Coupé^{d}, M Ries^{e}, L Facq^{a} and CTW Moonen^{e}

^{a} *University of Bordeaux, IMB, UMR CNRS 5251, Talence, France, F-33405 Talence Cedex, France*

^{b} *INRIA Project team Monc, Talence, France, F-33405 Talence Cedex, France*

^{c} *Department of Radiotherapy, UMC Utrecht, Heidelberglaan 100, 3508 GA, The Netherlands*

^{d} *CNRS, University of Bordeaux, Bordeaux INP, "Laboratoire Bordelais de la Recherche Informatique" (LaBRI), UMR5800, F-33400 Talence, France*

^{e} *Imaging Division, UMC Utrecht, Heidelberglaan 100, 3508 GA, Utrecht, The Netherlands*

ARTICLE INFO

ABSTRACT

Real-time MR-imaging has been clinically adapted for monitoring thermal therapies since it can provide on-the-fly temperature maps simultaneously with anatomical information. However, proton resonance frequency based thermometry of moving targets remains challenging since temperature artifacts are induced by the respiratory as well as physiological motion. If left uncorrected, these artifacts lead to severe errors in temperature estimates and impair therapy guidance.

In this study, we evaluated deep learning for on-line correction of motion related errors in abdominal MR-thermometry. For this, a convolutional neural network (CNN) was designed to learn the apparent temperature perturbation from images acquired during a preparative learning stage prior to hyperthermia. The input of the designed CNN is the most recent magnitude image and no surrogate of motion is needed. During the subsequent hyperthermia procedure, the recent magnitude image is used as an input for the CNN-model in order to generate an on-line correction for the current temperature map.

The method's artifact suppression performance was evaluated on 12 free breathing volunteers and was found robust and artifact-free in all examined cases. Furthermore, thermometric precision and accuracy was assessed for *in vivo* ablation using high intensity focused ultrasound. All calculations involved at the different stages of the proposed workflow were designed to be compatible with the clinical time constraints of a therapeutic procedure.

1. Introduction

MRI is used for monitoring thermal therapies since it can provide on-line anatomical informations (given by the spatial distribution of the magnitude of the MR-signal) together with temperature mapping [10] [25] [29]. Many approaches have been developed for MR-thermometry and the Proton-Resonance-Frequency shift (PRF) technique is widely used [3] [6] [19]. In the PRF approach, the phase component φ of the MR-signal, which is acquired using gradient echo sequences, is directly used to estimate voxel-wise temperature variations [3] [12] [15]. Due to the spatial phase variations, this signal component needs to be measured on a voxel-per-voxel basis. Let $\vec{r} = (x, y, z) \in \Omega$ be the voxel coordinates, Ω being the image coordinates domain. An estimate of the temperature change (noted ΔT) at a spatial location $\vec{r}$ and at instant t is obtained by comparing a baseline phase signal acquired at a reference instant t_0 to the phase signal acquired and at t, as follows:

$$\Delta T(\vec{r}, t_n) = (\varphi(\vec{r}, t_0) - \varphi(\vec{r}, t)) \times k \tag{1}$$

k is a constant parameter, more details on its determination can be found in [18]. Note that phase wraps need to be compensated on a voxel-by-voxel basis by adding (resp. substracing) 2π when $\varphi(\vec{r}, t_0) - \varphi(\vec{r}, t) < -\pi$ (resp. $\varphi(\vec{r}, t_0) - \varphi(\vec{r}, t) > \pi$).

While this approach works well on static objects, the application of PRF thermometry to moving targets remains challenging since additional variations of the phase component are induced by: (i) moving the observed tissue through an inhomogeneous magnetic field ; (ii) deforming/changing the tissue so that the demagnitisation field of the tissue changes, which are both a consequence of the patient's physiological activity and the associated organ motion [3]. If left uncorrected, these additional phase variations enter Eq. (1) in full and could lead to severe thermometric errors, leading to abolute errors exceeding the true temperature difference by more than a magnitude.

As a mitigation strategy, one of the first proposed approaches has been respiratory gating. Respiratory gating consists of intermittent acquisitions performed in each exhalation phase of the respiratory cycle [16]. As a trigger for the gating, several types of respiratory motion descriptors have been proposed [14], ranging from external pressure sensors [16], dedicated 1D MR navigator echoes [11] to self gated sequences based on MR magnitude images [5]. Although gating is generally a robust solution to avoid motion induced thermometric errors, it is nevertheless hampered by two drawbacks. First and foremost, the observed motion pattern must be strictly repetitive/periodical and secondly considering a good spatial coverage of the heated region, the achiveable temporal resolution is generally limited to a range of 3 to 6 s [28] [17] [21].

In particular the latter motivated the development of non-gated MR-thermometry correction strategies, which are able to selectively remove motion-induced phase changes from the MR-phase and thus to provide artefact-free temperature maps in real-time. However, the required precise modeling of the inhomogeneous magnetic field *in vivo* and the motion associated phase variations, in particular under real-time conditions for therapy guidance, has been difficult to achieve. Most of these early correction strategies can be coarsely classified into two different types, which are generally referred to as "Referenceless" and "Multi-baseline" PRF thermometry. The interested reader is referred to [7] for a pragmatic analysis of inherent advantages and drawbacks associated with these two correction strategies:

In referenceless PRF thermometry, the baseline phase signal used to compute the current temperature map is directly estimated from the current MR phase image. To this end, the phase signal of non-heated surrounding tissues is used to extrapolate a baseline phase signal in the targeted area [20] [9] [23]. This approach relies on an a priori choice of a region of interest (ROI) and the quality of the thermometry highly depends on an optimal ROI placement. In practice, the fitting ROI has to: (i) encompass — at least to some extent — the ablation area ; (ii) be sufficiently close the target area to allow a precise estimate of the background phase there ; (iii) be sufficiently far from the heating zone to be unaffected by heat diffusion and conduction ; (iv) not encompass areas prone to strong local susceptibility variations.

In multi-baseline PRF thermometry (illustrated in figure 1) a look-up table obtained in absence of temperature variations establishes a relation between the phase variations associated with the patient's physiological motion and a descriptor/detector [27]. The benefit of this approach is evident for application scenarios that do not permit a placement of the fitting ROI fulfilling all above-mentioned four conditions simultaneously. This is generally the case for minimally invasive ablations, or interventions at the boundary of organs. Both MR images and descriptors of motion patterns are continuously and simultaneously acquired during a period covering several respiratory cycles. A look-up table can then be used to store each pair of MR phase image/motion surrogate. During heating, phase artifacts due to the periodical motion of the respiration cycle are addressed by calculating a baseline phase image based on a model of the phase dependence of the current motion descriptor (red block in figure 1). Using multi-baseline strategies, the stability of MR-thermometry largely depends on: (i) the determination of an accurate and precise motion surrogate ; (ii) the accuracy of the model used to address susceptibility related phase changes, especially in regions with complex susceptibility distributions or signal discontinuities.

More recent approaches proposed to fuse these two largely complementary approaches to combined correction strategies, which compensate the respective weaknesses in order to achieve both increased accuracy and a less convoluted work-flow for clinical applications [8] [32].

The contribution of the current study is threefold:

1. We introduce the use of deep learning for on-line correction of motion related errors in abdominal MR-thermometry. The existing multi-baseline strategy is extended by a convolutional neural net-

work (CNN) which learns the apparent temperature perturbation from images acquired during the preparative learning stage. The input of the designed CNN is the current magnitude image and as a consequence no surrogate of motion-state is needed. During the hyperthermia procedure, the most recent magnitude image is used as an input for the pre-built CNN-model in order to generate a correction for the most recent temperature map.

2. Frequently, inherent computational costs are a major difficulty when dealing with deep learning in applications requiring adaptive / on-the-fly training. In order to mitigate this problem for clinical applications of MR-thermometry, a fine-tuning strategy is proposed to accelerate calculations during the preparative learning stage. Moreover, to meet computational requirements for real-time MR-thermometry, which requires that all calculations have to be completed within the interval between two successive image acquisitions, we propose to process all temperature images in a sliding temporal window within one single CNN-model call.

3. The ability of the proposed approach to remove thermometry artifacts is demonstrated for dynamic MRI datasets of the the liver of 12 healthy volunteers in absence of temperature changes. We demonstrate that the amount of learning images and the CNN training time can be optimized to the point that clinical thermotherapy interventions are feasible. Thermometric precision and accuracy is demonstrated with a heating experiment performed on a porcine liver using high intensity focused ultrasound (HIFU) [1]. The proposed method is compared to the two most frequently employed multi-baseline strategies in terms of temperature precision, without penalty in accuracy.

2. Materials and Methods

2.1. Method overview

The proposed method is detailed in figure 2: thermal maps with motion related temperature artifacts and magnitude images (noted M) are combined to establish prior to hyperthermia a CNN-based temperature correction model (noted g). During hyperthermia, the incoming magnitude image is used as an input for the pre-built CNN-model to generate in real time a temperture correction for the current temperature map. Differences with existing multi-baseline approaches (figure 1) are: (i) the input of the correction model is the most recent magnitude image, which eliminates the need for surrogates/sensors ; (ii) the fitted data is the apparent (artifacted) temperature ; (iii) the model is a CNN. The benefit of each of these aspects is discussed later in the manuscript.

2.2. Datasets

Dynamic MR-imaging was performed on a Philips Achieva 1.5 T (Philips Healthcare, Best, The Netherlands) under real-time conditions. The method was evaluated in 2D and the effect of through plane motion was reduced by setting the imaging plane direction parallel to the principal axis of the organ displacement.

2.2.1. Volunteer study.

An imaging frame rate of 10 Hz was maintained during 5 minutes on the abdomen of 12 healthy human volunteers under free-breathing conditions. The MR-protocol was composed of a learning step of 20 s dedicated to the acquisition of the training data, followed by 4 min-40 s devoted to mimic an interventional procedure. The MR-sequence was a single-shot gradient recalled echo-planar with the following parameters: one coronal slice, repetition time (TR)=100 ms, echo time (TE)=26 ms, bandwidth in readout direction=2085 Hz, flip angle=35°, field of view (FOV)=256 × 168 mm^2, slice thickness=6 mm, matrix=128 × 84, using a four element phased array body coil. The volunteer studies depicted various SNR conditions: over the twelve volunteers, the SNR was evaluated as 7 ± 3 (min=4, max=14) in the liver.

2.2.2. In vivo heating study in a porcine liver.

MRI guided HIFU was performed *in vivo* in the liver of a pig under general anesthesia and artificial breathing. The MR sequence employed the following parameters: single-shot, gradient recalled, echo-planar imaging, 1000 dynamic sagittal images, five slice, TR=250 ms, TE=33 ms, flip angle=40°, in-plane FOV=370 × 162 mm^2, voxel size=2.89 × 2.89 × 7 mm^3 using the integrated three elements phased array coil of the HIFU system. A MR compatible HIFU ablation system (Sonalleve, Profound Medical, Helskinki,
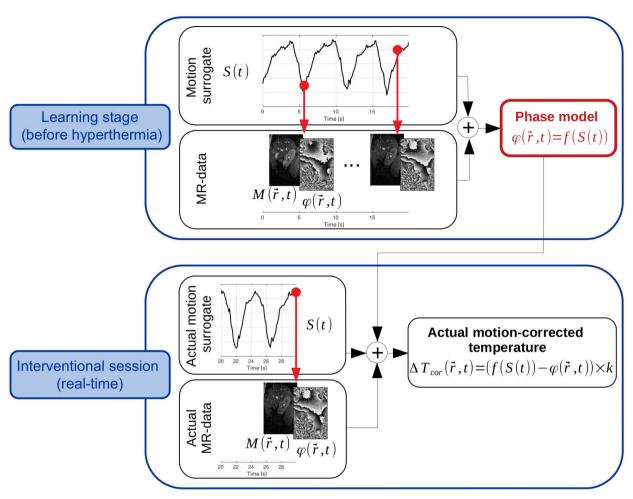
Figure 1: Illustration of a typical multi-baseline correction scheme. Both motion surrogate(s) and MR-images are used for this strategy. Both the motion surrogate $S(t)$ and the MR-images are acquired simultaneously in a training phase before the hyperthermia procedure. A multi-baseline collection is used to store each pair of motion surrogate/phase image. During hyperthermia, thermometry artifacts due to the periodical motion of the respiration cycle are addressed by calculating a baseline phase (noted $f(S(t))$) based on the current motion surrogate and the training phase images.

Finland) composed of a table top containing a 256 elements HIFU transducer, integrated in the 1.5 T Achieva-Intera MRI was used to perform a temperature elevation. The transducer radius and aperture were 120 mm and 126 mm, respectively, creating an ellipsoid focal point ($1 \times 1 \times 7$ mm^3). The animals were placed in the prone position so that the liver was accessible through an unobstructed beam-path directly below the rib-cage. MR-guided hyperthermia was performed for a duration of 4 minutes on the liver with HIFU power of 160 W during 100 s. All animal studies were performed under an approved Animal Care and Use protocol.

2.3. Proposed CNN-based correction
2.3.1. Learning motion-related errors in MR-thermometry.

Motion-related errors in MR-thermometry were learned during a preparative learning stage performed before hyperthermia. This step is based on a training set of N dynamically acquired data (each dynamic data is composed by the magnitude, the phase and the apparent temperature map calculated with Eq. (1)). The motion cycle has to be sampled with a sufficient density in order to avoid discretization errors. With a sufficient imaging frame rate of 5-10 Hz and a respiration frequency of 3-6 seconds this pre-treatment step can be completed in a relatively short duration of 15-20 seconds. For the volunteer study, we tested various
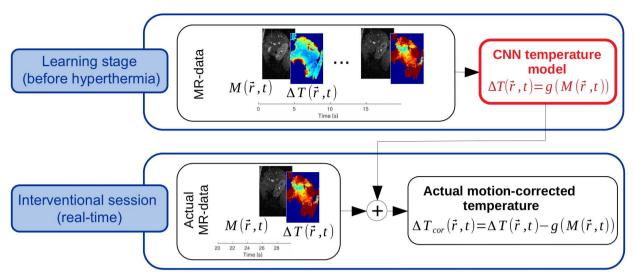
Figure 2: Illustration of the proposed CNN correction scheme. In this approach only MR-images are required and no motion surrogate(s)/sensors are needed. MR-training data are acquired prior to hyperthermia in a training phase. Subsequently, the motion-artifacted thermal maps and the magnitude images are combined to establish a CNN-based correction model of the temperature (red block). During the hyperthermia procedure, the most recent magnitude image is used as an input for the pre-built CNN-model in order to generate a correction map (noted $g(M)$) for the current temperature map.

(reconstructed) imaging frame rates — ranking from 1 Hz (*i.e.*, $N = 20$) to 10 Hz (*i.e.*, $N = 200$) — to train our CNN-model. For the *in vivo* heating study, we used $N = 200$.

2.3.2. Preprocessing of input images.

During both learning and hyperthermia (thermotherapy) stages, all incoming images (*i,e.* anatomical and temperature images) were preprocessed on-the-fly as follows. First, anatomical (magnitude) image intensities were normalized (z-scoring) using the mean and standard deviation within the complete image field-of-view. Second, thermal maps were registered onto a common reference position in order to allow kinetic analysis. Note that this registration step also allows for cumulative thermal dose calculations, which may be beneficial for on-line assessment of the therapy endpoint [24] [26]. In the scope of this manuscript, we registered all incoming phase images using motion estimates of a real-time image optical-flow (OF) algorithm applied to magnitude images. Additional details about the employed image registration algorithm can be found in [30]. Temperature calculations were performed using Eq. (1) applied to the registered phase maps.

2.3.3. Implemented deep neural network model.

Figure 3 describes the architecture of the proposed deep neural network model (noted g) designed to learn motion-related artifacts in MR-thermometry prior to hyperthermia (red block in figure 2). The input of the model is a magnitude image M and the output is a correction (noted $g(M)$) for the corresponding temperature image (*i.e.*, ΔT). Note that potential mis-corrected phase wraps in Eq. (1) may have a direct negative impact on the CNN-model optimization process. To mitigate this drawback, temperature maps used for training were filtered using a median filter (kernel 5×5). We used a convolutional encoder with 3 layers per resolution level, using a basis of 24 filters of 3×3 (*i.e.*, 24 filters for the first layer, 48 for the second and so on). We empirically optimized this setting for reduced memory consumption without impacting performance. Each block was composed of batch normalization, convolution and ReLU activation. We employed the following parameters: batch size = 1, optimizer = Adam with default parameters, epoch = 100, loss = Mean Square Error (MSE) and dropout = 0.5 after each block. We used one single input channel (*i.e.*, the actual magnitude image). The implemented CNN-model is detailed in the supplementary material of the manuscript. The output shape and the number of parameters involved in each layer of the CNN-model are given.
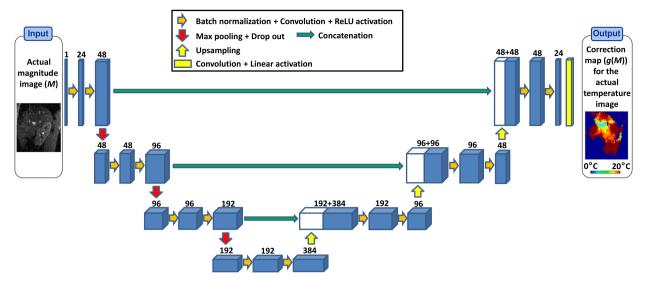
Figure 3: Architecture of the deep neural network proposed for learning motion-related errors in MR-thermometry before heating (red block in figure 2). The most recent magnitude image (*i.e.,* M) is used as a single input channel. The CNN-model constructs a correction map (noted $g(M)$) for the current temperature map. Each block of the CNN is composed of batch normalization, convolution and ReLU activation. The number of 3×3 filters is indicated on the top of each block.

2.3.4. Proposed fine-tuning strategy.

CNN optimization is a complex iterative process which is inherently time consuming and depends on initialization. To reduce these two issues and to meet clinical constraints of a therapeutic workflow, we propose a fine-tuning strategy. To this end, a pre-built CNN model (*i.e.,* a CNN-model trained on several data sets) — is loaded and used as a starting point for the actual model optimization (*i.e.,* the red block in figure 2).

2.3.5. On-line CNN-correction of temperature maps.

At this point we have a model g designed to predict the actual temperature perturbation $g(M)$ given the current anatomical image M. The motion compensated temperature image at instant t can be obtained as follows (see figure 2):

$$\Delta T_{cor}(\vec{r}, t) = \Delta T(\vec{r}, t) - g\left(M(\vec{r}, t)\right) \tag{2}$$

2.3.6. Correction of time-persistent offsets.

Once corrected according to section 2.3.5, the temperature in a voxel at location $\vec{r}$ is prone to a time dependent offset arising from the presence of noise in the baseline phase image (*i.e.,* $\varphi(\vec{r}, t_0)$ in Eq. (1)). To compensate for this offset, we assumed that the temperature change has to be identically equal to 0 during the learning stage (no hyperthermia). Practically, a pre-built temperature time average map — based on data acquired before heating — was subtracted from the actual motion compensated temperature image $\Delta T_{cor}(\vec{r}, t)$.

2.4. Implementation details

We evaluated the computational burden of our proposed method using an Intel Xeon E5-2683 2.4 GHz (2 Hexadeca-core) with 256 GB of RAM equipped by a GPU Nvidia Tesla V100. Our implementation was performed using Tensorflow 1.4 and Keras 2.2.4.

Each of the 13 tested data sets — 12 volunteer data sets + 1 *in vivo* heating data set — were processed without fine-tuning (each data set was processed independently of each other) and with fine-tuning (leave-one-out strategy).

To reduce the computation time during the interventional procedure, all data in a sliding temporal window (*i.e.*, a pack of most recent consecutive dynamically acquired images. We denote by δ the number of dynamic images) were processed simultaneously within one single CNN-model call. In the scope of this study, we tested δ-values in the following set: $\{1, 2, 16\}$.

The usefulness of the two above-mentioned implementation strategies is analyzed in the discussion section.

2.5. Validation framework
2.5.1. Implemented competitive approaches.

The performance of two existing multi-baseline solutions — referred to as "look-up-table" approach (or LUT) [27] and "linear model" approach (or LM) [22] throughout the rest of the manuscript — were also evaluated on the same data sets. In the learning stage of both LUT and LM, MR-images (magnitude and phase) were included over several respiratory cycles (identical number of training data N were employed for both LUT, LM and the proposed CNN method). The position in the respiration cycle (*i.e.*, the motion surrogate denoted by $S(t)$ in figure 1) was monitored using a Principal Component Analysis (PCA) applied to the above-mentioned OF-motion estimates, as described in [4]. The baseline phase image needed for the calculation of the actual temperatures maps with Eq. (1) was calculated in the following two ways:

Look-up-table approach (LUT): Each collected baseline phase image during the learning stage was indexed in a look-up-table according to its estimated position within the breathing cycle given by $S(t)$. During the intervention stage, the baseline phase image in Eq. (1) was a linear interpolation between the closest two reference phase images allowed for reconstructing a baseline phase image for the current position in the respiration cycle.

Linear model approach (LM): The overall phase variation (denoted by $f(S(t))$ in figure 1) was approximated by linear phase changes of the motion surrogate $S(t)$ on a voxel-by-voxel basis as described in [4].

2.5.2. Statistical analysis.

In the volunteer study, it was assumed that the temperature change has to be identically equal to 0 during the testing session (no hyperthermia was performed).

First, the temperature precision was evaluated for each volunteer by computing on a voxel-by-voxel basis the temporal temperature standard deviation (noted SD) within a manually defined mask (noted Γ, $\Gamma \subset \Omega$) encompassing the liver and over the duration of the interventional session (*i.e.*, from the starting instant $t_s = 20$ s to the final instant $t_f = 5$ min):

$$\text{SD}(\vec{r}) = \sigma\left(\Delta T(\vec{r}, t)\right) \qquad t \in [t_s, t_f], \quad \vec{r} \in \Gamma \tag{3}$$

Second, the temperature accuracy was evaluated for each volunteer by computing on a voxel-by-voxel basis the mean absolute temperature error (noted MAE) within Γ and over the interventional step:

$$\text{MAE}(\vec{r}) = \left| \frac{1}{t_f - t_s} \int_{t=t_s}^{t_f} \Delta T(\vec{r}, t) \mathrm{d}t \right| \qquad \vec{r} \in \Gamma \tag{4}$$

The same analysis was performed for the *in vivo* heating study to assess the thermometry precision and the accuracy outside the heated region.

For the volunteer study, a paired Wilcoxon test was carried out in order to study whether SD and MAE differences are statistically significant between LUT-, LM- and CNN-corrected data sets. A significance threshold of $p = 0.025$ was used. The power of the statistical analysis has been carried out, as described in [2].

3. Results

3.1. Volunteer study

Figure 4 shows an example of MR-thermometry results obtained in one volunteer of the examined group (volunteer #2). The leftmost image (4a) depicts the anatomy. The temperature precision (resp. accuracy)

is reported in the upper row (resp. bottom row) for each tested correction solution. Thermometry artifacts caused by motion-related susceptibility variations were compensated using LUT (first column), LM (second column) and CNN (third column). The temperature precision is visually better in the major part of the liver using LM method as compared to LUT (see arrow #1 in 4b and 4c). The best precision is however observable using the CNN method. It can also be noticed that large susceptibility artifacts render the temperature correction difficult in the upper part of the liver (see arrow #2 in 4b, 4c and 4d) and in the vicinity of hepatic arteries (see arrow #3 in 4b, 4c and 4d). In these regions, an improvement of the thermometry precision by up to 2 °C could be obtained using CNN as compared to both LUT and LM. This precision was achieved without negative impact on the accuracy, from a visual point of view, as shown in 4e-g.

Temperature maps from the individual volunteers were pooled in order to obtain a group set containing the temperature precision (resp. accuracy) from all volunteers. The distribution of the temperature precision (resp. accuracy) for the group set is reported in figure 5a (resp. 6a) using LUT, LM and CNN. The distribution of temperature precision (resp. accuracy) is also detailed for each volunteer in 5b (resp. 6b). The paired Wilcoxon test showed that the temperature precision was significantly better using LM as compared to LUT (p<0.001/statistical power=1). Furthermore, the temperature precision was significantly better using CNN as compared to LUT and LM (p<0.001/statistical power=1). Besides, the temperature accuracy was significantly better using CNN as compared to LM (p=0.008/statistical power=1).

Figure 7 analyzes the impact of the amount of learning images: compared to the original 10 Hz imaging frame rate, a 2 Hz frame rate deteriorated moderately thermometric precision and accuracy (by less than 30%).
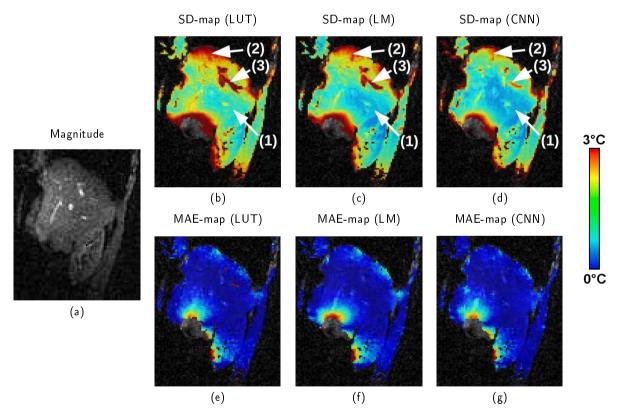

Figure 4: Typical temperature stability maps obtained in the abdomen of volunteer #2 using two existing multi-baseline approaches (*i.e.,* LUT and LM) and using the CNN approach: (a) anatomic image, (upper row) the temperature standard deviation map obtained with the LUT (b), the LM (c) and the CNN method (d), (lower row) the temperature mean absolute error map obtained with LUT (e), LM (f) and CNN (g). $N = 200$ images were used for training (an imaging frame rate of 10 hz was maintained during 20 seconds)
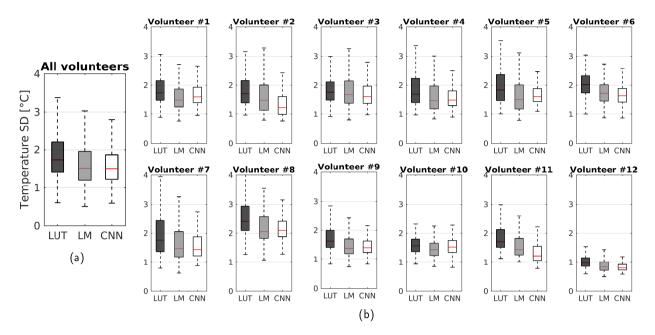
Figure 5: Comparison of temperature precision obtained using LUT, LM and CNN in the liver of free-breathing healthy volunteers during 4 minutes and 40 seconds of MR-thermometry. Similar to figure 4, $N = 200$ images were used for training. Box-and-whisker plots of the temporal temperature standard deviation are shown using LUT (dark gray box), LM (light gray box) and CNN (white box): (a) group analysis over the 12 volunteers, (b) volunteer-wise analysis. The median is shown by the central mark, the first and the third quartiles are reported by the edges of the box, the whiskers extend to the most extreme time points that are not considered as outliers.

3.2. *In vivo* heating study in a porcine liver

Figure 8 shows MR thermometry results obtained on a porcine liver during HIFU heating. Thermal maps are reported after 80 s of sonication using LUT (8a), LM (8b) and CNN (8c). In absence of any correction strategy, apparent temperature fluctuations of up to 13°C (peak-to-peak) were observed in the target area. The heated region using the LUT appears slightly elongated as compared to LM and CNN. Residual thermometry artifacts are observable in the upper part of the liver with LUT (see 8a). These apparent temperature fluctuations are however greatly reduced using LM (8b), and even more using CNN (8c). This visual observation is confirmed in associated SD-maps: in most of the voxels located in the upper part of the liver, a temperature standard deviation higher than 3°C using LUT (8d) decreased until 2°C using LM (8e), and reached 1°C using CNN (8f). This precision gain with CNN was achieved without creating any additional offset, as shown by the MAE-maps (see 8h, 8i and 8j). The evolution of the temperature is shown in a single voxel located at the focal point position using LUT (8k), LM (8l) and CNN (8m). Higher residual temporal temperature fluctuations are observable using the LUT correction as compared to the other two correction approaches. LM and CNN approaches lead to a comparable observation of the temperature evolution: a temperature increase of 12°C was reached after 80 s of HIFU sonication.

3.3. Benchmark

During the learning stage, around 1.5 s were required in average for the accomplishment of one epoch. Figure 9 shows the loss metric as a function of the number of epochs without and with the proposed fine-tuning strategy. It can be observed that the use of fine-tuning stabilized and accelerated the convergence of the optimization process, and this for all data sets involved in this study.
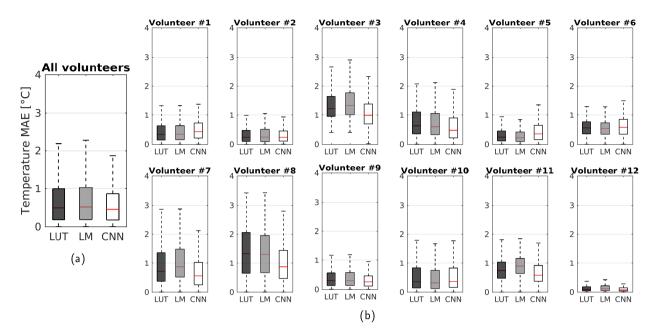
Figure 6: Comparison of temperature accuracy obtained using LUT, LM and CNN in the liver of free-breathing healthy volunteers during 4 minutes and 40 seconds of MR-thermometry. Similar to figures 4 and 6, $N = 200$ images were used for training. Box-and-whisker plots of the temporal temperature mean absolute error are shown using LUT (dark gray box), LM (light gray box) and CNN (white box): (a) group analysis over the 12 volunteers, (b) volunteer-wise analysis.
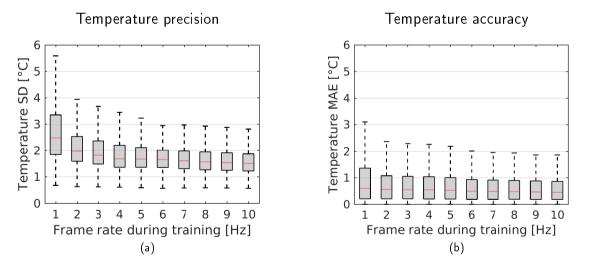

Figure 7: Analysis of the impact of the amount of learning images on the temperature stability. The temperature precision (a) and accuracy (b) were evaluated using various imaging frame rates for training. Box-and-whisker plots of the temporal temperature standard deviation (a) and mean absolute error (b) obtained over the 12 volunteers are reported for imaging frame rates ranking between 1 and 10 Hz. Note that the number N of dynamic images used to train the CNN-model was equal to 20 (resp. 40, 60, ..., 200) when an imaging frame rate of 1 Hz (resp. 2, 3, ... 10 Hz) was employed.
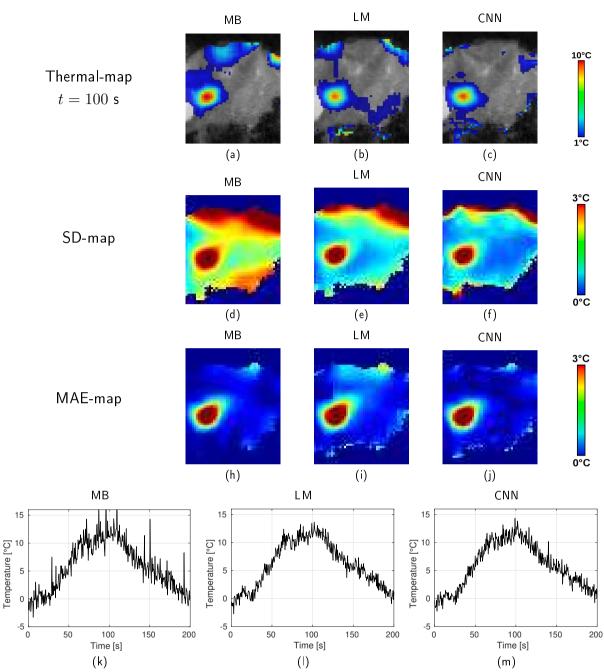
Figure 8: MR thermometry results obtained in a porcine liver during HIFU heating: (first row) Temperature maps obtained after 80 s of heating ($t = 100s$) overlaid on the anatomic image, (second row) temporal standard deviation map, (third row) mean absolute error map, and (bottom row) temporal evolution of the temperature in a single voxel located at the focal point position. Results are reported using LUT (left), LM (middle) and CNN (right).

During the hyperthermia session, 150 ms were required to generate one single motion compensated temperature map. In such a case, the GPU usage was however highly under-exploited: interestingly, the calculation of a pack of $\delta = 16$ maps could be also accomplished within 150 ms.
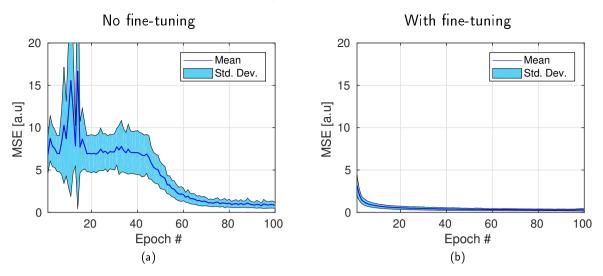
No fine-tuning

(a)

With fine-tuning

(b)

Figure 9: Loss metric as a function of the number of epochs obtained without (a) and with (b) the use of fine-tuning. For each number of epochs, the mean and standard deviation of 13 MSE values (*i.e.*, 12 values for the volunteer study + 1 value for the heating study) are reported.

4. Discussion

The proposed method is designed to remove motion-related susceptibility effects induced by breathing in real-time abdominal MR-thermometry. To this end, the existing multi-baseline strategy is extended using a deep neural network: a CNN learns the apparent temperature perturbation during a preparative learning stage performed before hyperthermia.

A major difficulty when dealing with deep learning is the inherent computational cost. One goal of this study was to investigate if (i) the amount of learning images and (ii) the CNN training time can be optimized to the point that clinical thermotherapy interventions are feasible. For this, all calculations at the different stages of the proposed workflow (*i.e.*, learning stage, CNN model optimization, determination of an offset map and interventional session) were designed to be compatible with the constraints of clinical thermotherapy procedures (see section 3.3). With respect to the learning stage, 20 epochs using fine-tuning provide a loss similar to the one achieved by 100 epochs (see figure 9). The calculation of the CNN-model (red block in figure 2) could thus be accomplished within less than 30 s (resp. less than 6 s) using our test platform with a training frame rate of 10 Hz (resp. 2 Hz). The compensation of the time-persistent offset (as described in section 2.3.6), which relied on a CNN-correction for each of the N training images, could be accomplished within less than a second in all presented experiments. Regarding the interventional session, it is imperative that all calculations have to be done within the interval of subsequent image aquisitions in order to prevent back-log. The use of a sliding temporal window (as introduced in section 2.4) of size $\delta = 2$ dynamic was mandatory in the volunteer study to cope with a 10 Hz imaging frame rate.

A second major challenge is the presence of noise and wraps in MR-phase images, which hampers the CNN optimization process (red block in figure 2). This drawback was first partially addressed by the use of temperature maps as inputs for the CNN instead of the phase images. However, a voxelwise time-persistent offset remained, induced by the the presence of noise in the baseline phase image $\varphi(\vec{r}, t_0)$ in Eq. (1). This issue was addressed using an offset correction (as described in section 2.3.6), and no additional penalty in the accuracy was observable using CNN as compared to LUT and LM.

Using LM, a linear phase model is derived from the resolution of an overdetermined system of N reference images. In comparison with LUT, noise may be reduced on the resynthesized baseline phase image in Eq. (1). Ideally, a "noise-free" resynthesized baseline phase images is produced, and the noise contribution on temperature uncertainty is reduced by a factor $\sqrt{2}$. Moreover, while LUT intrinsically cannot correct for motion amplitudes higher than the ones observed in the learning stage, LM can still provide an extrapolation of the reference phase. As a consequence, regarding the precision of MR-thermometry, LM

outperformed LUT in all presented experiments. Our findings show that the multi-baseline strategy can be further improved using CNN. It must be underlined that an inherent drawback with CNN lies in the high complexity of the fitted model, which makes it difficult to interpret. However, CNN was able to cope with complex motion-related thermometry artifacts (as encountered in the upper part of the liver), for which an explicit modeling is very challenging on-line.

The CNN method provided accuracte temperature measurements for imaging frame rate above 2 Hz. The proposed CNN method was also demonstrated to be compatible with fast MR acquisition schemes of up to 10 Hz. Real-time MR-thermometry may thus be advantageously combined with any suitable real-time temporal filtering to further improve the measurement precision, as described in [22] or [31].

It should be noted that LUT and LM rely on the on-line determination of a motion surrogate, which can be provided by various types of sensors such as breathing belt or MR-/ultrasound-based surrogates. Using CNN, no motion surrogate is required since motion patterns are implicitly extracted from the actual magnitude image. CNN thereby provides an independent thermal information with that provided by LUT and LM (among others), which opens great perspectives for the use of the method as a "Watchdog" for on-line quality control (QC).

Several other machine learning models have been considered (such as Logistic Regression, Naïve Bayes, Random Forest and Support Machine Vector) to learn motion related errors in abdominal MR-thermometry in the current study. However, using such algorithms, the computation cost was consistently much higher than LM, which relies on a very simple linear model. Moreover, the above-mentioned machine learning models showed difficulties to interpolate/extrapolate positions not observed during training. The LM model, as implemented in the current study, appeared more appropriate for this task, since the pixel wise temperature variation with respiratory motion in the abdomen can be efficiently approximated in first order with a linear term, as shown in [13]. As a consequence, we decided to limit the scope of this paper to a comparison of the two already published multi-baseline methods: LUT and LM.

The main limitation of the proposed method — as it is common with multi-baseline strategies — is its inability to compensate for thermometry artifacts related to motion / deformation(s), which has not been observed during the training period. In practice, this can be encountered due to physiological drift or spontaneous motion. If during hyperthermia bulk shofts or major drifts from the calibration position are observed, a recalibration of the correction data is then required.

5. Conclusion

PRF-based MR-thermometry is complicated in abdominal organs by displacement of the target and surrounding tissues, which hampers direct voxel-by-voxel comparisons. Strong temperature artifacts are introduced by motion-induced additional phase variations via an inhomogeneous and time-variant magnetic field. The proposed approach extends the existing multi-baseline strategy using CNN to address such artifacts in abdominal organs due to breathing. A workflow is proposed to solve inherent issues with CNN associated to computational burden for training. The proposed method outperformed two existing multi-baseline strategies in terms of temperature precision, especially in regions prone to strong susceptibility artifacts as encountered in the upper part of the liver. This was achieved without noteworthy additional penalty in the temperature accuracy. Moreover, we have demonstrated that, even under clinical conditions, the method was found robust and artifact-free in all examined cases and well able to follow the temperature evolution of an *in vivo* HIFU ablation.

References

[1] H. Cline, J. Schenck, K. Hynynen, R. Watkins, S. Souza, and F. Jolesz. MR-guided focused ultrasound surgery. *J Comput Assist Tomogr*, 16(6):956–965, 1992.

[2] J. Cohen. *Statistical Power Analysis for the Behavioral Sciences*. Lawrence Erlbaum Associates, 1988.

[3] J. De Poorter, C. De Wagter, Y. De Deene, C. Thomson, F. Stahlberg, and E. Achten. The proton resonance frequency shift method compared with molecular diffusion for quantitative measurement of two dimensional time dependent temperature distribution in phantom. *Journal of Magnetic Resonance Imaging*, 103:234–241, 1994.

[4] B. Denis de Senneville, A. El Hamidi, and C. T. W. Moonen. A direct PCA-based approach for real-time description of physiological organ deformations. *IEEE Transactions on Medical Imaging*, 34(4):974–982, 2015.

[5] B. Denis de Senneville, C. Mougenot, and C. Moonen. Real time adaptive methods for treatment of mobile organs by MRI controlled high intensity focused ultrasound. *Magnetic Resonance in Medicine*, 57(2):319–330, 2007.

[6] B. Denis de Senneville, B. Quesson, and C. T. W. Moonen. Magnetic resonance temperature imaging. *International Journal of Hyperthermia*, 21(6):515–531, 2005.

[7] B. Denis de Senneville, S. Roujol, C. T. W. Moonen, and M. Ries. Motion correction in MR thermometry of abdominal organs: A comparison of the referenceless vs. the multibaseline approach. *Magnetic Resonance in Medicine*, 64(5):1373–1381, 2010.

[8] W. A. Grissom, V. Rieke, A. B. Holbrook, Y. Medan, M. Lustig, J. Santos, M. V. McConnell, and P. K. B. Hybrid referenceless and multibaseline subtraction MR thermometry for monitoring thermal therapies in moving organs. *Medical Physics*, 37(9):5014–5026, 2010.

[9] A. B. Holbrook, J. M. Santos, E. Kaye, V. Rieke, and K. Butts Pauly. Real-time MR thermometry for monitoring HIFU ablations of the liver. *Magnetic Resonance in Medicine*, 63(2):365–373, 2009.

[10] F. A. Jolesz and K. Hynynen. Magnetic resonance image-guided focused ultrasound surgery. *Cancer J*, 8(1):100–12, 2002.

[11] M. Kohler, B. Denis de Senneville, B. Quesson, C. T. W. Moonen, and M. Ries. Spectrally selective pencil-beam navigator for motion compensation of MR-guided high-intensity focused ultrasound therapy of abdominal organs. *Magnetic Resonance in Medicine*, 66(1):102–111, 2011.

[12] K. Kuroda. Non-invasive MR thermography using the water proton chemical shift. *International Journal of Hyperthermia*, 21:547–560, 2005.

[13] G. Maclair, B. Denis de Senneville, M. Ries, B. Quesson, P. Desbarats, J. Benois-Pineau, and C. T. W. Moonen. PCA-based magnetic field modeling : Application for on-line MR temperature monitoring. In N. Ayache, S. Ourselin, and A. Maeder, editors, *Medical Image Computing and Computer-Assisted Intervention, MICCAI 2007*, pages 411–419. Springer Berlin Heidelberg, 2007.

[14] J. R. McClelland, D. J. Hawkes, T. Schaeffter, and A. P. King. Respiratory motion models: a review. *Medical Image Analysis*, 17(1):19–42, 2013.

[15] N. McDannold. Quantitative MRI-based temperature mapping based on the proton resonant frequency shift: review of validation studies. *International Journal of Hyperthermia*, 21(6):533–546, 2005.

[16] S. Moricawa, T. Inubushi, Y. Kurumi, S. Naka, V. Seshan, and T. Tsukamoto. Feasibility of simple respiratory triggering in MR-guided interventional procedures for liver tumors under general anesthesia. In *ISMRM, 10th Annual Meeting, Hawaï*, 2002.

[17] A. Okada, T. Murakami, K. Mikami, H. Onishi, N. Tanigawa, T. Marukawa, and H. Nakamura. A case of hepatocellular carcinoma treated by MR-guided focused ultrasound ablation with respiratory gating. *Magnetic Resonance in Medical Sciences*, 5(3):167–171, 2006.

[18] R. D. Peters, R. Hinks, and R. M. Henkelman. Ex vivo tissue-type independence in proton-resonance frequency shift MR thermometry. *Magnetic Resonance in Medicine*, 40:454–459, 1998.

[19] V. Rieke, K. Pauly, and R. M. Henkelman. MR thermometry. *Journal of Magnetic Resonance Imaging*, 27(2):376–390, 2008.

[20] V. Rieke, K. Vigen, G. Sommer, B. Daniel, J. Pauly, and K. Butts. Referenceless PRF shift thermometry. *Magnetic Resonance in Medicine*, 51(6):1223–31, 2004.

[21] E. J. Rijkhorst, D. Heanes, F. Odille, D. Hawkes, and B. D. Simulating dynamic ultrasound using MR-derived motion models to assess respiratory synchronisation for image-guided liver interventions. *Information Processing in Computer-Assisted Interventions: Springer*, pages 113–123, 2010.

[22] S. Roujol, M. Ries, B. Quesson, C. T. W. Moonen, and B. Denis de Senneville. Real-time MR-thermometry and dosimetry for interventional guidance on abdominal organs. *Magnetic Resonance in Medicine*, 63(4):1080–7, 2010.

[23] R. Salomir, M. Viallon, A. Kickhefel, J. Roland, D. R. Morel, L. Petrusca, V. Auboiroux, T. Goget, S. Terraz, C. D. Becker, and P. Gross. Reference-free PRFS MR-thermometry using near-harmonic 2-d reconstruction of the background phase. *IEEE Transactions on Medecal Imaging*, 31(2):287-301, 2012.

[24] S. Sapareto and W. Dewey. Thermal dose determination in cancer therapy. *Int J Radiation Oncology Biol Phys*, 10:787–800, 1964.

[25] R. J. Stafford and J. Hazle. Magnetic resonance temperature imaging for focused ultrasound surgery: a review. *Top Magn Reson Imaging*, 17(3):153–63, 2006.

[26] S. Toupin, P. Bour, M. Lepetit-Coiffé, V. Ozenne, B. Denis de Senneville, R. Schneider, A. Vaussy, A. Chaumeil, H. Cochet, F. Sacher, P. Jaïs, and B. Quesson. Feasibility of real-time MR thermal dose mapping for predicting radiofrequency ablation outcome in the myocardium in vivo. *Journal of Cardiovascular Magnetic Resonance*, 19(14), 2017.

[27] K. K. Vigen, B. L. Daniel, J. Pauly, and K. Butts. Triggered, navigated, multi-baseline method for proton resonance frequency temperature mapping with respiratory motion. *Magnetic Resonance in Medicine*, 50(5):1003–1010, 2003.

[28] C. Weidensteiner, N. Kerioui, B. Quesson, B. Denis de Senneville, H. Trillaud, and C. T. W. Moonen. Stability of real-time MR temperature mapping in healthy and diseased human liver. *Journal of Magnetic Resonance Imaging*, 19(4):438–446, 2004.

[29] J. Yuan, C. S. Mei, L. P. Panych, N. J. McDannold, and B. Madore. Towards fast and accurate temperature mapping with proton resonance frequency-based MR thermometry. *Quantitative Imaging in Medicine and Surgery*, 2(1):21–32, 2012.

[30] C. Zachiu, N. Papadakis, M. Ries, C. T. W. Moonen, and B. Denis de Senneville. An improved optical flow tracking technique for real-time MR-guided beam therapies in moving organs. *Physics in Medicine and Biology*, 60(23):9003, 2015.

[31] C. Zachiu, M. Ries, C. T. W. Moonen, and B. Denis de Senneville. An adaptive non-local-means filter for real-time

MR-thermometry. *IEEE Transactions on Medical Imaging*, 36(4):904–916, 2017.

[32] Y. Zhang, S. Chen, K. Deng, B. Chen, X. Wei, J. Yang, S. Wang, and K. Ying. Kalman filtered bio heat transfer model based self-adaptive hybrid magnetic resonance thermometry. *IEEE Transactions on Medical Imaging*, 36(1):194–202, 2017.