Zero-Shot Unsupervised Motion Estimation for Motion-Corrected Cardiac T1 Mapping

Mara Guastini, Jeanette Schulz-Menger, Tobias Schaeffter, Simone Hufnagel, Andreas Kofler, Christoph Kolbitsch

***Abstract*— *Objective:* Cardiac quantitative MRI (qMRI) is a powerful imaging technique for diagnosing pathologies such as diffuse myocardial fibrosis. One main challenge is cardiac motion, which requires synchronization of data acquisition with the heartbeat, leading to long scan times. We present a novel deep learning-based image registration method for cardiac qMRI that enables non-rigid motion correction of data acquired continuously over multiple cardiac cycles, thereby reducing scan times. *Methods:* Our method is a zero-shot approach that utilizes the physical qMRI signal model for accurate motion estimation. Non-rigid motion of dynamic images is estimated with a U-Net-based architecture. This exploits the intrinsic smoothness of cardiac motion, allowing sharing information between neighboring images. The approach is robust to undersampling artifacts, enabling motion estimation from dynamic images reconstructed from very few k-space data even without advanced image reconstruction methods. *Results:* We evaluated the method for fast cardiac T1 mapping using a Golden radial sampling scheme on numerical simulations and in-vivo acquisitions. On numerical simulations, our method achieved a 61.64% improvement in T1 accuracy. On in-vivo data, our approach yielded a 45.13% improvement in sharpness of T1 maps, and temporal image alignment of motion-corrected dynamics improved on average by 11.78%. *Conclusion:* Our method enables accurate non-rigid motion correction of highly undersampled cardiac qMRI data obtained from continuously acquired data. *Significance:* As our method is individually optimized for each scan without the need for training on large datasets, it can easily be adapted to other cardiac qMRI approaches.**

***Index Terms*— cardiac qMRI, deep learning, image registration, zero-shot, dataset-free, motion correction, T1 mapping**

Date of submission: April 29th 2025. This paper was supported by the German Research Foundation (GRK2260–BIOQIC, CRC1340 Matrix in Vision).

M. Guastini, S. Hufnagel, A. Kofler, and C. Kolbitsch are with the Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany. (e-mails: mara.guastini@ptb.de, simonehufnagel@outlook.de, andreas.kofler@ptb.de, christoph.kolbitsch@ptb.de)

J. Schulz-Menger is with: the Charité Medical Faculty University Medicine, Berlin, Germany; Working Group on CMR, Experimental and Clinical Research Center (ECRC), Charité Humboldt University Berlin; DZHK (German Centre for Cardiovascular Research), partner site Berlin, Germany; the Department of Cardiology and Nephrology, HELIOS Klinikum Berlin-Buch, Berlin, Germany. (email: jeanette.schulz-menger@charite.de)

T. Schaeffter is with the Physikalisch-Technische Bundesanstalt (PTB), Braunschweig and Berlin, Germany, the Department of Medical Engineering, Technical University Berlin, Germany, and the Einstein Centre Digital Future, Berlin, Germany. (email: tobias.schaeffter@tu-berlin.de)

I. Introduction

CARDIAC quantitative MRI (qMRI) is a powerful imaging technique for diagnosing heart pathologies by evaluating the myocardial tissue, based on changes in relaxation times — T1, T2 or T2* [1]. Among these, T1 times are of particular clinical importance [2], [3], aiding in the diagnosis of conditions such as diffuse myocardial fibrosis [4].

A major challenge in cardiac qMRI is the presence of physiological and pathological (e.g., arrhythmia) motion, which is an open research area in medical imaging. Movements of the heart occurring during acquisition can lead to motion artifacts in the reconstructed images and, consequently, in the estimated parametric maps. These artifacts may result in poor image quality and inaccurate quantification of the underlying tissue parameters, leading to false diagnoses.

The main sources of motion artifacts are respiratory and cardiac motion [5]. While breath-holding techniques can often compensate for the former, cardiac motion poses a more significant challenge due to the complex nature of the beating heart. In qMRI, cardiac motion is commonly addressed with ECG-triggered acquisitions [6]–[11]. This technique reduces motion by synchronizing the image acquisition with specific phases of the cardiac cycle, ensuring consistency across scans over multiple heartbeats. However, while this approach reduces motion artifacts caused by the heartbeat, it leads to long acquisition times and may still require post-processing motion correction (MoCo). Waiting for ECG triggers and the correct cardiac phase limits data acquisition to a small fraction of the scan time, making it inefficient. Alternatively, continuous acquisition across the full cardiac cycle achieves 100% scan efficiency, acquiring the same amount of data in less scan time, but introduces a more complex MoCo challenge to estimate the full cardiac motion.

Furthermore, in qMRI, image contrast is not constant but changes over time. This is required to make the data acquisition sensitive to different quantitative parameters. Therefore, changes between images are due to a mixture of both complex non-rigid deformations from the heartbeat and signal variations required to estimate quantitative parameters.

MoCo methods in qMRI have been developed and proposed over the years, compensating for motion during image reconstruction or as a post-processing step. MoCo is typically based on image registration, which consists of the alignment of the anatomical structure of two images by applying a motion

transformation to one of them. Such alignment corresponds to the optimization problem of a function that quantifies the image match. This problem is ill-posed, as there is generally no unique motion field solution to refer to. Classical methods usually apply to standard acquisition sequences, adopting techniques such as registering images to a mean space [12], computing segmentations to align contours [13], or applying multiple registration steps for robustness [14], [15]. The method proposed by [16] is tailored to continuously acquired cardiac qMRI.

Over the past years, deep learning (DL) methods have made significant advances in image registration [13], [17]–[20], outperforming traditional algorithms in both accuracy and efficiency [21]. A notable example is the VoxelMorph model [17], which uses a U-Net-like architecture [22] to learn motion fields from moving and fixed image pairs. It is considered a golden standard method for image registration due to its simplicity and adaptability, as well as the elimination of the requirement of ground truth motion fields, whose lack in real-world scenarios presented a challenge in medical image registration. Building on this framework, SynthMorph [23] generalizes registration by training a CNN without real data, using only synthetic image pairs for robust multi-modal alignment.

In cardiac registration, existing MoCo methods generally fall into two categories. The first includes DL-based techniques for unimodal imaging, such as cine MRI, that estimate full cardiac motion (for 2D or 3D motion [24], [25], requiring fully sampled data [26], or jointly integrating segmentation with additional annotated data [27]). The second group targets MoCo in cardiac T1 mapping and is applied to standard sequences. Thus, these methods focus on correcting misalignments across diastolic phases [18], [28] or respiratory motion [19], [29], rather than capturing the full continuous cardiac motion.

Some further limitations of DL-based methods are that they typically require datasets for training and validation, which are challenging to obtain in medical imaging, and have mainly been shown to work well on high-quality images. These limitations are especially challenging for cardiac qMRI, where limited breath-hold durations and heart rate variations make the acquisition of high-quality reference images impossible. Finally, undersampling is often used to speed up data acquisition. Hence, the reconstructed images can exhibit undersampling artifacts, which impair the image quality.

In this study, we propose a novel DL non-rigid image registration method for cardiac qMRI data acquired continuously that combines the strengths of deep learning while addressing the mentioned limitations. To the best of our knowledge, this is the first method to enable DL-based non-rigid registration across the full cardiac cycle in qMRI. Our proposed registration method employs a zero-shot approach, as the network weights are optimized for each example without any pre-training, thus eliminating the need for possibly large training datasets. In addition, the underlying signal models of qMRI are exploited to overcome the challenge of contrast-changing dynamics. To estimate the motion fields, we use a U-Net architecture, similar to that of VoxelMorph. However, our approach works directly on dynamic image sequences covering, for example, multiple cardiac cycles, rather than individual image pairs, so that the network takes into account the information of neighboring time points to produce a more consistent sequence of motion fields.

The paper is structured as follows. First, Section 2 introduces the mathematical framework of the proposed approach. In Section 3, we then present our method, where the algorithm for the registration task is explained, and details on the network, minimizing loss, and generation of motion-free synthetic references are provided. Also, the motion-corrected parameter estimation process is explained. Sections 4 and 5 introduce, respectively, the experimental setup and the obtained results. Finally, the main findings from the results and the contributions of the proposed method are discussed in Section 6.

II. Problem Formulation

In qMRI, data are obtained with different image contrasts over time, depending on MR acquisition parameters and tissue properties. A *dynamic* $\mathbf{x} = (\mathrm{x}_1, \ldots, \mathrm{x}_T)^\top \in \mathbb{R}^{T \cdot N}$, for $T \geq 1$ times and $N \geq 1$ voxels, is a sequence of images, commonly reconstructed at a generalized set of acquisition times $\mathcal{T} = \{t_1, \ldots, t_T\}$, where the evolution of contrast can described by a non-linear *signal model* $\mathbf{Q}_\mathcal{T}$

$$\mathbf{Q}_\mathcal{T} : \mathbb{R}^{P \cdot N} \to \mathbb{R}^{T \cdot N}, \qquad \mathbf{p} \mapsto \mathbf{Q}_\mathcal{T}(\mathbf{p}). \tag{1}$$

The model $\mathbf{Q}_\mathcal{T}$ is a mapping from the parametric space into the image space, and maps P parameters $\mathbf{p} = (\mathrm{p}_1, \ldots, \mathrm{p}_P)^\top$ into a dynamic $\mathbf{Q}_\mathcal{T}(\mathbf{p})$.

In an ideal motion-free scenario, reconstructed dynamics are entirely described by an appropriate physical model, i.e. $\mathbf{x} = \mathbf{Q}_\mathcal{T}(\mathbf{p})$, for some unknown parameters $\mathbf{p}$. However, in real cases, image content changes due to the motion that appears during the acquisition. Therefore, the actual reconstructed moving dynamic $\mathbf{x}^\mathrm{m}$ is described by

$$\mathbf{x}^\mathrm{m} = \mathbf{M}_{\boldsymbol{\phi}} \mathbf{Q}_\mathcal{T}(\mathbf{p}), \tag{2}$$

where $\mathbf{M}_{\boldsymbol{\phi}}$ is the *motion operator* that applies a sequence of motion fields $\boldsymbol{\phi} = (\phi_1 \ldots, \phi_T)^\top \in \mathbb{R}^{T \cdot N \cdot D}$ to a dynamic using local spatial interpolation kernels, with $D = 2, 3$ differing between 2D and 3D motion fields. Here we assume that the motion only occurs between acquisition time points but not during data acquisition. This approximation is justified as the acquisition time steps are usually short compared to the physiological motion process. For a dynamic $\mathbf{x}$, each image is warped by the respective motion field, i.e. $\mathbf{M}_{\boldsymbol{\phi}}\mathbf{x} := (\mathbf{M}_{\phi_1}\mathrm{x}_1, \ldots, \mathbf{M}_{\phi_T}\mathrm{x}_T)^\top$. $\mathbf{M}_{\boldsymbol{\phi}}$ is also known as a spatial transformer, and we refer the reader to [30, Section 3] for further details. If these motion transformations are not taken into consideration during the estimation of the parameters $\mathbf{p}$, motion artifacts will occur.

The complexity of $\mathbf{Q}_\mathcal{T}$ together with $\mathbf{M}_{\boldsymbol{\phi}}$ results in an ill-posed problem when estimating the motion directly from (2). Thus, we proceed in the following way to simplify the problem formulation.

We start by considering a reconstructed moving dynamic $\mathbf{x}^\mathrm{m}$. The goal is to estimate the motion present in $\mathbf{x}^\mathrm{m}$ over

time, by aligning the dynamic to a motion-free setting. Thus, in particular, we want to recover motion fields ϕ^*, with

$$\mathbf{M}_{\phi^*}\mathbf{x}^{\mathrm{m}} = \mathbf{x}^{\mathrm{r}}, \tag{3}$$

for $\mathbf{x}^{\mathrm{r}} := \mathbf{Q}_{\mathcal{T}}(\mathbf{p})$ being the ideal motion-free underlying dynamic.

In this context, motion fields ϕ^* can now be thought of as the inverse motion present in (2). However, no inverse property is assumed since, in real-case scenarios, such an assumption may not hold. For instance, when dealing with 2D cardiac images, estimating 2D in-plane motion fields does not consider existing through-plane motion taking place during heart systole.

Estimating the motion fields in (3) corresponds to optimize the alignment of $\mathbf{M}_{\phi}\mathbf{x}^{\mathrm{m}}$ to $\mathbf{x}^{\mathrm{r}}$ through the minimization of an appropriate energy function f with respect to motion fields ϕ:

$$\phi^* \in \underset{\phi \in \mathbb{R}^{T\cdot N\cdot D}}{\arg\min}\ f\left(\mathbf{M}_{\phi}\mathbf{x}^{\mathrm{m}}, \mathbf{x}^{\mathrm{r}}, \phi\right). \tag{4}$$

The choice of f depends on the applications. In general, it aims at addressing two main aspects: quantifying the image alignment and controlling the spatial regularity of the fields ϕ to maintain the most natural and realistic deformations possible, and ensuring robustness in the presence of noise and artifacts.

Problem (4) is also called a registration problem, and the resulting $\mathbf{M}_{\phi^*}\mathbf{x}^{\mathrm{m}}$ is the registered dynamic $\mathbf{x}^{\mathrm{m}}$ aligned to $\mathbf{x}^{\mathrm{r}}$.

III. Proposed Method

The presented method addresses solving the registration problem (4) with a dataset-free zero-shot unsupervised DL iterative approach (Section A). The registration takes place simultaneously for all time steps of a given moving dynamic $\mathbf{x}^{\mathrm{m}}$, resulting in a sequence of optimal motion fields ϕ^* after an iterative process.

A U-Net architecture is used to estimate the motion fields, with weights initialized for each scan. Indeed, optimal fields ϕ^* are parameterized as the output of a CNN $\mathbf{u}_{\theta}$ (Section B) that takes as input the moving dynamic $\mathbf{x}^{\mathrm{m}}$, i.e.

$$\phi^* := \mathbf{u}_{\theta^*}(\mathbf{x}^{\mathrm{m}}), \tag{5}$$

for θ^* being optimal network weights. In the network optimization, a loss function is selected to ensure both accurate image similarity and motion regularity (Section C). In particular, the alignment term of the loss uses some motion-compensated reference dynamic $\mathbf{x}^{\mathrm{r}}$ that is calculated prior to the image registration by exploiting the underlying physical signal model and computing preliminary parameters $\tilde{\mathbf{p}}$ (Section D).

The method is applied to quantitative parameters by formulating a fitting problem for motion-corrected parametric maps estimation using the recovered optimal motion fields ϕ^* (Section E).

Fig. 1 shows a depiction of the method, comprising both the registration process and motion-corrected parameter estimation.

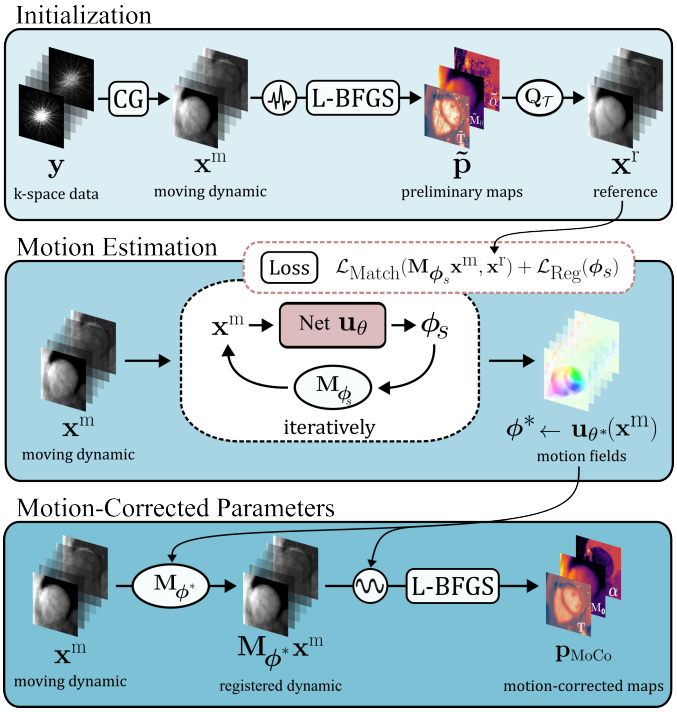

Fig. 1: Overview of the proposed registration method. The first row shows the initialization steps. Firstly, image reconstruction is carried out with the iterative SENSE method [31] solving $\min_{\mathbf{x}}\|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2^2$, for k-space data $\mathbf{y}$ and acquisition operator $\mathbf{A}$. Secondly, preliminary parameters $\tilde{\mathbf{p}}$ are estimated to compute $\mathbf{x}^{\mathrm{r}}$ by solving (10). The second row is a depiction of the registration algorithm. From the moving dynamic $\mathbf{x}^{\mathrm{m}}$ as input, the network $\mathbf{u}_{\theta}$ is trained iteratively as explained in Algorithm 1 yielding optimal motion fields ϕ^*. The last row describes the motion-corrected parameter estimation. 15% of the motion-corrected dynamic $\mathbf{M}_{\phi^*}\mathbf{x}^{\mathrm{m}}$ comprising the systolic phase are excluded from the fitting problem to estimate $\mathbf{p}_{\mathrm{MoCo}}$.

A. Registration Process

The registration occurs iteratively for E epochs by optimizing weights θ through gradient descent on the loss function $\mathcal{L}$. At the end of the iterations, the method returns the optimal motion fields $\phi^* = \mathbf{u}_{\theta^*}(\mathbf{x}^{\mathrm{m}})$ for a given dynamic $\mathbf{x}^{\mathrm{m}}$. Through this method, the weights θ are optimized specifically for each input dynamic $\mathbf{x}^{\mathrm{m}}$ without requiring any training dataset.

The input for the network $\mathbf{u}_{\theta}$ is initialized with $\mathbf{x}^{\mathrm{m}}$ at the beginning of every epoch, each composed of S consecutive steps, where the input are the intermediate warped dynamics. During each step s, the current input is fed into the neural network, and motion fields (ϕ_s) are estimated. The loss is computed, and the weights are updated accordingly. In the subsequent step $s+1$, the input to the network becomes the dynamic $\mathbf{x}^{\mathrm{m}}$ warped with the previous network's output ($\mathbf{M}_{\phi_s}\mathbf{x}^{\mathrm{m}}$), detached from its previous gradient. This allows the network to improve learning of the existing motion by also considering intermediate steps, i.e. the network eventually learns large deformations by a series of small transformations. Motion fields are nevertheless always applied to the original input dynamic $\mathbf{x}^{\mathrm{m}}$, so no concatenation of motion fields is

needed, which could lead to error accumulation. Note that, in Algorithm 1, although the input to the network changes at each iteration, the loss, defined in (III-C), is always computed by applying the predicted motion field to the initial moving images $\mathbf{x}^{\mathrm{m}}$. Thus, the task of the network remains consistent throughout training. If the loss does not decrease for a fixed number of C consecutive steps, the next epoch is started. At each new epoch, the input is initialized again as $\mathbf{x}^{\mathrm{m}}$, and the weights of $\mathbf{u}_\theta$ continue to be further updated.

By the end of this process, the network has gradually learned to estimate the motion from its input dynamic. The method returns the optimal motion fields. The full process is outlined in Algorithm 1.

Algorithm 1 Registration Algorithm

Initialization:
Given: epochs E, steps S, and input dynamic $\mathbf{x}^{\mathrm{m}}$
Compute reference dynamic $\mathbf{x}^{\mathrm{r}}$ from (10)
for each epoch $e = 1, \dots, E$ **do**
  Set initial input: $\mathbf{x}^{\mathrm{m}}_0 \leftarrow \mathbf{x}^{\mathrm{m}}$
  for each step $s = 1, \dots, S$ **do**
    Compute motion fields: $\boldsymbol{\phi}_s \leftarrow \mathbf{u}_\theta(\mathbf{x}^{\mathrm{m}}_{s-1})$
    Update parameters: $\theta \leftarrow \theta - \nabla_\theta \mathcal{L}(\mathbf{u}_\theta(\mathbf{x}^{\mathrm{m}}_{s-1}))$
    Update network input: $\mathbf{x}^{\mathrm{m}}_s \leftarrow \mathbf{M}_{\boldsymbol{\phi}_s}\mathbf{x}^{\mathrm{m}}$
    Early step stopping based on loss improvement
  end for
end for
Get optimal motion fields $\boldsymbol{\phi}^* \leftarrow \mathbf{u}_{\theta^*}(\mathbf{x}^{\mathrm{m}})$
Return optimal fields $\boldsymbol{\phi}^*$

B. Network Structure

The network $\mathbf{u}_\theta$ takes a 2D dynamic sequence $\mathbf{x}^{\mathrm{m}}$ as input and outputs a sequence of 2D motion fields $\boldsymbol{\phi}$. It uses a 2D Attention U-Net architecture [22], [32], where each block contains a 2D convolution, batch normalization, and LeakyReLU activation. Temporal frames in $\mathbf{x}^{\mathrm{m}}$ are treated as input channels (shape $(1, T, H, W)$), with the number of input/output channels determined by the number of frames. Additive attention gates help localize motion across spatial and temporal dimensions. As in VoxelMorph [17], three final convolutions and a scaling-and-squaring step ensure diffeomorphic outputs.

We use magnitude dynamics, though real and imaginary components could also be provided as separate channels.

C. Loss Function

To update the network weights θ, we use the following loss function. Given a moving dynamic $\mathbf{x}^{\mathrm{m}}$, motion fields $\boldsymbol{\phi}$, and a reference dynamic $\mathbf{x}^{\mathrm{r}}$, let $\mathcal{L}$ be defined as

$$\mathcal{L}(\boldsymbol{\phi}) := \mathcal{L}_{\text{Match}}\left(\mathbf{M}_{\boldsymbol{\phi}}\mathbf{x}^{\mathrm{m}}, \mathbf{x}^{\mathrm{r}}\right) + \mathcal{L}_{\text{Reg}}\left(\boldsymbol{\phi}\right). \tag{6}$$

Accordingly, finding a solution of (4) corresponds to finding optimal motion fields $\boldsymbol{\phi}^* = \boldsymbol{\phi}_{\theta^*}(\mathbf{x}^{\mathrm{m}})$ corresponding to optimal weights θ^* such that

$$\theta^* \in \arg\min_\theta \mathcal{L}\left(\mathbf{u}_\theta(\mathbf{x}^{\mathrm{m}})\right). \tag{7}$$

The two terms in (6), which address the image alignments and the regularity of the motion fields, are discussed next.

1) Matching Term: The alignment between two dynamics is quantified by comparing images at corresponding times. The matching term in (6) is defined as

$$\mathcal{L}_{\text{Match}}(\mathbf{M}_{\boldsymbol{\phi}}\mathbf{x}^{\mathrm{m}}, \mathbf{x}^{\mathrm{r}}) = \sum_{t=1}^{T} \mathrm{MI}\left(\mathbf{M}_{\phi_t}\mathrm{x}^{\mathrm{m}}_t; \mathrm{x}^{\mathrm{r}}_t\right), \tag{8}$$

where MI denotes the mutual information loss between two images [33]–[36].

2) Regularization Term: The second term in (6) deals with the spatial regularity of the motion fields $\boldsymbol{\phi}$. In particular, the intention is to encourage smooth and natural deformations, ensuring robustness to noise and potential image artifacts.

The regularization term is defined as

$$\mathcal{L}_{\text{Reg}}(\boldsymbol{\phi}) = \sum_{t=1}^{T} \left(\lambda \left\|\nabla\phi_t\right\|_2^2 + \mu \left\|\nabla\cdot\phi_t\right\|_2^2\right), \tag{9}$$

with $\lambda, \mu > 0$, ∇ and $\nabla\cdot$ being the gradient and divergence operators computed via finite differences. These two norms address the elastic properties of a displacement field ϕ differently. The gradient $\nabla\phi$ shows the rate of directional change, reflecting how much neighboring points move relative to each other, affecting the shape of the warped object. Thus, penalizing the gradient results in suppressing sharp and unrealistic local distortions. On the other hand, the divergence $\nabla\cdot$ captures the local changes in the area of the warped object [37, Part III, Chapter 8], either by contraction (negative divergence) or by expansion (positive divergence), and penalizing it reduces unphysiological motion transformations such as folding.

D. Motion-Compensated References Computation

A motion-compensated reference dynamic $\mathbf{x}^{\mathrm{r}}$ is computed in the initialization steps of Algorithm 1 to be used in the loss function (6). The following approach builds upon previous work estimating parameter maps with motion compensation [16] using info from ECG recorded during scan time.

The idea is to exploit physical knowledge coming from the acquisition, by finding some preliminary parameters $\tilde{\mathbf{p}}$ to generate some synthetic dynamic starting from the moving dynamic $\mathbf{x}^{\mathrm{m}}$.

As the motion in $\mathbf{x}^{\mathrm{m}}$ leads to blurring artifacts in the solution $\tilde{\mathbf{p}}$, a subset of the dynamic $\mathbf{x}^{\mathrm{m}}$ is used for this preliminary fitting, consisting of images from $\mathbf{x}^{\mathrm{m}}$ in a similar motion state (e.g. diastole, which can be detected from the ECG).

Ultimately, we carry out a voxel-wise fit on the following minimization problem

$$\tilde{\mathbf{p}} \in \arg\min_{\mathbf{p}\in\mathcal{A}} \left\|\mathbf{Q}_{\mathcal{T}_{\text{sel}}}(\mathbf{p}) - \mathbf{x}^{\mathrm{m}}_{\text{sel}}\right\|_2^2. \tag{10}$$

In (10), $\mathcal{A}$ denotes a set of admissible values for the parameters $\mathbf{p}$, with admissibility being context-dependent (e.g., T1 times must be positive). The subset of acquisition times $\mathcal{T}_{\text{sel}} \subseteq \mathcal{T}$ consists of $T_{\text{sel}} < T$ acquisition times, chosen to exclude phases with large motion deformations (e.g., systole). Finally, $\mathbf{x}^{\mathrm{m}}_{\text{sel}}$ is defined from $\mathbf{x}^{\mathrm{m}}$ by selecting only the times from $\mathcal{T}_{\text{sel}}$. Given the differentiability and non-linearity of the signal

model $\mathbf{Q}_\mathcal{T}$, a solution of (10) can be handled for instance with the limited memory BFGS method (L-BFGS) [38].

This fit uses only a small portion of diastolic frames selected via ECG, yielding a preliminary T1 map less affected by motion. As a result, the predicted vector of parameters $\tilde{\mathbf{p}}$ can be used to generate a motion-compensated dynamic $\mathbf{x}^\mathrm{r}$ to be used as reference for the registration task, with $\mathbf{x}^\mathrm{r} := \mathbf{Q}_\mathcal{T}(\tilde{\mathbf{p}})$ sharing similar contrast change with $\mathbf{x}^\mathrm{m}$, but with strongly reduced motion artifacts.

E. Motion-Corrected Parameter Estimation

In the process of estimating parametric maps from the registered dynamic $\mathbf{M}_{\boldsymbol{\phi}^*}\mathbf{x}^\mathrm{m}$, 15% of the dynamic around systole are excluded from the fitting problem, as suggested by [16], [39] and based on assumptions from [40], [41], to reduce inaccuracies in T1 values caused by through-plane motion during systole, which can reach approx. 7 mm at the mid-ventricle and 13 mm at the base [42]. Such exclusion is carried out automatically exploiting the fact that highest amplitudes of the estimated motion fields $\boldsymbol{\phi}^*$ help to identify the systolic motion. To achieve this, the mean amplitudes $A = (A_1, \dots, A_T)^\top \in \mathbb{R}^T$ of the respective motion fields $\boldsymbol{\phi}^* = (\phi_1^*, \dots, \phi_T^*)^\top$ are first computed. In particular, for each time point t, we have

$$A_t := \frac{1}{N}\sum_{x=1}^{N} \|M_\mathrm{corr}\phi_t^*(x)\|_2^2\,, \tag{11}$$

where $M_\mathrm{corr} = \|\frac{1}{T}\sum_{t=1}^{T-1}\phi_t^*\phi_{t+1}^*\|_2^2$ is a map used to focus the amplitude values on regions where the fields show temporal correlation, under the assumption that the motion in the heart region is more temporally dependent. Secondly, we proceed with a sinusoidal fit of the amplitudes A and 15% of the dynamic $\mathbf{M}_{\boldsymbol{\phi}^*}\mathbf{x}^\mathrm{m}$ closest to the sinusoidal peaks are excluded, assuming that the peaks reflect the systole. We call the remaining 85% of the registered dynamic $[\mathbf{M}_{\boldsymbol{\phi}^*}\mathbf{x}^\mathrm{m}]_{85\%}$, with respective acquisition times $\mathcal{T}_{85\%}$.

Finally, the motion-corrected parameters $\mathbf{p}_\mathrm{MoCo}$ are given by

$$\mathbf{p}_\mathrm{MoCo} \in \underset{\mathbf{p}\in\mathcal{A}}{\arg\min} \left\|\mathbf{Q}_{\mathcal{T}_{85\%}}(\mathbf{p}) - [\mathbf{M}_{\boldsymbol{\phi}^*}\mathbf{x}^\mathrm{m}]_{85\%}\right\|_2^2, \tag{12}$$

where $\mathcal{A}$ denotes the same constraint values as in (10). Again, problem (12) is solved via the L-BFGS. Indeed, the objective function in (12) inherits its non-linearity and differentiability with respect to $\mathbf{p}$ from the signal model $\mathbf{Q}_\mathcal{T}$. This, together with the computability of its gradients, makes the problem well-suited for a quasi-Newton method like L-BFGS.

IV. Experiments

A. Datasets

Our method was evaluated for cardiac T1 mapping on simulated as well as on single-slice 2D and multi-slice 2D in-vivo acquisitions. Data acquisition was carried out on a 3T (Verio, Siemens Healthineers, Erlangen, Germany) and ECG was recorded during scan.

A 2D Golden-angle radial acquisition scheme was used to continuously obtain data for 2.3 s after a single inversion pulse. The acquisition parameters were: 5° flip angle, 2.03 ms TE, and 5 ms TR. The field of view (FOV) was 320 mm × 320 mm, with a spatial resolution of 1.3 mm × 1.3 mm and a slice thickness of 8 mm. The beginning of the data acquisition was triggered to mid-diastole. Dynamics consisted of 54 images for the simulated data and the single-slice 2D dataset, and of 49 images for the multi-slice 2D datasets.

1) Simulated Data: XCAT, a framework for numerical phantoms, was used to create a tissue segmentation in short-axis orientation and the corresponding motion fields ($\boldsymbol{\phi}^\mathrm{XCAT}$) [43], [44], describing the heart deformation during a cardiac cycle at 100 bpm. Quantitative parameters were assigned to the tissue segmentation provided by XCAT and, after transforming them to signal intensities using $\mathbf{Q}_\mathcal{T}$, the motion fields $\boldsymbol{\phi}^\mathrm{XCAT}$ were used to transform the images to different cardiac motion states. MR data acquisition was then simulated to create k-space data corresponding to an in-vivo data acquisition.

2) Single-slice 2D SA: For 10 healthy volunteers (4 females, 6 males, age: 32 ± 8 y), 2D short-axis (SA) slices in the mid-ventricular myocardium were acquired in a single breath-hold with non-selective inversion pulse [16]. Three volunteers were scanned twice.

3) Multi-slice 2D SA and LA: Data from a group of 10 healthy subjects (7 males, 3 females, age: 30.3 ± 2.28 y) were acquired. In each subject, 12 interleaved stacks of 5 two dimensional slices each were acquired continuously after slice-selective RF inversion pulses in a single breath-hold, with the same acquisition parameters as above, except for a flip angle of 9° and a TE of 2.2 ms. Data were acquired as SA stacks covering the apex, apical, mid-cavity, and basal views and long-axis (LA) stacks, covering four-chambers (4CH) and left and right two-chambers (2CH) ventricle views [45].

As for T1 values references, 3(3)3(3)5 Modified Look-Locker Inversion Recovery (MOLLI) [7] measurements were also obtained for all in-vivo datasets, with acquisition time of 17 s at a heart rate of 60 bpm. with acquisition parameters: 2.7 ms TR, 1.12 ms TE, and 35° flip angle. The spatial resolution was of 2.1 mm × 1.4 mm × 6.0-8.0 mm and the FOV was 360 mm × 307-323 mm.

The research was approved by the institution's ethical committee ('Ethikkommission der PTB'). The research was conducted in accordance with the principles embodied in the Declaration of Helsinki and in accordance with local statutory requirements. All subjects gave written informed consent before participation.

B. Signal Model

A Look-Locker model [46] $\mathbf{Q}_\mathcal{T}$ was used to estimate the synthetic motion-compensated reference $\mathbf{x}^\mathrm{r}$ and parameters $\mathbf{p}_\mathrm{MoCo}$. It is a function of the parameters $\mathbf{p}$, which consist of the equilibrium magnetization $\mathbf{M}_0$, readout flip angle $\boldsymbol{\alpha}$, and $\mathbf{T}_1$ values. At acquisition time t, $\mathbf{Q}_t$ was defined as the following three-parameter model

$$\mathbf{Q}_t(\mathbf{p}) = \mathbf{M}_0^* - (\mathbf{M}_0 + \mathbf{M}_0^*)\, e^{-t\frac{1}{\mathbf{T}_1^*}} \tag{13}$$

$$\mathbf{T}_1^* = \left[\mathbf{T}_1^{-1} - T_R^{-1}\log(\cos\boldsymbol{\alpha})\right]^{-1} \tag{14}$$

$$\mathbf{M}_0^* = \mathbf{T}_1^{-1}\mathbf{M}_0\mathbf{T}_1^*, \tag{15}$$

for $\mathbf{p} = (\mathbf{M}_0, \boldsymbol{\alpha}, \mathbf{T}_1)^\top$ and repetition time T_R.

Thus motion-corrected parametric maps $\mathbf{p}_{\text{MoCo}}$ consisted of $\mathbf{p}_{\text{MoCo}} = (\mathbf{M}_{0,\text{MoCo}}, \boldsymbol{\alpha}_{\text{MoCo}}, \mathbf{T}_{1,\text{MoCo}})^\top$. In the following, we focus our evaluations specifically on the motion-corrected T1 maps $\mathbf{T}_{1,\text{MoCo}}$ as it is the clinically relevant parameter.

C. Synthetic References

To compute synthetic dynamics from (10), the model $\mathbf{Q}_\mathcal{T}$ from (13) together with L-BFGS as optimization routine.
However, in the multi-slice 2D SA and LA datasets, the slice-selective inversion pulse led to the mixing of inverted and incoming non-inverted blood, resulting in a contrast that could not be captured well by the original model in (13). For this reason, in the computation of the reference dynamics of the multi-slice 2D SA and LA datasets, we used a corrected model $\mathbf{Q}_{t,\boldsymbol{\beta}}$ to avoid errors in the registration task due to the presence of non-inverted blood flux. In this context, we used a modified model $\mathbf{Q}_{t,\boldsymbol{\beta}}$, defined as

$$\mathbf{Q}_{t,\boldsymbol{\beta}}\left(\mathbf{M}_0, \boldsymbol{\alpha}, \mathbf{T}_1, \boldsymbol{\beta}\right) = \mathbf{M}_0^* - \left(\boldsymbol{\beta}\mathbf{M}_0 + \mathbf{M}_0^*\right) e^{-t\frac{1}{\mathbf{T}_1^*}}, \quad (16)$$

where $\boldsymbol{\beta} < 1$ describes a mix of inverted and inflowing non-inverted blood.

D. Experimental Setup

For all datasets, moving dynamics were reconstructed using the iterative SENSE algorithm [31], a standard conjugate gradient (CG) method for image reconstruction without any further regularization. Motion estimation was performed within the imaging plane, using bicubic interpolation for the motion operator.
For the parameters fitting of motion-compensated references $\mathbf{x}^{\text{r}}$, values were constrained to $\boldsymbol{\alpha} \in (1°, 9°)$, $\mathbf{T}_1 \in (0.18\text{ s}, 6.66\text{ s})$.
In the registration process, the network $\mathbf{u}_\theta$ (64 filters, 5 stages, 49 or 54 input channels – depending on the dataset) was trained for a fixed number of iterations ($E = 80$ epochs and $S = 20$ steps), with loss function's weights $\lambda = 0.05$, $\mu = 0.01$, Adam optimizer, and learning rate of 0.0001.

E. Evaluation

1) Method of Comparison: We compared the proposed registration method to the motion estimation of the method presented in [16]. Their motion estimation consisted of three main steps. For each scan, a dynamic was first reconstructed with spatial and temporal total variation (TV) regularization [47] to deal with minimized undersampling artifacts. Then, a synthetic reference dynamic was computed using the exponential signal models by including diastolic phases from the reconstructed dynamic. Finally, the estimation of motion fields was carried out via the MIRTK Toolkit [48]. This method assumed regular heartbeats by considering the motion vector field of a cardiac phase to be the same in each cardiac cycle. Thus, motion fields referred to an averaged cardiac cycle and not to different acquired images. In the following, we refer to this method as MIRTK-T1Reg.

2) Ablation Study: We conducted an ablation study focusing on the two main components of our method: the signal model and the CNN-based parametrization of motion fields. First, we kept the parametrization but removed the signal model (i.e., removed external reference $\mathbf{x}^{\text{r}}$). For this, we used a 2D SynthMorph, the best choice available in this context, to register $\mathbf{x}^{\text{m}}$ to a diastolic frame. Second, we kept the signal model but replaced the CNN by directly optimizing the loss function with respect to the motion field tensors (DirectOpt).

3) Metrics: To assess the registration quality, we evaluated two metrics within a manually selected region of interest (ROI) around the ventricular septum in SA view. The ROI ensured that our analysis focused specifically on motion artifacts within the heart. The sharpness of $\mathbf{T}_{1,\text{MoCo}}$ was quantified using the Tenengrad metric (T) [49], a well-established approach for evaluating image sharpness through finite differences in the absence of ground truth images [50], [51].

To evaluate the temporal consistency of the registered dynamics, we computed temporal TV [52]. Given a dynamic $\mathbf{x} \in \mathbb{R}^{T \cdot N}$, temporal TV is defined as

$$\text{TV}_{\text{Time}}(\mathbf{x}) = \frac{1}{N(T-1)} \sum_{t=1}^{T-1} \sum_{i=1}^{N} |\text{x}_t(i) - \text{x}_{t+1}(i)|.$$

Since dynamics consist of regions with intensities that change smoothly over time, an effective registration should yield a low TV_{Time}. For comparison, we analyzed the relative change in both metrics, computing $(\text{T} - \text{T}_{\text{NM}})/\text{T}_{\text{NM}}$ for sharpness and $(\text{TV}_{\text{Time,NM}} - \text{TV}_{\text{Time}})/\text{TV}_{\text{Time,NM}}$ for temporal consistency, where T_{NM} and $\text{TV}_{\text{Time,NM}}$ refer to the respective metrics on the data without motion correction (No MoCo). This formulation ensures that improvements in both metrics yield positive values.

V. Results

A. Numerical Simulation

Fig. 2 illustrates a comparison of the motion fields $\boldsymbol{\phi}^*$ computed using the proposed method against the motion fields $\boldsymbol{\phi}^{\text{XCAT}}$ from the XCAT model for simulated data. For three different acquisition times, corresponding to different phases during the cardiac cycle, images of the moving dynamic ($\mathbf{x}^{\text{m}}$) and the registered dynamics ($\mathbf{M}_{\boldsymbol{\phi}^*}\mathbf{x}^{\text{m}}$, $\mathbf{M}_{\boldsymbol{\phi}^{\text{XCAT}}}\mathbf{x}^{\text{m}}$) are shown.

In Fig. 3, a comparison of T1 maps is presented between T1 maps without motion correction and the motion-corrected maps $\mathbf{T}_{1,\text{MoCo}}$, obtained registering the dynamic with MIRTK-T1Reg, the proposed method and, finally, the reference motion fields from XCAT. The second row displays the difference between the recovered maps and the reference T1 map, providing a quantitative evaluation of the accuracy of each MoCo approach. On a manually selected ROI around the entire heart, the mean difference is -74.06 $\pm$ 91.81 ms without motion correction, -28.24 $\pm$ 43.60 ms registering with MIRTK-T1Reg, and -28.41 $\pm$ 46.04 ms registering with the proposed approach.

B. In-Vivo

Fig. 4 shows a comparison between preliminary and final motion-corrected T1 values.

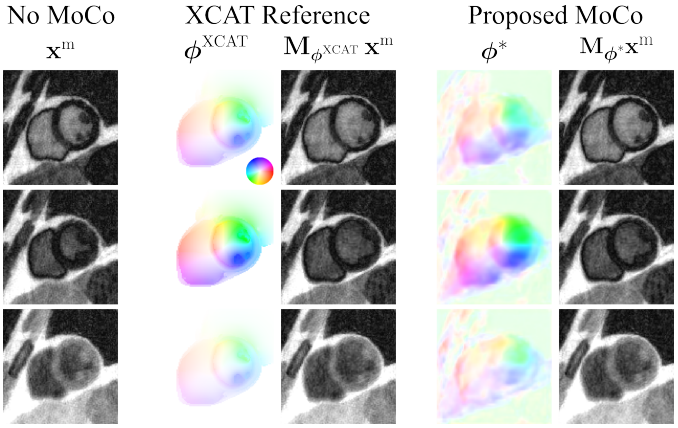

Fig. 2: Images from the input dynamic ($\mathbf{x}^{\mathrm{m}}$), motion fields and motion-transformed images shown for the numerical simulation using XCAT (ϕ^{XCAT}, $\mathbf{M}_{\phi^{\mathrm{XCAT}}}\mathbf{x}^{\mathrm{m}}$) and the proposed method (ϕ^{*}, $\mathbf{M}_{\phi^{*}}\mathbf{x}^{\mathrm{m}}$) . Motion fields are visualized using a color wheel representation, which qualitatively indicates the direction (color) and amplitude (intensity) of motion. Sharp changes in the color values of ϕ^{XCAT} reflect the irregularity of the fields at the boundaries.

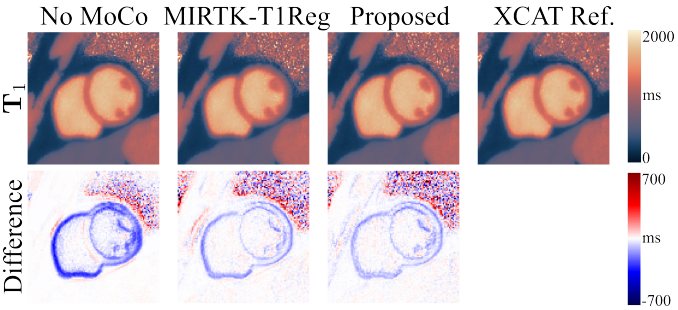

Fig. 3: T1 maps estimation on simulated data. T1 maps are displayed without and with motion correction ($\mathbf{T}_{1,\mathrm{MoCo}}$), by registering with MIRTK-T1Reg, the proposed method, and the motion fields from XCAT (used as reference). The second row shows the differences to the reference T1 map.

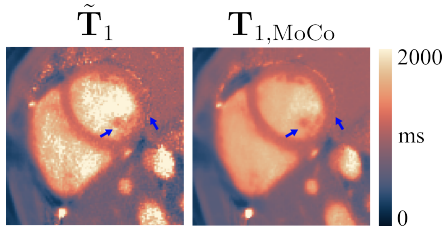

Fig. 4: Example of T1 values achieved from the preliminary fitting (10), with $\tilde{\mathbf{p}} = (\tilde{\mathbf{T}}_1, \tilde{\mathbf{M}}_0, \tilde{\boldsymbol{\alpha}})$, and from the final motion-corrected fitting (12). Arrows indicate improved details.

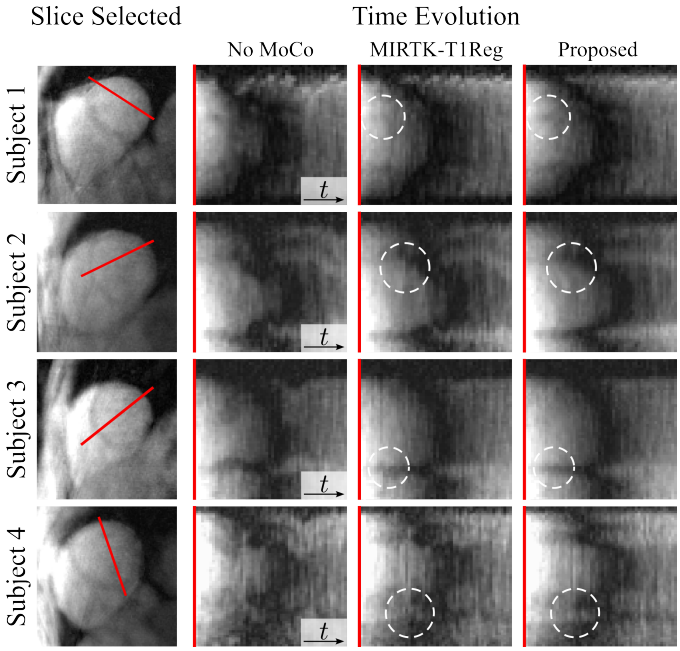

Fig. 5: Slice-view of registered dynamics of the single-slice 2D SA dataset for four different volunteers. The red line indicates the selected voxel whose temporal evolution is shown for the case without motion correction, registered with MIRTK-T1Reg, and with the proposed approach. Circles highlight the improved alignment achieved by the proposed method, where residual motion artifacts remain visible in the MIRTK-T1Reg registration.

Fig. 5 shows the results of the registration on the single-slice 2D SA dataset by displaying the temporal evolution of a line profile indicating the selected voxel for four different subjects, with voxel intensities shown across all consecutive time points in the respective dynamics. From left to right, the time evolution is presented for the moving dynamic $\mathbf{x}^{\mathrm{m}}$ (showing the contraction and expansion of the heart during the cardiac cycles), followed by the registration results obtained with MIRTK-T1Reg, and those from the proposed method.

T1 maps for four healthy volunteers are presented in Fig. 6. Here, we compare maps without motion correction, $\mathbf{T}_{1,\mathrm{MoCo}}$ using MIRTK-T1Reg and the proposed approach, respectively, and T1 maps from the clinical reference sequence (MOLLI). Fig. 6A) shows results from the single-slice 2D SA dataset, while Fig. 6B) from the multi-slice 2D SA dataset. Fig.7 shows the results of the ablation study.

Moreover, Fig. 9 presents the registration results for a multi-slice acquisition of the same subject in both 4CH and 2CH views. Different time steps are shown for images of both the moving and registered dynamics. Also, a slice segment is selected for both views to illustrate its temporal evolution. To further assess the generalization of the method, Fig. 10 compares T1 maps without motion correction to motion-corrected T1 maps $\mathbf{T}_{1,\mathrm{MoCo}}$ for the multi-slice 2D SA dataset of one healthy volunteer, covering basal, mid-ventricular, apical, and apex views. Similarly, Fig. 11 displays such comparison of T1 maps for the multi-slice 2D LA dataset over four healthy volunteers, including both 4CH and 3CH views.

For the evaluation of the registration methods, we assessed performance on single-slice 2D SA dataset (10 cases) and mid-ventricular slices from the multi-slice 2D SA dataset (120 cases). MIRTK-T1Reg was compared to the proposed approach, with Wilcoxon tests and t-tests performed to evaluate the differences (on single- and multi-slice 2D SA dataset,

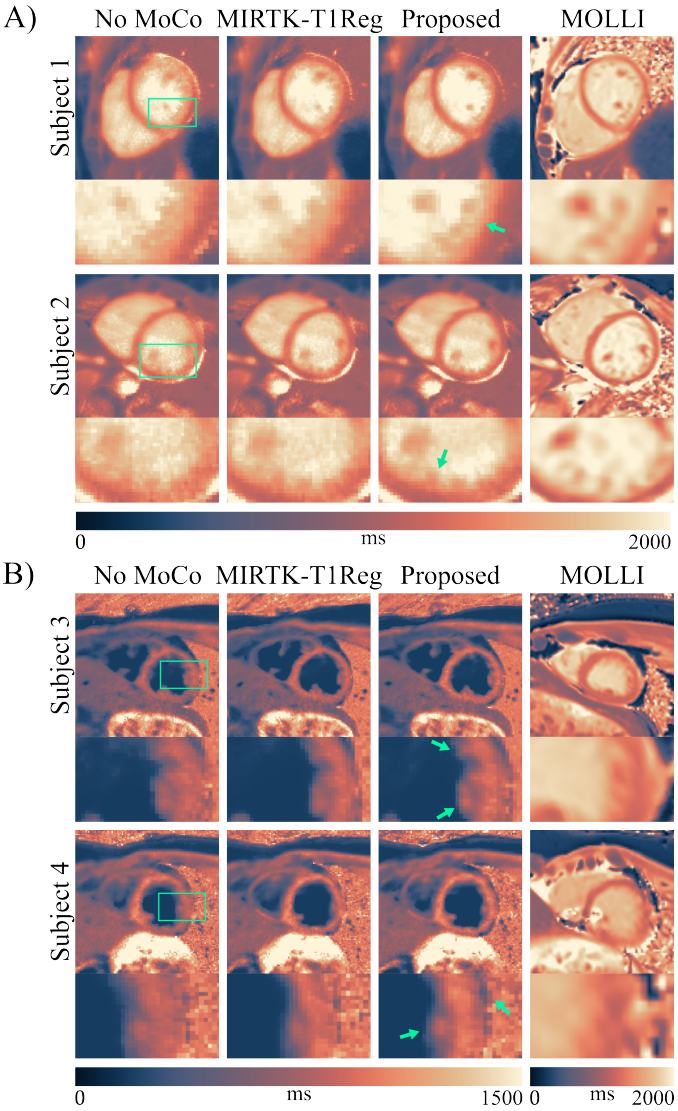

Fig. 6: T1 maps in short-axis view for four volunteers are shown: without motion correction, with $\mathbf{T}_{1,\text{MoCo}}$ using motion fields from MIRTK-T1Reg, and with the proposed method. MOLLI maps are included for reference, requiring approx. 7.4× longer scan time. Green arrows hihglight improved detail in zoomed-in areas. A) shows single-slice 2D SA results; B) shows multi-slice 2D SA. Since data in B) use a slice-selective inversion pulse, blood T1 values are underestimated. Thus the colorbar in B) is adjusted to better visualize tissue contrast.

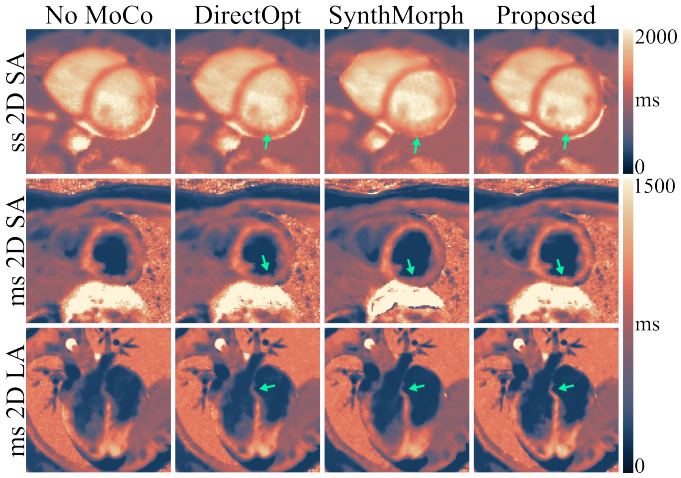

Fig. 7: Results from the ablation study for volunteers from the single-slice 2D SA dataset and the multi-slice 2D SA and LA dataset. From left to right: T1 maps estimated without motion correction (No MoCo), and with motion fields from DirectOpt, SynthMorph, and the proposed method. Arrows indicate the most affected details.

respectively).

Fig. 8A) illustrates the relative temporal consistency of the registered dynamics compared to the moving dynamics for both the single-slice and multi-slice 2D SA datasets. Using the proposed method, we observed improvements of 8.39% and 15.18% for the single- and multi-slice 2D SA datasets, respectively, with statistical significance ($p < 0.01$ and $p < 0.001$). Additionally, compared to the registration performed with MIRTK-T1Reg, the proposed approach yielded improvements of 8.40% and 14.85% ($p < 0.002$ and $p < 0.001$). Fig. 8B) presents the relative image sharpness of $\mathbf{T}_{1,\text{MoCo}}$, estimated using motion fields from both methods, compared to the T1 maps without motion correction. The metric shows comparable results between the two methods in the single-slice 2D SA dataset. For the multi-slice 2D SA dataset, the proposed method achieved a significant improvement in sharpness of 45.13% ($p < 0.001$). Additionally, compared to the motion-corrected $\mathbf{T}_{1,\text{MoCo}}$ obtained with MIRTK-T1Reg, the proposed approach yielded a further improvement of 19.64% in the multi-slice 2D SA dataset ($p < 0.001$).

Regarding computation time per dynamic, the proposed approach required 30 ± 1.8 s for CG reconstruction on the CPU and 81.14 ± 3.62 s for motion estimation on the GPU (NVIDIA GeForce RTX 4090). In contrast, MIRTK-T1Reg took 160 ± 12.14 s for TV reconstruction and 118 ± 8.51 s for motion estimation, both performed on the CPU.

To assess reproducibility, we applied our method to subjects scanned twice and compared the final T1 maps, with and without motion correction, across the two scans. Results are shown in Suppl. Fig. 2.

We analyzed deformation fields for three subjects from the 2D short-axis dataset by plotting mean motion amplitudes (11) and the percentage of negative Jacobian determinants over time within a manually selected cardiac ROI (Suppl. Fig. 1).

VI. Discussion

In this paper, we proposed a deep learning-based image registration technique for cardiac qMRI data acquired continuously over multiple cardiac cycles. Without the need for a training dataset, this method enables accurate motion estimation across various cardiac views. The motion is estimated from CG-reconstructed dynamics, which show undersampling artifacts, thus not requiring high-quality reconstructions.

One key feature of our approach is its zero-shot design, in which the registration process is perfomed for each input individually in an iterative unsupervised way. This is particularly advantageous in medical imaging, where acquiring large

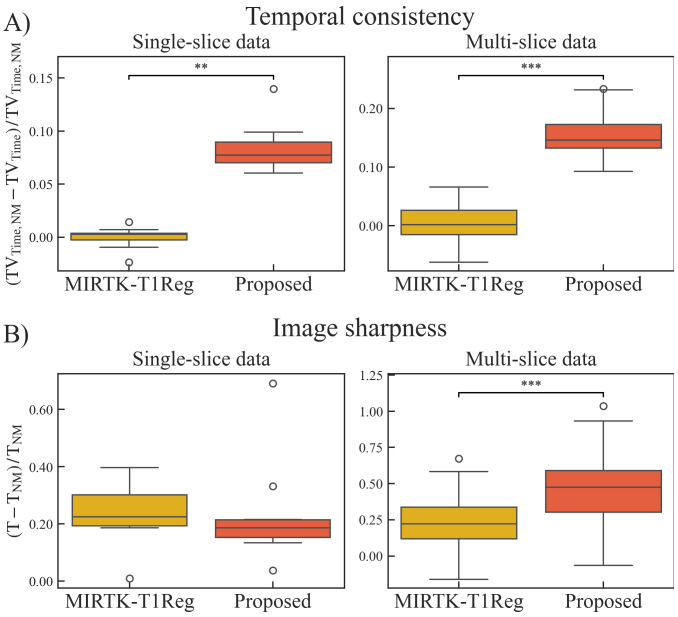

Fig. 8: Box plots comparing relative metrics for single-slice 2D SA (10 maps) and multi-slice 2D SA (mid-ventricular slices subset, 120 maps) datasets, evaluating registration with MIRTK-T1Reg and our proposed method. A) assesses temporal consistency and B) image sharpness. Statistical significance from Wilcoxon tests for single-slice data and t-tests for multi-slice data is marked as ** ($p < 0.01$) and *** ($p < 0.001$).

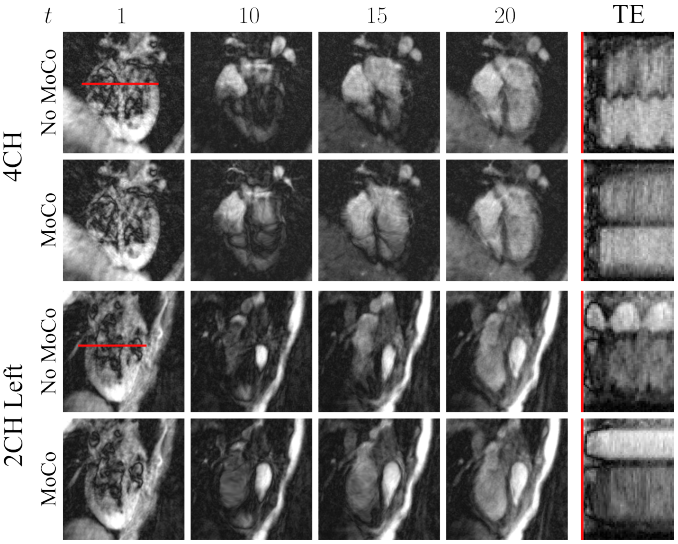

Fig. 9: The proposed registration is shown for one volunteer from the multi-slice 2D LA dataset (4CH and 2CH views). Images of the dynamics are shown for different time steps (1, 10, 15, and 20). The first row is without motion correction (No MoCo), while the second row with the proposed MoCo. The red line indicates the selected voxel whose temporal evolution (TE) is shown in the last column.

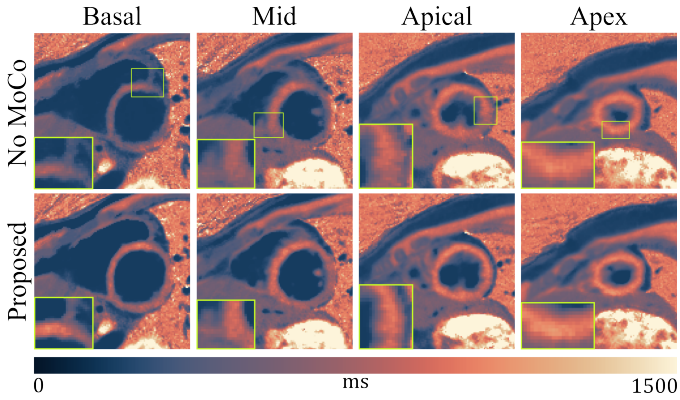

Fig. 10: T1 maps are shown for one volunteer from the multi-slice 2D SA dataset over four different views: basal, mid-ventricular, apical, and apex. T1 maps without motion correction (No MoCo) and the proposed $\mathbf{T}_{1,\text{MoCo}}$ are displayed. Improved visualization of different features can be seen, such as papillary muscles (mid-ventricular and apical sections) and RV walls (basal and mid-ventricular).

datasets is challenging. By iteratively adapting the network weights to each dynamic $\mathbf{x}^{\text{m}}$, our approach ensures optimal registration tailored to the specific input data.

Furthermore, information on the underlying physics is used to compute a motion-compensated reference based on a preliminary parametric fit of the dynamic to the signal model (shown in Fig. 4). It is then incorporated into the registration process as part of the matching term in the loss function. This makes the approach easily applicable to different qMRI approaches. The preliminary T1 map uses only a fraction of the available data (data only acquired in diastole). The T1 map shows undersampling artifacts and overestimation of T1 values for long T1 times (blood). The proposed approach utilizes much more data and leads to higher quality.

In contrast to other state-of-the-art DL-based image registration approaches like VoxelMorph, our model processes the full dynamic as input, rather than just image pairs, with the time frames as input channels. This allows for the model $\mathbf{u}_\theta$ to rely also on the temporal correlations between frames, by applying the same convolutional layers over the entire dynamic. Moreover, unlike VoxelMorph or SynthMorph, which require both moving and fixed images as inputs, our network parametrizes the motion fields only from the moving dynamic, since the motion-compensated reference is used only in the loss computation.

Regarding the results on simulated data, Fig. 2 shows that our computed motion fields align with the reference motion fields ϕ^{XCAT} in both direction and amplitude. Our approach also reduces the difference to the reference T1 values compared to the case without motion correction, improving the accuracy of T1 by 61.64% (Fig. 3), while achieving comparable results with better generalizability compared to classic registration algorithms. Note that, the pixel-wise difference between the XCAT reference and motion corrected T1 maps is limited by the fact that the different motion-correction approaches might transform the final image to a

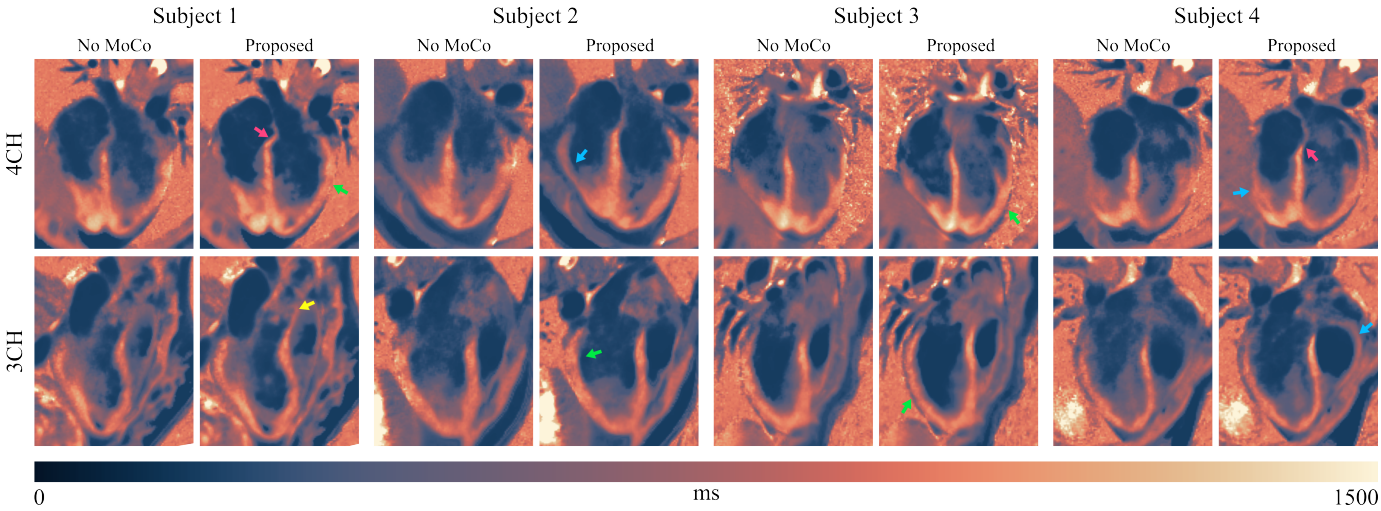

Fig. 11: T1 maps are shown for four different volunteers from the multi-slice 2D LA dataset. For each subject, 4CH and 3CH views are displayed. The maps were estimated without motion correction (No MoCo) and with the proposed MoCo method. Improved visualization of the ventricular septum (pink arrows), LV and RV walls (green and blue arrows, respectively), and the aorta (yellow arrow) can be seen.

slightly different reference motion state.

It is worth mentioning that the motion fields provided by the XCAT framework cannot be considered as ground truth as they may not be physiologically plausible. However, they still offer a useful means of quantifying the quality of motion estimation. Since the numerical XCAT model consists of distinct discrete shapes, motion transformations may exhibit discontinuities along their boundaries, leading to inconsistencies in the regularity of the fields, particularly along tissue boundaries, as shown in their color wheel representation (Fig. 2). Our approach assumes spatial regularity in the motion fields, which is enforced by the regularity term (9) in the loss function. As a result, a direct quantitative comparison between the two is not possible.

Furthermore, the estimated motion fields show that peaks in motion amplitude correlate with increased negative Jacobian percentages, suggesting that larger motion, due to systolic contraction and through-plane displacement not captured in 2D, results in more non-positive Jacobians. This supports excluding 15% of frames before final fitting to mitigate through-plane effects (Suppl. Fig. 1).

The CNN parametrization plays a key role in motion correction. The ablation study (Fig. 7) shows that removing it (DirectOpt) fails to correct artifacts and produces T1 maps similar to uncorrected ones (green arrows). On the other hand, removing the dependency on the signal model (SynthMorph) also performs worse than the proposed model, proving unsuitable for data with varying contrast, low quality, and strong motion by failing to align key structures or producing inaccurate T1 values (green arrows).

Regarding the registration quality on in-vivo data, the temporal profiles in Fig. 5 visually display that our method achieves better motion alignment than MIRTK-T1Reg in short-axis views. Specifically, the proposed registration yields improved border alignment in areas such as the papillary muscles and myocardium (white circles), resulting in a smoother correction of motion and removing residual motion artifacts still present in the registration performed with MIRTK-T1Reg. The evaluation of temporal consistency in 8A) confirms our visual observations, as our method improves motion correction of the registered dynamics by 8.39% in the single-slice 2D SA dataset when compared to the moving dynamics. An improvement of 15.18% is also achieved in the multi-slice 2D SA dataset, further demonstrating that our registration leads to more accurate alignment over time. This improvement is achieved because our registration process estimates motion from the entire dynamic $\mathbf{x}^{\mathrm{m}}$, rather than registering each image individually. Moreover, MIRTK-T1Reg assumes that motion fields are shared across the same phase in different cardiac cycles, which limits its ability to capture the full motion variability over time.

In addition, the motion estimated with MIRTK-T1Reg relies on higher-quality TV-reconstructed images, whose reconstruction requires the choice of a regularization parameter, whereas our method achieves comparable results directly from faster CG-reconstructed dynamics, further highlighting the robustness of our approach.

As for the motion-corrected T1 map estimation, Fig. 6 illustrates that our method provides comparable overall image quality to the motion-corrected maps estimated using MIRTK-T1Reg in the single-slice 2D SA dataset (first column), while also offering improved depiction of finer structural details, such as papillary muscles and fat tissue (green arrows). A more noticeable improvement is observed in the multi-slice 2D SA dataset (second column), which aligns with the quantitative results shown in Fig. 8B), confirming comparable sharpness in the single-slice 2D SA dataset. This was expected as MIRTK-T1Reg was specifically optimized for this type of acquisition. In contrast, for the multi-slice 2D SA dataset, our method outperformed the registration with MIRTK-T1Reg and improved sharpness by 45.13% compared to the T1 maps without motion correction, indicating that our proposed approach can

generalize better to different types of acquisition.

From the results across different cardiac orientations, we can see that our method performs well on short-axis as well as on long-axis views (Fig. 9, 4CH and 2CH). This versatility highlights the flexibility of the method, which is also reflected in the motion-corrected T1 maps. For instance, in the stack of short-axis views (Fig. 10), our method ensures an improved depiction of the left ventricular (LV) myocardium across all slices, as highlighted in the close-ups. Additionally, better-defined details, such as the papillary muscles and right ventricular (RV) walls, are visible. Finally, Fig. 11 shows the results of T1 maps in long-axis views, where the proposed approach improves the visualization of the ventricular walls, papillary muscles, and also other cardiac features, such as the aortic outflow tract.

The reproducibility experiment showed that motion correction consistently improved sharpness (evaluated in ROIs around the ventricular septum) across repeated scans, supporting the robustness of our method. Sharpness increases were 13.92% and 12.93% for Subject 1, 24.30% and 18.38% for Subject 2, and 18.25% and 20.66% for Subject 3.

In terms of computation time, although motion estimation times are not directly comparable due to the different hardware (CPU vs. GPU), the key advantage of the proposed method lies in achieving motion estimation with a nevertheless faster image reconstruction. Our method can recover structural details with only a CG reconstruction ($\approx$ 30 s), whereas MIRTK-T1Reg requires higher-quality images from a TV reconstruction ($\approx$ 2.5 min), which also involves selecting a suitable regularization parameter.

One limitation of this study is that no ground truth T1 maps were available. MOLLI uses a different sequence and different signal model, thus yielding different T1 values. Furthermore, MOLLI T1 maps were obtained using the standard clinical parameters, leading to maps with different resolutions and each acquisition is obtained in a different breath-hold, yielding changes in the slice position. Therefore, a direct comparison between these two acquisitions is not possible.

Another limitation of our study is that we were only able to compare our method to the registration performed with MIRTK-T1Reg. To the best of our knowledge, no other existing method allows for non-rigid cardiac motion estimation across the entire cardiac cycle in the presence of varying contrast. Both classical (e.g., [15], [53], [54]) and DL-based (e.g., [18], [20]) non-rigid MoCo approaches for cardiac T1 mapping have been developed for ECG-triggered sequences and therefore assume only small misalignments between images or breathing motion. Furthermore, the DL method proposed by [55], which operates in k-space, assumes consistent contrast between images, thus preventing its application to our problem of contrast-changing dynamic images. Because of these factors, a broader comparison with other registration methods was not feasible.

Finally, the network architecture could be further explored, as recent structures (e.g., [56], [57]) have shown to perform well in registration tasks. Similarly, incorporating an iterative process between motion and parametric estimation (e.g., [15]) may improve results. In this work, we prioritized a faster approach in both network and algorithm design, but further investigation could yield improved results. Also, T1 values were estimated using a standard method, but advanced techniques could further improve accuracy [58].

While further patient studies are necessary, as our method is subject-specific, it can adapt to different anatomies and individual cases. As no assumptions about the cardiac cycle are made, future work could involve testing this approach on more complex motion types, to access its adaptability to patient-specific heart rhythms, such as variations in heart rates or motion irregularities (e.g., ectopic beats, septal flashes).

VII. Conclusion

Our method provides an effective solution for cardiac motion estimation in continuously acquired qMRI data, enabling registration of image dynamics over full cardiac cycles despite contrast changes. It employs a zero-shot dataset-free framewok, computing optimal motion fields on a per-subject basis by relying solely on the moving dynamic, and integrating the qMRI signal model directly into the motion estimation process. The approach was evaluated on both short- and long-axis datasets. On simulated data, T1 accuracy improved by 61.64% when compared to reference values. On acquired data, our method showed improved sharpness in the motion-corrected T1 maps (by 45.13%) and enhanced temporal image alignment in the registered dynamics (by 11.78%, on average). Additionally, it outperformed a recent method designed for motion estimation under similar acquisition assumptions. To our knowledge, this is the first registration method to perform non-rigid registration across full cardiac cycles in continuous qMRI acquisition. While further studies on patient data are needed, our approach does not make any assumptions about heartbeat regularity or anatomy, which makes the method well-suited for broader clinical applications.

Acknowledgment

The authors gratefully acknowledge the co-funding from the German Research Foundation (GRK2260–BIOQIC, CRC1340 Matrix in Vision).

References

[1] A. Seraphim et al. Quantitative cardiac MRI. *Journal of Magnetic Resonance Imaging*, 51(3):693–711, 2020.

[2] R. Guo et al. Emerging Techniques in Cardiac Magnetic Resonance Imaging. *Journal of magnetic resonance imaging : JMRI*, 55(4):1043–1059, April 2022.

[3] N. Al-Wakeel-Marquard et al. Diffuse myocardial fibrosis by T1 mapping is associated with heart failure in pediatric primary dilated cardiomyopathy. *International Journal of Cardiology*, 333:219–225, June 2021.

[4] D. Radenkovic et al. T(1) mapping in cardiac MRI. *Heart failure reviews*, 22(4):415–430, July 2017.

[5] D. R. Messroghli et al. Human myocardium: Single-breath-hold MR T1 mapping with high spatial resolution–reproducibility study. *Radiology*, 238(3):1004–1012, March 2006.

[6] J. P. Ridgway. Cardiovascular magnetic resonance physics for clinicians: Part I. *Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance*, 12(1):71, November 2010.

[7] D. R. Messroghli et al. Modified Look-Locker inversion recovery (MOLLI) for high-resolution T_1 mapping of the heart. *Magnetic Resonance in Medicine*, 52(1):141–146, July 2004.

[8] S. K. Piechnik et al. Shortened Modified Look-Locker Inversion recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat breathhold. *Journal of cardiovascular magnetic resonance : official journal of the Society for Cardiovascular Magnetic Resonance*, 12(1):69, November 2010.

[9] O. P. Simonetti et al. "Black blood" T2-weighted inversion-recovery MR imaging of the heart. *Radiology*, 199(1):49–57, April 1996.

[10] M. Westwood et al. A single breath-hold multiecho T2* cardiovascular magnetic resonance technique for diagnosis of myocardial iron overload. *Journal of magnetic resonance imaging : JMRI*, 18(1):33–39, July 2003.

[11] J. I. Hamilton et al. MR fingerprinting for rapid quantification of myocardial T(1) , T(2) , and proton spin density. *Magnetic resonance in medicine*, 77(4):1446–1458, April 2017.

[12] Q. Tao et al. Robust motion correction for myocardial T(1) and extracellular volume mapping by principle component analysis-based groupwise image registration. *Journal of magnetic resonance imaging : JMRI*, 47(5):1397–1405, May 2018.

[13] H. El-Rewaidy et al. Nonrigid active shape model–based registration framework for motion correction of cardiac T_1 mapping. *Magnetic Resonance in Medicine*, 80(2):780–791, August 2018.

[14] S. Tilborghs et al. Robust motion correction for cardiac T1 and ECV mapping using a T1 relaxation model approach. *Medical Image Analysis*, 52:212–227, February 2019.

[15] H. Xue et al. Motion correction for myocardial T1 mapping using image registration with synthetic image estimation. *Magnetic Resonance in Medicine*, 67(6):1644–1655, June 2012.

[16] K. M. Kerkering et al. Motion-corrected model-based reconstruction for 2D myocardial T1 mapping. *Magnetic Resonance in Medicine*, 90(3):1086–1100, 2023.

[17] G. Balakrishnan et al. VoxelMorph: A learning framework for deformable medical image registration. *IEEE Transactions on Medical Imaging*, 38(8):1788–1800, 2019.

[18] R. A. Gonzales et al. MOCOnet: Robust Motion Correction of Cardiovascular Magnetic Resonance T1 Mapping Using Convolutional Neural Networks. *Frontiers in cardiovascular medicine*, 8:768245, 2021.

[19] Y. Li et al. Motion correction for native myocardial T(1) mapping using self-supervised deep learning registration with contrast separation. *NMR in biomedicine*, 35(10):e4775, October 2022.

[20] D. Arava et al. Deep-learning based motion correction for myocardial T1 mapping. In *2021 IEEE International Conference on Microwaves, Antennas, Communications and Electronic Systems (COMCAS)*, pp. 55–59, 2021.

[21] Y. Fu et al. Deep learning in medical image registration: A review. *Physics in Medicine & Biology*, 65(20):20TR01, October 2020.

[22] O. Ronneberger et al. U-Net: Convolutional Networks for Biomedical Image Segmentation. In N. Navab et al., editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 234–241, Cham, 2015. Springer International Publishing.

[23] M. Hoffmann et al. SynthMorph: Learning Contrast-Invariant Registration Without Acquired Images. *IEEE Transactions on Medical Imaging*, 41(3):543–558, March 2022.

[24] M. Morales et al. Implementation and validation of a three-dimensional cardiac motion estimation network. *Radiology: Artificial Intelligence*, 1:e180080, July 2019.

[25] H. Qi et al. End-to-end deep learning nonrigid motion-corrected reconstruction for highly accelerated free-breathing coronary MRA. *Magnetic resonance in medicine*, 86(4):1983–1996, October 2021.

[26] A. Ghoul et al. Attention-aware non-rigid image registration for accelerated MR imaging. *IEEE Transactions on Medical Imaging*, 43(8):3013–3026, 2024.

[27] C. Qin et al. Joint Learning of Motion Estimation and Segmentation for Cardiac MR Image Sequences. In A. F. Frangi et al., editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, volume 11071, pp. 472–480. Springer International Publishing, Cham, 2018.

[28] Y. Zhang et al. Deep-learning-based groupwise registration for motion correction of cardiac T1 mapping. In *Medical Image Computing and Computer Assisted Intervention – MICCAI 2024: 27th International Conference, Marrakesh, Morocco, October 6–10, 2024, Proceedings, Part II*, pp. 586–596, Marrakesh, Morocco and Berlin, Heidelberg, 2024. Springer-Verlag.

[29] E. Hanania et al. MBSS-T1: Model-based subject-specific self-supervised motion correction for robust cardiac T1 mapping. *Medical Image Analysis*, 102:103495, 2025.

[30] M. Jaderberg et al. Spatial transformer networks. In C. Cortes et al., editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

[31] K. P. Pruessmann et al. Advances in sensitivity encoding with arbitrary *k* -space trajectories. *Magnetic Resonance in Medicine*, 46(4):638–651, October 2001.

[32] O. Oktay et al. Attention U-Net: Learning Where to Look for the Pancreas, 2018.

[33] F. Maes et al. Multimodality image registration by maximization of mutual information. *IEEE transactions on medical imaging*, 16(2):187–198, April 1997.

[34] P. Viola and W. Wells. Alignment by maximization of mutual information. In *Proceedings of IEEE International Conference on Computer Vision*, pp. 16–23, Cambridge, MA, USA, 1995. IEEE Comput. Soc. Press.

[35] C. Studholme et al. An overlap invariant entropy measure of 3D medical image alignment. *Pattern Recognition*, 32(1):71–86, January 1999.

[36] C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3):379–423, July 1948.

[37] L. Evans. *Partial Differential Equations*. Graduate Studies in Mathematics. American Mathematical Society, 2010.

[38] D. C. Liu and J. Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical Programming*, 45(1-3):503–528, August 1989.

[39] K. M. Becker et al. Fast myocardial T(1) mapping using cardiac motion correction. *Magnetic resonance in medicine*, 83(2):438–451, February 2020.

[40] A. D. Scott et al. Motion in cardiovascular MR imaging. *Radiology*, 250(2):331–351, February 2009.

[41] R. P. Lewis et al. A critical review of the systolic time intervals. *Circulation*, 56(2):146–158, August 1977.

[42] W. J. J. Rogers et al. Quantification of and correction for left ventricular systolic long-axis shortening by magnetic resonance tissue tagging and slice isolation. *Circulation*, 84(2):721–731, August 1991.

[43] W. P. Segars et al. 4D XCAT phantom for multimodality imaging research. *Medical Physics*, 37(9):4902–4915, September 2010.

[44] L. Wissmann et al. MRXCAT: Realistic numerical phantoms for cardiovascular magnetic resonance. *Journal of Cardiovascular Magnetic Resonance*, 16(1):63, December 2014.

[45] S. Hufnagel et al. 3D whole heart k-space-based super-resolution cardiac T1 mapping using rotated stacks. *Physics in Medicine & Biology*, 69(8):085027, April 2024.

[46] R. Deichmann and A. Haase. Quantification of T1 values by SNAPSHOT-FLASH NMR imaging. *Journal of Magnetic Resonance (1969)*, 96(3):608–612, 1992.

[47] K. T. Block et al. Undersampled radial MRI with multiple coils. Iterative image reconstruction using a total variation constraint. *Magnetic Resonance in Medicine*, 57(6):1086–1098, June 2007.

[48] D. Rueckert et al. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, August 1999.

[49] E. Krotkov. Focusing. *International Journal of Computer Vision*, 1(3):223–237, October 1988.

[50] V. Spieker et al. Deep learning for retrospective motion correction in MRI: A comprehensive review. *IEEE Transactions on Medical Imaging*, 2023.

[51] M. Zhu et al. Review: A survey on objective evaluation of image sharpness. *Applied Sciences*, 13(2652), 2023.

[52] L. I. Rudin et al. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

[53] S. Zhang et al. Cardiac magnetic resonance T1 and extracellular volume mapping with motion correction and co-registration based on fast elastic image registration. *Magma (New York, N.Y.)*, 31(1):115–129, February 2018.

[54] S. Roujol et al. Adaptive registration of varying contrast-weighted images for improved tissue characterization (ARCTIC): Application to T_1 mapping. *Magnetic Resonance in Medicine*, 73(4):1469–1482, April 2015.

[55] T. Küstner et al. LAPNet: Non-rigid registration derived in k-Space for magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 40(12):3686–3697, 2021.

[56] J. Chen et al. Transmorph: Transformer for unsupervised medical image registration. *Medical Image Analysis*, 82:102615, 2022.

[57] M. Meng et al. Correlation-aware coarse-to-fine mlps for deformable medical image registration. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9645–9654, 2024.

[58] R. Guo et al. Accelerated cardiac T1 mapping in four heartbeats with inline MyoMapNet: A deep learning-based T1 estimation approach. *Journal of Cardiovascular Magnetic Resonance*, 24(1):6, January 2022.