Deep-Learning based Motion Correction for Myocardial T₁ Mapping

Dar Arava, Mohammad Masarwy
Andrew and Erna Viterbi Faculty of Electrical and Computer Engineering
Technion – Israel Institute of Technology Haifa, Israel
{aravadar,mohammad.m}@campus.technion.ac.il

Sam Khawaled *, Moti Freiman†
* Department of Applied Mathematics
† Department of Biomedical Engineering
Technion – Israel Institute of Technology Haifa, Israel
ssamakh@campus.technion.ac.il, moti.freiman@bm.technion.ac.il

Abstract—Myocardial T₁ mapping is a cardiac MRI technique, used to assess myocardial fibrosis. In this technique, a series of T₁-weighted MRI images are acquired with different saturation or inversion times. These images are fitted to the T₁ model to estimate the model parameters and construct the desired T₁ maps. In the presence of motion, the different T₁-weighted images are not aligned. This, in turn, will cause errors and inaccuracies in the final estimation of the T₁ maps. Therefore, motion correction is a necessary process for myocardial T₁ mapping. We present a deep-learning (DL) based system for cardiac T₁-weighted MRI images motion correction. When applying our DL-based motion correction system we achieve a statistically significant improved performance by means of R² of the model fitting regression, in compared to the model fitting regression without motion correction (0.52 vs 0.29, p<0.05).

Index Terms—Cardiac MRI, Deep Learning, Registration, Motion Correction, Myocardial T1 mapping, T1 Maps

I. INTRODUCTION

Myocardial T₁ mapping is a cardiac magnetic resonance imaging (MRI) technique, which shows early clinical promise in the assessment of cardiac fibrosis and inflammation by characterizing changes in the myocardial extracellular water (edema, focal, or diffuse fibrosis), fat, iron, and amyloid protein content. This imaging technique uses a series of T₁-weighted images acquired with different saturation or inversion times. The intensity of the T₁-weighted images behaves according to the following signal decay model:

\[ I = M_0(1 - e^{-\frac{t}{T_1}}) \]  

where \( T_1 \) and \( M_0 \) are the model parameters. \( M_0 \) and the native T₁ maps are later estimated by a voxel-wise fitting of the aforementioned intensity signal decay model.

To prevent motion disruption while acquiring MRI images for T₁ mapping, patients are often requested to hold-breath during the scan. There have been attempts for reconstruction of T₁ maps from free-breathing sequences, for example by using a slice tracking respiratory navigator, however, these methods still suffer from motion artifacts (caused by respiratory drifting or limitations of the prospective slice-tracking technique to reach full registration).

Respiratory and cardiac motion cause voxel misalignment between the different T₁-weighted images. This, in turn, will cause errors, artifacts, and inaccuracies in the estimation of T₁ maps. Therefore, motion correction is essential in myocardial T₁ mapping. T₁-weighted images should be aligned by a registration technique prior to performing the T₁ mapping. Post-processing registration methods find the deformation field that map between a pair of fixed and moving images into a combined coordinate system. Then, they register the moving image by warping it with the corresponding deformation field thereby reducing motion artifacts and misalignment between the two images.

In [6], a method of simulating synthetic free-motion T₁ images according to an initial noisy T₁ map is proposed. This synthetic series of images is used for the estimation of the deformation field by comparing it with the series of original T₁ images (with motion). This approach does not consider the T₁ model variations between patients. In [7], a modified optical flow energy function is defined, including an additional term meant to prevent the formation of transient structures from through-plane motions. In this approach, estimated motion parameters were effected by signal-to-noise and contrast-to-noise ratios of the acquired images. Both of the aforementioned methods [6], [7] were performed on breath-holding acquisitions (which minimize the effect of motion) and require high computation resources. In [8], a strategy of non-rigid registration of free-breathing T₁ images based on active shape models is proposed. This method, however, requires a new optimization calculation of the motion deformation for each new previously unseen data input at inference phase.

To tackle this obstacle, we propose a deep-learning (DL) approach for the motion correction of T₁ images. Utilization of a DL-based system in such post-processing registration tasks enables a faster registration, since once the network is trained, the deformation field is estimated by feed-forwarding the pair of fixed and moving images into the network. Unlike classical methods, solving an optimization problem for an unseen input of images is not needed at inference time. Additionally, we employ an unsupervised DL system to obtain a free-form deformation field, which is not limited to a certain model and can characterize the deformations caused
Deformable registration can be formulated as an optimization problem. Let us denote the pair of fixed and moving images by $I_f$ and $I_M$, respectively. $\Phi$ is the deformable transformation, which accounts for mapping the grid of $I_M$ to the grid of $I_f$. Then, the energy functional that we aim to optimize is:

$$\arg \min_{\Phi} S(I_f, I_M \circ \Phi) + \lambda R(\Phi)$$

(2)

where $I_M \circ \Phi$ denotes the result of warping the moving image with $\Phi$. $S$ is a dissimilarity metric which quantifies the resemblance between the resulted image and the fixed input, and $R$ is a regularization term that penalizes the deformation smoothness. The scalar $\lambda$ is a tuning parameter that accounts for balancing between the two terms and it controls the resemblance between the resulted image and the fixed input, $\Phi$.

The dissimilarity metric, $S$, is based on the unsupervised DL model of [9]. Our model was trained and evaluated on myocardial $T_1$-weighted images, which were acquired with a free-breathing scan.

**II. Motion correction: Proposed method**

Deformable registration can be formulated as an optimization problem. Let us denote the pair of fixed and moving images by $I_f$ and $I_M$, respectively. $\Phi$ is the deformable transformation, which accounts for mapping the grid of $I_M$ to the grid of $I_f$. Then, the energy functional that we aim to optimize is:

$$\arg \min_{\Phi} S(I_f, I_M \circ \Phi) + \lambda R(\Phi)$$

(2)

where $I_M \circ \Phi$ denotes the result of warping the moving image with $\Phi$. $S$ is a dissimilarity metric which quantifies the resemblance between the resulted image and the fixed input, and $R$ is a regularization term that penalizes the deformation smoothness. The scalar $\lambda$ is a tuning parameter that accounts for balancing between the two terms and it controls the resemblance of the resulting deformation. Popular dissimilarity metrics are the mean squared difference/error (MSE), the cross-correlation, and the mutual information (MI). We use the negative MI, as a dissimilarity loss between the resulted registered warped image and the fixed image. Local MI is empirically calculated on blocks of size $16 \times 16 \times 4$

$$S(I_f, I_M \circ \Phi) = -\sum_{x \in I_f} \sum_{y \in I_M \circ \Phi} p(x, y) \log \left( \frac{p(x, y)}{p(x)p(y)} \right)$$

(3)

As opposed to other common loss functions, such as the MSE, the MI preserves the natural image contrast change over time, which better suits the $T_1$ decay model mentioned in (1).

Additionally, we set the regularization term to be equal to the $l_2$ norm of the deformation gradients:

$$R(\Phi) = \sqrt{\sum_{i,j} |\nabla \Phi(i, j)|^2}$$

(4)

In DL-based registration, the task of the model is to predict the deformation: $\hat{\Phi} = f_\theta(I_M, I_F)$, where $\theta$ are the parameters of the network and $I_M, I_F$ are the input images. Then, the network predicts the deformation, $\Phi$, by optimizing the following:

$$\hat{\Phi} = \arg \min_{\theta} \sum_{I_F} S(I_F, I_M \circ f_\theta(I_M, I_F)) + \lambda R(f_\theta(I_M, I_F))$$

(5)

The scheme of our registration system is given in Fig. 1. Given a pair of moving ($I_M$) and target ($I_F$) images as a 2-channel input, the model predicts the deformation field, $\Phi = f_\theta(I_M, I_F)$. The deformation field is obtained by a pretrained U-Net model. Lastly, it maps each voxel, $p$, in the moving image to $\Phi(p)$ by applying the linear spatial interpolation according to the corresponding deformation field.

**Network architecture:** Our main building block is a U-Net based model, consisting of encoder and decoder layers with skip connections. The architecture of the U-Net is based on the VoxelMorph model [9]. Both the encoder and decoder layers consist of convolutional neural network (CNN) layers with kernel size $3 \times 3 \times 3$ followed by Leaky ReLU activation functions. The U-Net includes propagation of features learned in the encoder stage to the decoder layers, with the number of hidden layers as highlighted in Fig. 2. The input of the U-Net is a pair of 3D fixed and moving images taken at consecutive inverse times. The output of the U-Net is a 3 directional deformation field used in warping the moving image. The implementation of our registration network is based on an open source model of VoxelMorph [9].

**III. Experiments and Results**

**A. Database and Preprocessing**

The data set used in this work is adopted from [8]. This data set consists of 3D free-breathing $T_1$-weighted images from...
210 subjects. Cardiac MRI images are of size $320 \times 320 \times 5$, where the third dimension consists of five short axial slices located along the left ventricular. For each slice location, 11 $T_1$-weighted images were captured at the time points: ($T_1^1$, $T_1^1+RR$, $T_1^1+2RR$, ... $T_1^1+4RR$, $T_2^1$, $T_2^1+RR$, ..., $T_2^1+4RR$, $\infty$). The initial inversion times $T_1^1$, $T_2^1$ are set to 135[ms], 350[ms] and RR is the duration of one heart-beat [3]. Further details about the $T_1$ images can be found in [8].

Images were initially cropped to the central region of interests (ROIs). Then, they were down-sampled and zero-padded, in z dimension only, to be of size $128 \times 128 \times 8$. Cropping and down-sampling in x and y dimensions were applied to decrease the calculation complexity and for adjusting the size of the U-Net inputs to be of powers of 2. In addition, images were normalized to gray levels within the range $[0,1]$. Lastly, the dataset was randomly divided into training and test sets with ratios of 80% and 20%, respectively.

B. Results

We performed motion correction for image pairs with consecutive inverse times. Fig. [3] depicts the registration results. The resulting warped image has a resemblance in both shape and coordinate system to that of the fixed image. Further, the motion correction preserved the contrast and intensities of the moving image.

To align between the series of $T_1$-weighted images of all 11 time points, we cascaded the registration systems recursively. Fig. [4] shows the cascading scheme for our model. Firstly, we register between $T_1$-weighted images of times $t = 0$ and $t = 1$. Then, registration is performed in a recursive manner by forwarding the registered image from the block $n$ as a fixed image input to the next registration unit, where the image of time $t = n + 1$ is used as the moving input.

Lastly, after having the series of $T_1$-weighted images aligned, we perform quantitative $T_1$ mapping. $T_1$ maps are constructed by a voxel-wise fitting of a signal-decay model [1]. We investigate the effect of motion correction on the resulting $T_1$ maps by comparing the estimated regression results before and after applying the alignment of the $T_1$-weighted images. We noticed improvement in fitting to the $T_1$ decay model, as we achieved an increased R$^2$ of the regression after registration. The average R$^2$ and the standard deviation (std) is calculated over the whole test set. We obtained mean R$^2$ of 0.52 with std of 0.09 after applying motion correction. However, estimation $T_1$ without motion correction yields mean R$^2$ of 0.29 with standard deviation of 0.06. Fig. [5] illustrates an example of $T_1$ mapping results before and after applying motion correction. The fitting of the intensity levels of the images to the signal-decay model are performed per voxel for each of the images in the test set. Voxel-wise scatter plots of the regression and the $T_1$ decay model before and after motion correction are presented inFig. [6].

Although we achieved an increased R$^2$ in the fitting to the $T_1$ decay model after registration, there is a place for improvement in the performance of fitting. This slight improvement (R$^2$ is smaller than 0.95 , Fig. [5]) is caused by artificial distortion in the resulting registered image that were incorporated during cascading of the registration systems for 11 times. In fact, cascading registration models causes accumulating of registration error that lead to a considerable visible artifacts to the final warped image (See Fig. [7]). To inspect this effect, the MSE between the fixed image, image at time $t = 0$, to each one of the other images in times $t = 1, ..., 10$ is calculated.
Fig. 6. Scatter plots of the voxel-wise regression to T1 decay model, before and after motion correction. Motion correction improves the fitting to the model and better regression results are achieved (higher R²).

Fig. 7. Example of distorted image after cascading the model six times. From left to right: T1-weighted image at time t = 0, which is used as the input moving image, the fixed image (at time t = 6) and the resulted warped after cascading the registration blocks six times.

At initial time point, motion correction leads to lower values of MSE, compared to MSE before registration. However, at a certain time point (t = 6) accumulating registration error and distortions lead to higher MSE values (See Fig. 8).

Fig. 8. MSE of registration over time, before and after motion correction. The MSE is calculated between a reference image taken from the T1-weighted images at time t = 0 and each one of the other images at times t = 1, ..., 10. At initial time points, we notice an improvement in MSE after performing motion correction (orange curve), compared with that obtained without motion correction (blue curve). However, as a result of the accumulated error in the cascaded scheme the MSE increases starting from t = 6.

Another attempt to extend the motion correction model to deal with larger set of images is to train the network on image pairs consisting of images at t0 (as the fixed image) and another randomly selected image from times t = 1, ..., 10. This training fashion enables estimation of deformation fields that characterize larger movements. However, due to the intensity decaying behavior and gray-level variation between images with large time ranges, this led to visible artifacts in the registered image.

IV. CONCLUSIONS

In this paper, we performed a motion-correction for cardiac T1-weighted images for quantitative T1 mapping. Our motion correction system based on the unsupervised DL registration system, VoxelMorph [9]. T1 maps are reconstructed by voxel-wise signal-decay model fitting. Our DL-based motion correction system shows an improved performance in the estimation of T1 maps, as it provides better model fitting in terms of R² of the regression.

A simple cascading approach of our trained model showed a slight improvement in the fit to the T1 model, however, it led to artifacts in the final registered image due to the registration error that accumulated during cascading of the units. To overcome this challenge, one may incorporate the T1 decaying model into the loss of the DL network, and the whole series of T1-weighted images should be utilized as an input rater that only a pair of fixed and moving images.

ACKNOWLEDGMENTS

Khawaled, S. is a fellow of the Ariane de Rothschild Women Doctoral Program. Freiman, M. is a Taub fellow (supported by the Taub Family Foundation, Technion’s program for leaders in Science and Technology). We wish to thank Mr. Johanan Erez, Ms. Ina Talmon and the Vision and Image Sciences Laboratory (VISL) staff for their ongoing assistance, support, and providing technical tools for this project.

REFERENCES

[1] Christine L. Jellis, Deborah H. Kwon. Myocardial T1 mapping: modalities and clinical applications. Cardiovasc Diagn Ther. 2014 Apr; 4(2): 126–137. doi: 10.3978/j.issn.2223-3652.2013.09.03.
[2] Dina Radenkovic, Sebastian Weingärtner, Lewis Ricketts, James C. Moon, Gabriella Captur. T1 mapping in cardiac MRI. Heart Fail Rev. 2017; 22(4): 415–430. Published online 2017 Jun 16. doi: 10.1007/s10741-017-9627-2.
[3] Weingärtner S, Roujol S, Acker, Masha TA, Nezafat R. Freebreathing multislice native myocardial T1 mapping using the slice-interleaved T1 (STONE) sequence. Magn Reson Med 2015;74:115–124.
[4] Piechnik SK, Ferreira VM, Dell’Armellina E, Cochlín LE, Greiser A, Neubauer S, Robson MD. Shortened Modified Look-Locker Inversion recovery (ShMOLLI) for clinical myocardial T1-mapping at 1.5 and 3 T within a 9 heartbeat breathhold. J Cardiovasc Magn Reson 2010; 12:69.
[5] Marks B, Mitchell DG, Simelaró JP. Breath-holding in healthy and pulmonary-compromised populations: effects of hyperventilation and oxygen inspiration. J Magn Reson Imaging 7:595–597.
[6] Xue H, Shah S, Greiser A, Guettler C, Littmann A, Jolly M-P, Arai AE, Zuehlke S, Guehring J, Kellman P. Motion correction for myocardial T1 mapping using image registration with synthetic image estimation. Magn Reson Med 2012;67:1644–1655.
[7] Roujol S, Foppa M, Weingärtner S, Manning WJ, Nezafat R. Adaptive registration of varying contrast-weighted images for improved tissue characterization (ARCTIC): application to T1 mapping. Magn Reson Med 2015;73:1469–1482.
[8] Hossam El-Rewaiedy, Maryam Nezafat, Jihye Jang, Shiro Nakamori, Ahmed S. Fahmy, and Reza Nezafat. Nonrigid Active Shape Model-Based Registration Framework for Motion Correction of Cardiac T1 Mapping. Magnetic Resonance in Medicine 2018:80:780–791
[9] Guha Balakrishnan, Amy Zhao, Mert R. Sabuncu, John Guttag, Adrian V. Dalca. An Unsupervised Learning Model for Deformable Medical Image Registration. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018:9252-9260

[10] Guo, Courtney K. Multi-modal image registration with unsupervised deep learning. MEng. Thesis

[11] Unsupervised Learning of Probabilistic Diffeomorphic Registration for Images and Surfaces Adrian V. Dalca, Guha Balakrishnan, John Guttag, Mert R. Sabuncu MedIA: Medial Image Analysis. 2019. eprint arXiv:1903.03545