

Groupwise Deformable Registration of Diffusion Tensor Cardiovascular Magnetic Resonance: Disentangling Diffusion Contrast, Respiratory and Cardiac Motions

Fanwen Wang^{1,2}, Yihao Luo¹, Ke Wen^{2,3}, Jiahao Huang^{1,2}, Pedro F. Ferreira^{2,3}, Yaqing Luo^{2,3}, Yinzhe Wu^{1,2}, Camila Munoz^{2,3}, Dudley J. Pennell^{2,3}, Andrew D. Scott^{2,3}, Sonia NIELLES-Vallespin^{2,3} and Guang Yang^{1,2,3,4}

¹ Bioengineering Department and Imperial-X, Imperial College London, UK

² Cardiovascular Magnetic Resonance Unit, Royal Brompton Hospital, Guy's and St Thomas' NHS Foundation Trust, UK

³ National Heart and Lung Institute, Imperial College London, UK

⁴ School of Biomedical Engineering and Imaging Sciences, King's College London, UK

Abstract. Diffusion tensor based cardiovascular magnetic resonance (DT-CMR) offers a non-invasive method to visualize the myocardial microstructure. With the assumption that the heart is stationary, frames are acquired with multiple repetitions for different diffusion encoding directions. However, motion from poor breath-holding and imprecise cardiac triggering complicates DT-CMR analysis, further challenged by its inherently low SNR, varied contrasts, and diffusion-induced textures. Our solution is a novel framework employing groupwise registration with an implicit template to isolate respiratory and cardiac motions, while a tensor-embedded branch preserves diffusion contrast textures. We've devised a loss refinement tailored for non-linear least squares fitting and low SNR conditions. Additionally, we introduce new physics-based and clinical metrics for performance evaluation. Access code and supplementary materials at: <https://github.com/ayanglab/DTCMR-Reg>

Keywords: Registration · Diffusion · Deep Learning · Motion Correction.

1 Introduction

Diffusion tensor based cardiovascular magnetic resonance (DT-CMR) uniquely visualizes in vivo myocardial microstructure by assuming a stationary heart and employing multiple repetitions across diffusion encoding directions to track water molecule motion [11]. Despite using breath-holding and cardiac triggering to mitigate physiological motion, prospective motion correction still necessitates image registration.

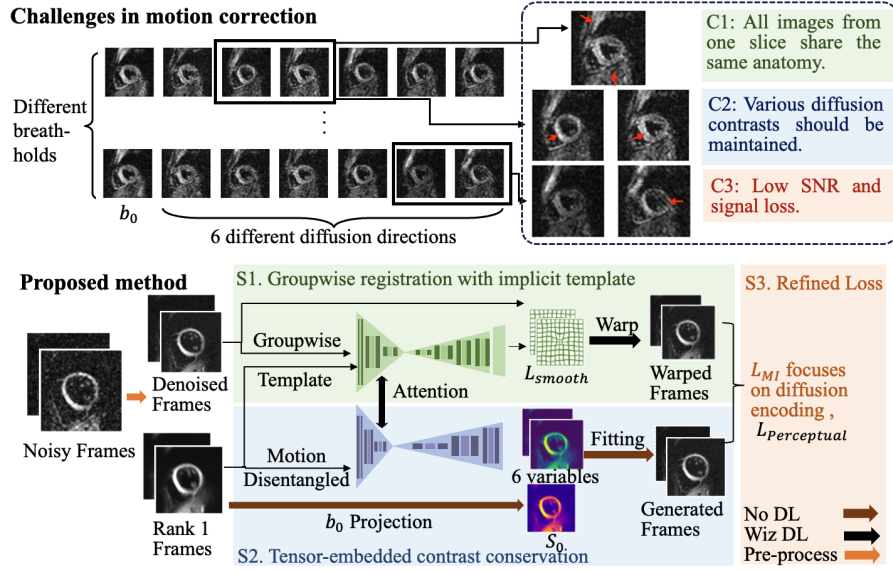


Fig. 1. The challenges in motion correction in DT-CMR and the proposed method with corresponding solutions. Given multiple scans of a single slice with varying contrasts (C1), our groupwise registration uses an implicit template to isolate motions (S1). To address diverse diffusion contrasts (C2), we integrated tensor data to maintain contrast and created pseudo frames for enhanced registration (S2). We further constrained our model with S_0 and b_0 information (S3). A unique differentiable mutual information loss is crafted for noise-reduced warped frames against generated ones (C3), fine-tuning our approach.

Registration is a technique that aligns disparate images into a unified coordinate system, based on the intensity similarity or the theoretical likelihood. This criteria fails on DT-CMR since the textural information embedded should be retained for tensor fitting. Challenges amplify due to noise, varied diffusion contrasts, and surrounding tissues (Fig. 1). Rigid registration is the most commonly applied method to correct for in-plane shifts but fail on the physiological motion. Nevertheless, cardiac and respiratory motions in the low SNR diffusion frames still call for deformable registration [4].

In our study, we introduce an innovative registration technique for correcting non-rigid motion in DT-CMR acquisitions. To manage the challenge of acquiring a single slice through multiple repetitions with varied diffusion contrasts, we employ a group-wise registration strategy anchored by an implicit template [2], which effectively separates respiratory motion and local deformations. To preserve diffusion contrast and encoding details, we incorporate tensor information matching the anatomy and generate pseudo-diffusion frames within a generation branch to facilitate registration. This approach, distinct from those in [8] and [17], introduces additional tensor constraints to enhance frame fitting under

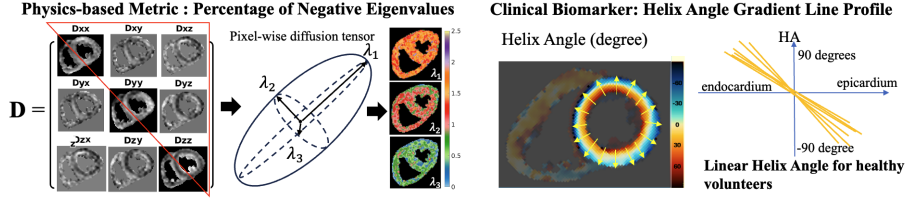


Fig. 2. Our evaluation introduces novel metrics grounded in physical properties and clinical markers. Each pixel’s three eigenvalues, indicating water molecule motion, are expected to be inherently positive. The helix angle represents the cardiomyocytes’ helical structure, which in healthy individuals, should display a linear gradient. To quantify this helical arrangement’s consistency, we employ coefficient of determination (R^2) and root mean square error (RMSE) analyses of the helix angle gradient.

low SNR conditions, leveraging differentiable mutual information as the primary loss function.

In Fig. 2, we introduce two evaluations for DT-CMR: one quantifying physical integrity, and the other assessing clinical relevance. The first, the percentage of negative eigenvalues, measures noise-affected pixels, with valid diffusion tensors requiring three positive eigenvalues to reflect tissue water motion accurately. The second metric evaluates the consistency of cardiomyocyte orientation via the helix angle (HA). A linear gradient of this helical arrangement is indicative of a healthy heart, and non linearities may arise from residual motion and noise. To the best of our knowledge, this is the first physics-based metric and clinical biomarker designed to evaluate the performance of the registration.

Our proposed method shows superior performance when applied to DT-CMR assessed using two traditional and two deep-learning based methods. Novel evaluations on the tensor data also demonstrate the potential application of these methods.

2 Methods

2.1 Problem Formulation

The fitting of diffusion images follows the pixel-wise Stejskal-Tanner equation [14]:

$$S_k = S_0 e^{-b_i \mathbf{g}_k^T \mathbf{D} \mathbf{g}_k}, \quad (1)$$

where S_k is the diffusion images we acquired, S_0 is the image intensity with no diffusion weighting and b_i is the diffusion encoding strength. $\mathbf{D}$ is a symmetric 3×3 covariance matrix with 6 independent variables (Fig. 1). $\mathbf{g}_k$ is the diffusion encoding direction.

We hypothesize that generating variables in the diffusion covariance matrix concurrently with motion correction will enhance the accuracy of the final tensor fitting. Therefore, we formulate the problem as follows:

Table 1. Ablation studies over the whole test set. R^2 and RMSE are shown in (mean $\pm$ standard deviation). NE% is shown in (median with 25% and 75% quantiles). SX stands for the modules in solution X.

Model	R^2 of HAG($\uparrow$)	RMSE of HAG($\downarrow$)	NE%($\downarrow$)
(S2.a) S0 constrain	0.894 ± 0.056	6.876 ± 3.605	1.283 (0.477, 3.259)
(S2.b) No attention	0.894 ± 0.057	6.978 ± 3.673	1.492 (0.695, 4.154)
(S3.a) No denoising	0.898 ± 0.054	6.556 ± 3.407	1.301 (0.690, 3.731)
(S3.b) MSE loss	0.890 ± 0.058	7.217 ± 3.699	1.761 (0.750, 4.585)
(S4.a) Pure seg	0.891 ± 0.057	7.213 ± 3.704	1.661 (0.766, 4.016)
(S4.b) +Seg	0.896 ± 0.055	6.758 ± 3.567	1.301 (0.818, 4.460)
Proposed	0.903 ± 0.053	6.542 ± 3.423	1.244 (0.519, 3.424)

$$\hat{\Phi}, \mathbf{D} = \arg \min_{\Phi, \mathbf{D}} \sum_{i \in k} \mathcal{L}_{mi}(\phi_i \circ S_i - S_0^{-b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i}), \quad (2)$$

where ϕ_i is the i^{th} displacement field corresponding to the frame S_i . Specifically, S_0 is an extra constraint we put into the network. Six independent variables which stand for the upper triangular part of the diffusion covariance matrix in Fig. 2 can be derived from the network.

2.2 Pipeline

Solution 1 Reference choice: Groupwise-based registration with implicit template In our group-wise registration approach using an implicit template, crucial for DT-CMR where one slice is sampled $N_d \times N_{rep}$ times (N_d : diffusion frames, N_{rep} : repetitions), we tackle the challenge of selecting an appropriate fixed reference for deformable registration. The implicit template is constructed in Tucker form [16] as:

$$\mathbf{A} = \Phi_{xy} \cdot \Phi_{dynamic}, \quad (3)$$

with Φ_{xy}^R representing the image basis matrix and $\Phi_{dynamic}^R$ denoting the dynamic factor tensor ($N_d \times N_{rep}$), where R specifies the basis matrix rank. By setting R to 1, we generate a first-rank averaged projection tensor $\mathbf{A}^1 = \Phi_{xy}^1 \cdot \Phi_{dynamic}^1$, effectively separating respiratory and local deformations [10]. This template guides a b-spline registration network with diffeomorphic transformations to align frames and compute displacement fields [12].

Solution 2 Diffusion conservation: Tensor-embedded contrast conservation To conserve the textural information from diffusion encoding directions, we propose a tensor-embedded generation branch creating pseudo diffusion images with same anatomy but different contrast to guide the groupwise registration. Following Eq. 1, we use the averaged 1st-rank projection of tensor of b_0 frames as the S_0 . We avoid the blurring induced by misalignment of the original

input images and take this as a further constraint. Using $\mathbf{A}^1$ as the input, six independent variables from the diffusion covariance matrix in Fig. 2 are generated from an encoder-decoder structure. The variables were then combined with S_0 in an exponential way following Eq. 1 to generate pseudo-diffusion images. Besides, a cross-attention module is built between the generation and registration network in a forward manner for better connection.

Solution 3 Least-squares fitting: Refined loss on diffusion encoding information To mitigate the impact of low SNR in diffusion images, which can adversely affect metrics and lead to significant deformation, we employ a strategy of extracting different components from $\Phi_{dynamic}$ across varied diffusion encodings and repetitions for initial denoising. This process is enhanced with an auto-correlation function to better preserve diffusion contrast [7]. Additionally, to address the challenges posed by substantial intensity fluctuations from motion or acquisition issues, and to align with the principles of least-squares tensor fitting, we opt for a differentiable mutual information loss $\mathcal{L}_{mi}$ as our primary loss metric, diverging from the traditional mean squared error approach used in previous studies [8, 17]. The loss function consists of three parts:

$$\mathcal{L}_{total} = \lambda_1 \cdot \mathcal{L}_{mi} + \lambda_2 \cdot \mathcal{L}_{smooth} + \lambda_3 \cdot \mathcal{L}_{perceptual}, \quad (4)$$

in which $\mathcal{L}_{mi}$ is defined on the image domain:

$$\mathcal{L}_{mi} = \sum_{i \in k} \frac{H(S_0^{-b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i}) + H(\phi_i \circ S_i)}{H(S_0^{-b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i}, \phi_i \circ S_i)}, \quad (5)$$

where $H(S_0^{-b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i})$ and $H(\phi_i \circ S_i)$ represent the marginal entropies of the generated and the warped diffusion images and $H(S_0^{-b_i \mathbf{g}_i^T \mathbf{D} \mathbf{g}_i}, \phi_i \circ S_i)$ represent the joint entropy. Using a Gaussian-based Parzen window [15, 12] instead of a rectangular window to calculate the intensity bins, we make the main loss differentiable and easy to back propagate.

$\mathcal{L}_{smooth}$ is the $L2$ norm of the spatial gradient of the displacement field,

$$\mathcal{L}_{smooth} = \sum_{i \in k} \sum_{p \in \Omega} \|\nabla \phi_i(p)\|^2, \quad (6)$$

where Ω is the 2-dimensional image area in our case.

Furthermore, we incorporate a perceptual loss, $\mathcal{L}_{perceptual}$, to supervise the high-level feature similarity between pseudo-diffusion images and warped images, utilizing layers from the VGG16 model [13].

Solution 4 Contour guidance: Optional semi-supervised version A pre-trained segmentation network [4] can be utilized to provide the mask of the myocardium as an auxiliary loss in a semi-supervised manner. We also tried to

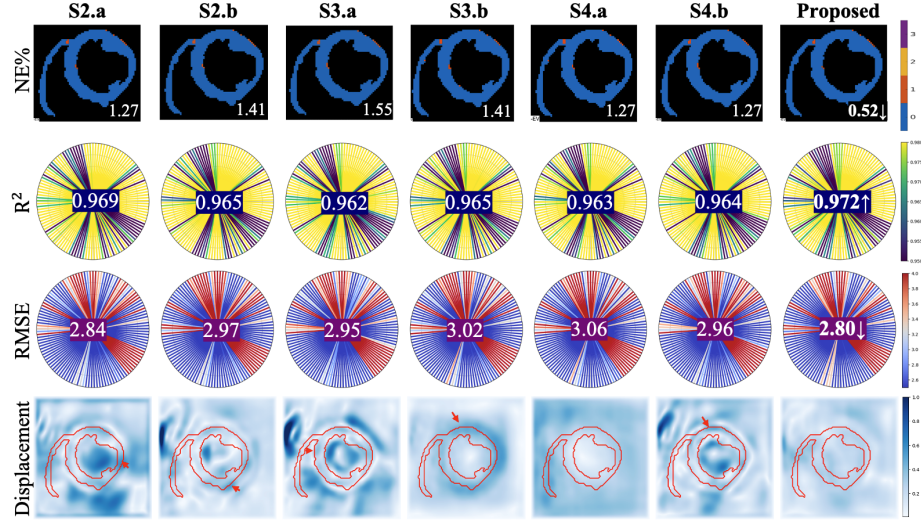


Fig. 3. The quantification of ablation studies with NE% listed on the right bottom, R^2 and RMSE of HAG and displacement field along rows. SX stands for the solution X. S2.a stands for using MSE on S_0 as constraint; S2.b is without attention module; S3.a takes noisy images as input; S3.b uses MSE as the main loss; S4.a just uses DICE on the segmented label to supervise; S4.b combines the segmentation DICE and the mutual information loss together. The unfavourable drastic deformation is pointed out by red arrows.

add a soft DICE between mask of the the intrinsic template of S_0 and the warped masks of different frames.

Notably, the displacement field ϕ_i inferred by the denoised input is applied on the original noisy frame in the following inference stage for a better and more robust fitting.

2.3 Datasets

Two datasets were collected from 20 healthy volunteers using Stimulated Echo Acquisition Mode (STEAM) based Echo Planar Imaging (EPI) at 3T and 1.5T scanners. **Dataset 1** comprises images at end-systole (ES) and end-diastole (ED) with a spatial resolution of $2.8 \times 2.8 \times 8.0 \text{ mm}^3$, including 10 repetitions with b_0 and six high b-value encodings at $b = 0$ (b_0), 150 (b_{150}) and 600 (b_{600}) s/mm^2 , resulting in $20 \text{ (subjects)} \times 2 \text{ (magnetic field)} \times 2 \text{ (ES/ED)} = 80$ cases. **Dataset 2** includes two short-axis slices (apical and basal) at ES, utilizing STEAM-EPI at 3T with twelve repetitions for b_0 , b_{150} and b_{600} . A number of $20 \text{ (subjects)} \times 2 \text{ (slices)} = 40$ cases are acquired after employing diaphragm positioning to reduce drastic through-plane motion artifacts. Hence, we got 120 cases in total. ECG-triggered breath-holding was applied during acquisitions. Data segmentation was performed using a pre-trained network [4], categorizing cases by segmentation

quality. Eighty cases with superior segmentation were allocated into training (80%) and validation (20%) subsets, while the remaining 40, characterized by lower segmentation quality, were designated for testing. This setup aimed to challenge and evaluate model performance primarily on data with lower image quality.

2.4 Experimental Details

Traditional rigid [6] and deformable registration method [9] were used for comparison. Two state-of-the-art (SOTA) deep learning based registration methods, namely MIDIR [12] and Transmorph [1] were also compared.

Ablation studies were performed to examine the various modules in our proposed method: (S2.a) MSE loss on S_0 instead of directly using the averaged projection of b_0 , (S2.b) no cross-attention module, (S3.a) noisy frames as input, (S3.b) MSE as loss, segmentation label incorporation with (S4.a) purely soft DICE on segmentation labels, (S4.b) combination of differentiable MI on the frames and soft DICE on the segmented labels, and the proposed. Details about the implementation and hyper parameter settings can be found in supplementary material.

2.5 Evaluation Methods

We compacted all the frames along a third dimension to examine the alignment of image stack. Besides, two extra quantification were applied as in Fig. 2:

Helix Angle Gradient (HAG) Line Profile In healthy volunteers, the helix angle (HA) reveals the alignment of cardiomyocytes, which ideally should be linear from the endocardium to the epicardium, as shown in Fig. 2. The mean HAG was calculated as degree of the alignment over the percentage of myocardial wall thickness. Radial line profiles with linear regression showing negative slopes and a fitting R^2 greater than 0.3 were included for analysis [5]. The R^2 values and RMSE of these fittings were evaluated.

Percentage of Negative Eigenvalues (NE%) As illustrated in Fig. 2, a rank-2 diffusion tensor and its eigenvalues were calculated for each pixel of the diffusion dataset [3]. These three eigenvalues indicate the magnitude of water molecule movement, which should ideally all be positive. However, noise, residual motion or misregistration artifacts may increase the NE%. We calculated the NE% in the left ventricular myocardium.

3 Results

3.1 Ablation and Comparison Study

In our study, we quantitatively compared various modules' performance on a representative subject, as shown in Fig.3. The proposed method, relying exclusively on the loss of generated frames, surpassed others, demonstrated alongside

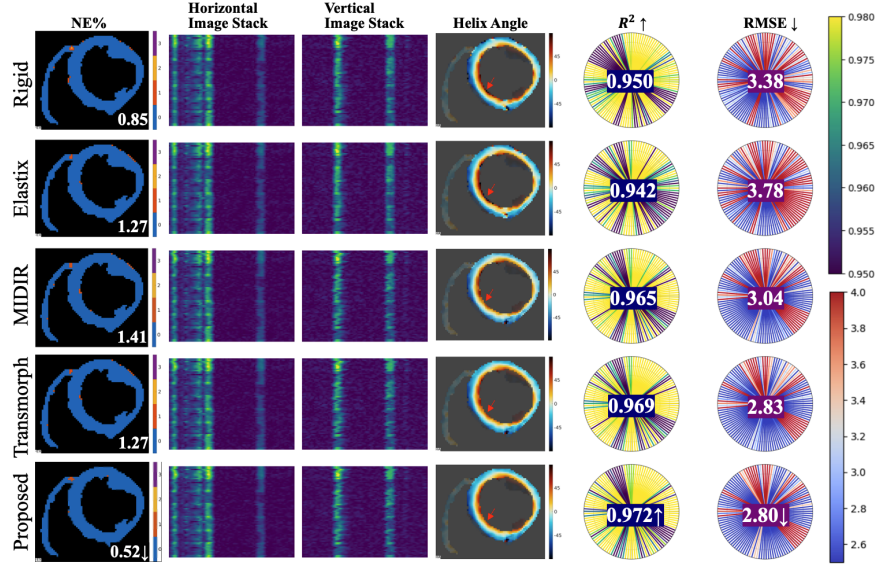


Fig. 4. The result of tensors of different comparison studies on one representative subject. NE%, image stack, R^2 and RMSE of HAG are shown. Differences along the edge of HA are highlighted by red arrows.

Table 2. Comparison studies over the whole test dataset. R^2 and RMSE are shown in (mean $\pm$ standard deviation). NE% is shown in (median with 25% and 75% quantiles)

Model	R^2 of HAG($\uparrow$)	RMSE of HAG($\downarrow$)	NE%($\downarrow$)
Rigid	0.873 $\pm$ 0.061	7.583 $\pm$ 3.739	1.251 (0.546, 4.078)
Deformable	0.881 $\pm$ 0.053	7.502 $\pm$ 3.544	1.459 (0.439, 4.136)
MIDIR	0.889 $\pm$ 0.059	7.266 $\pm$ 3.740	1.727(0.790, 4.835)
Transmorph	0.892 $\pm$ 0.057	7.144 $\pm$ 3.649	1.703 (0.508, 4.259)
Proposed	0.903 $\pm$ 0.053	6.542 $\pm$ 3.423	1.244 (0.519, 3.424)

an example displacement field. Issues such as drastic myocardium deformation, highlighted by red arrows, were observed with strategies like applying MSE to S_0 , using noisy frames as inputs, omitting the attention module, employing MSE loss, and integrating segmentation labels. These findings are further substantiated in Table 1.

Additionally, our approach was benchmarked against conventional and deep-learning based techniques, detailed in Fig. 4 and Table 2. It consistently outperformed competing methods, showcasing the lowest percentage of negative eigenvalues, highest R^2 , and minimal RMSE for helix angle gradient linear fitting, indicating superior performance across metrics for a representative subject.

4 Conclusion and Discussion

In this study, we introduce a groupwise registration method for DT-CMR to address challenges posed by multiple repetitions, diverse contrasts, and noise. An implicit template segregates respiratory and cardiac motions for group-wise frame registration. Our tensor-embedded network generates pseudo frames to differentiate diffusion contrasts, while a novel differentiable mutual information loss targets diffusion encoding to correct residual motions.

We validated our method with two new metrics, physical and clinical, demonstrating its superiority over traditional and deep learning-based registration methods. Our approach significantly reduces negative eigenvalues and improves the helix angle gradient in healthy volunteers. However, the proposed clinical biomarker of the helix angle gradient line profile is currently applicable only to healthy volunteers due to the linear features of cardiomyocytes in these individuals. To address this limitation, we plan to include additional metrics and expand our study to patients with various diseases in future research. This will enhance the applicability and robustness of our method.

Acknowledgments. This work was supported in part by ERC IMI (101005122), the H2020 (952172), the MRC (MC/PC/21013), the Royal Society (IEC\NSFC\211235), the NVIDIA Academic Hardware Grant Program, the SABER project supported by Boehringer Ingelheim Ltd, Wellcome Leap Dynamic Resilience, the UKRI Future Leaders Fellowship (MR/V023799/1) and the British Heart Foundation grant (RG/19/1/34160).

Disclosure of Interests. We have no competing interests in the paper as required by the publisher.

References

1. Chen, J., Frey, E.C., He, Y., Segars, W.P., Li, Y., Du, Y.: Transmorph: Transformer for unsupervised medical image registration. *Medical image analysis* **82**, 102615 (2022)
2. Christodoulou, A.G., Shaw, J.L., Nguyen, C., Yang, Q., Xie, Y., Wang, N., Li, D.: Magnetic resonance multitasking for motion-resolved quantitative cardiovascular imaging. *Nature biomedical engineering* **2**(4), 215–226 (2018)
3. Ferreira, P.F., Kilner, P.J., McGill, L.A., Nelles-Vallespin, S., Scott, A.D., Ho, S.Y., McCarthy, K.P., Haba, M.M., Ismail, T.F., Gatehouse, P.D., et al.: In vivo cardiovascular magnetic resonance diffusion tensor imaging shows evidence of abnormal myocardial laminar orientations and mobility in hypertrophic cardiomyopathy. *Journal of Cardiovascular Magnetic Resonance* **16**(1), 87 (2014)
4. Ferreira, P.F., Martin, R.R., Scott, A.D., Khalique, Z., Yang, G., Nelles-Vallespin, S., Pennell, D.J., Firmin, D.N.: Automating in vivo cardiac diffusion tensor post-processing with deep learning-based segmentation. *Magnetic resonance in medicine* **84**(5), 2801–2814 (2020)
5. Gorodezky, M., Scott, A.D., Ferreira, P.F., Nelles-Vallespin, S., Pennell, D.J., Firmin, D.N.: Diffusion tensor cardiovascular magnetic resonance with a spiral trajectory: An in vivo comparison of echo planar and spiral stimulated echo sequences. *Magnetic resonance in medicine* **80**(2), 648–654 (2018)
6. Guizar-Sicairos, M., Thurman, S.T., Fienup, J.R.: Efficient subpixel image registration algorithms. *Optics letters* **33**(2), 156–158 (2008)
7. Gurney-Champion, O.J., Collins, D.J., Wetscherek, A., Rata, M., Klaassen, R., Van Laarhoven, H.W., Harrington, K.J., Oelfke, U., Orton, M.R.: Principal component analysis for fast and model-free denoising of multi b-value diffusion-weighted mr images. *Physics in Medicine & Biology* **64**(10), 105015 (2019)
8. Hanania, E., Volovik, I., Barkat, L., Cohen, I., Freiman, M.: Pcmc-t1: Free-breathing myocardial t1 mapping with physically-constrained motion correction. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 226–235. Springer (2023)
9. Klein, S., Staring, M., Murphy, K., Viergever, M.A., Pluim, J.P.: Elastix: a toolbox for intensity-based medical image registration. *IEEE transactions on medical imaging* **29**(1), 196–205 (2009)
10. Nguyen, C.T., Christodoulou, A.G., Coll-Font, J., Ma, S., Xie, Y., Reese, T.G., Mekkaoui, C., Lewis, G.D., Bi, X., Sosnovik, D.E., et al.: Free-breathing diffusion tensor mri of the whole left ventricle using second-order motion compensation and multitasking respiratory motion correction. *Magnetic resonance in medicine* **85**(5), 2634–2648 (2021)
11. Nelles-Vallespin, S., Scott, A., Ferreira, P., Khalique, Z., Pennell, D., Firmin, D.: Cardiac diffusion: technique and practical applications. *Journal of Magnetic Resonance Imaging* **52**(2), 348–368 (2020)
12. Qiu, H., Qin, C., Schuh, A., Hammernik, K., Rueckert, D.: Learning diffeomorphic and modality-invariant registration using b-splines. In: *Medical Imaging with Deep Learning* (2021)
13. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014)
14. Stejskal, E.O., Tanner, J.E.: Spin diffusion measurements: spin echoes in the presence of a time-dependent field gradient. *The journal of chemical physics* **42**(1), 288–292 (1965)

15. Thévenaz, P., Unser, M.: Optimization of mutual information for multiresolution image registration. *IEEE transactions on image processing* **9**(12), 2083–2099 (2000)
16. Tucker, L.R.: Some mathematical notes on three-mode factor analysis. *Psychometrika* **31**(3), 279–311 (1966)
17. Zaffrani-Reznikov, Y., Afacan, O., Kurugol, S., Warfield, S., Freiman, M.: qdwi-morph: Motion-compensated quantitative diffusion-weighted mri analysis for fetal lung maturity assessment. In: *European Conference on Computer Vision*. pp. 482–494. Springer (2022)