Learning-based and unrolled motion-compensated reconstruction for cardiac MR CINE imaging

Jiazhen Pan1  Daniel Rueckert1,2  Thomas Küstner3⋆  Kerstin Hammernik1,2⋆

1 Klinikum Rechts der Isar, Technical University of Munich, Germany
2 Department of Computing, Imperial College London, United Kingdom
3 Medical Image And Data Analysis (MIDAS.lab), University Hospital of Tübingen, Germany

Abstract. Motion-compensated MR reconstruction (MCMR) is a powerful concept with considerable potential, consisting of two coupled sub-problems: Motion estimation, assuming a known image, and image reconstruction, assuming known motion. In this work, we propose a learning-based self-supervised framework for MCMR, to efficiently deal with non-rigid motion corruption in cardiac MR imaging. Contrary to conventional MCMR methods in which the motion is estimated prior to reconstruction and remains unchanged during the iterative optimization process, we introduce a dynamic motion estimation process and embed it into the unrolled optimization. We establish a cardiac motion estimation network that leverages temporal information via a group-wise registration approach, and carry out a joint optimization between the motion estimation and reconstruction. Experiments on 40 acquired 2D cardiac MR CINE datasets demonstrate that the proposed unrolled MCMR framework can reconstruct high quality MR images at high acceleration rates where other state-of-the-art methods fail. We also show that the joint optimization mechanism is mutually beneficial for both sub-tasks, i.e., motion estimation and image reconstruction, especially when the MR image is highly undersampled.

1 Introduction

Cardiac magnetic resonance imaging (CMR) plays an essential role in evidence-based diagnostic of cardiovascular disease [16] and serves as the gold-standard for assessment of cardiac morphology and function [18]. High-quality cardiac image reconstruction with high spatial and temporal resolution is an inevitable prerequisite for this assessment. Shorter scan times with higher spatio-temporal resolution are desirable in this scenario. However, this requires high acceleration rates which in turn is only achievable if sufficient spatio-temporal information linked by the cardiac motion is shared during reconstruction.

A large variety of CMR reconstruction methods have been dedicated to cope with cardiac motion during the reconstruction. They can be categorized into two

⋆ contributed equally
sections: Implicit and explicit motion correction. Implicit motion correction during the reconstruction sidesteps the non-rigid cardiac motion estimation, which still remains one of the most challenging problems in CMR. Most research in this section focused either on exploiting spatio-temporal redundancies in complementary domains [14,30], enforcing sparseness/low-rankness along these dimensions [13,19,26], or improving the spatio-temporal image regularization in an unrolled optimization model [17,31]. Yet in this sense, motion is only implicitly corrected without knowing or estimating the true underlying motion.

Motion can also be explicitly corrected during the reconstruction by applying motion estimation/registration models. The work of Batchelor et al. [6] has pioneered the field of motion-compensated MR reconstruction (MCMR). It proposed the idea of embedding the motion information as an explicit general matrix model into the MR reconstruction process and demonstrated some successful applications in leg and brain reconstruction. However, in CMR, the respiratory and cardiac motion is much more complex and therefore a more sophisticated motion model is required. Some endeavours were made first in the field of coronary magnetic resonance angiography (CMRA) to compensate respiratory motion [7,9] in which non-rigid motion models based on cubic B-Splines were employed [15,21]. However, these conventional registrations require substantial computation times in the order of hours making the practical implementation of image reconstruction infeasible. More recently, learning-based registration methods [4,5,28] have been proposed, which leverage trained neural networks to accelerate the estimation in inference. In [29] these learned registrations are embedded into a CMRA reconstruction for an unrolled CG-SENSE optimization [27] with a denoiser regularizer [1]. However, MCMR has been rarely studied in the context of cardiac CINE imaging [8].

MCMR can be recast as two codependent and intertwined sub-optimization problems: Image reconstruction and motion estimation. This stands in contrast to methods in which only the MR image is optimized [9,7,29,8], whereas the motion is pre-computed and assumed to be fixed during the whole optimization process. A reliable reconstruction relies on precise motion estimation, while the accuracy of the motion prediction is subject to the quality of images. Odille et al. [23,24] introduced an iterative, alternating approach to solve this problem. Concurrently, [3,32] proposed to solve this joint optimization problem by leveraging variational methods. However, to the best of our knowledge, this problem has not been formulated and explored using deep learning-based approaches. Furthermore, as algorithm unrolling has been successfully applied in modern reconstruction approaches [1,11] by learning image regularization from data, the possibility to unroll the MCMR optimization with motion estimation networks has not been explored.

In this work, we introduce a deep learning-based and unrolled MCMR framework to reconstruct CMR images in the presence of motion. The highlights of the work can be summarized as follows: First, we propose an unrolled MCMR optimization process with an embedded group-wise motion estimation network [12]. A joint optimization between the image reconstruction and motion estimation
is proposed, while the motion estimation is updated iteratively with the reconstruction progress. Second, in contrast to [7,8,29] our proposed approach does not require a pre-computed motion field, which is usually obtained from initial non-motion compensated reconstruction. Finally, instead of reconstructing only one frame by compensating motion from all frames to one target frame, our proposed method provides a motion-resolved image sequence in which all frames are motion-corrected. The proposed framework was trained and tested on in-house acquired 2D CINE data of 40 subjects and compared to state-of-the-art MCMR methods. The conducted experiments demonstrate that the proposed method outperforms the competing methods especially in high acceleration situations, concluding that embedding the dynamic motion into an unrolled MCMR can drastically improve the reconstruction performance.

2 Methods

Assume that $x^{(t)} \in \mathbb{C}^M$ is the $t$-th frame of the image sequence $x = [x^{(1)}, \ldots, x^{(N)}]$ with $M$ pixels and $y^{(t)} \in \mathbb{C}^{MQ}$ denotes its corresponding $t$-th undersampled k-space data with $Q$ coils. In the context of CMR, the MCMR of the whole MR sequence with $N$ temporal frames can be formulated as

$$
\min_{x, U} \sum_{t_1=1}^{N} \sum_{t_2=1}^{N} \| A^{(t_2)} U^{(t_1 \rightarrow t_2)} x^{(t_1)} - y^{(t_2)} \|_2^2,
$$

(1)

where $A^{(t)}$ denotes the MR multi-coil encoding operator with $A^{(t)} = D^{(t)} FC$ for coil-sensitivity maps $C$, Fourier transform $F$ and the undersampling matrix $D$. Based on the idea of [6], we build a motion matrix $U^{(t_1 \rightarrow t_2)}$ representing the warping from the temporal frame $t_1$ to $t_2$, which is obtained from the estimated motion fields $u^{(t_1 \rightarrow t_2)}$ (introduced in Section 2.2).

In traditional MCMR [7,8,29], only the images $x$ are optimized, whereas the motion is pre-computed from initial (non-motion compensated) reconstructed images just for once and assumed to be constant during the whole optimization process. However, motion estimation on the initial motion-corrupted and artifact-degraded images can be inaccurate, incurring error-propagation during the optimization. Therefore, a joint optimization as proposed in [24] for image $x$ and motion $U$ as described in Eq. 1 is desired.

2.1 Motion-compensated image reconstruction framework

In this work we unroll the MCMR in Eq. 1 with a motion estimation network and apply the joint optimization between the image reconstruction and motion estimation. The framework is illustrated in Fig. 1. First, the zero-filled undersampled image sequence $x_0$ is fed to the conjugate-gradient SENSE block [27] along with $y$, $C$, and $D$. The first SENSE block conducts a reconstruction without motion embedding. Afterwards, the reconstructed image sequence $x_1$ is passed
Fig. 1. The proposed unrolled MCMR framework. A dynamic joint optimization is performed between the image reconstruction in the CG-SENSE block and the motion correction from the motion estimator GRAFT. GRAFT is a group-wise motion estimation network leveraging the temporal redundancy to conduct a 1-to-$N$ motion estimation. This estimation is carried out $N$ times to accomplish an $N$-to-$N$ motion correction for all temporal frames.

to a motion estimation network (GRAFT, introduced in Section 2.2) as inputs and the first motion sequence $u_1$ comprising of all pairs of motion frames is predicted. This motion sequence is then applied to the input together with $y$, $C$ and $D$ in the next CG-SENSE block, while the previous reconstruction is used as a regularization. These CG-SENSE blocks solve Eq. 1 with an additional $\ell_2$ regularization by the image from the previous step and output a motion-corrected image sequence with temporarily freezing $U_{i-1}$ and $x_{i-1}$, following

$$x_i = \arg\min_x \sum_{t_1=1}^{N} \sum_{t_2=1}^{N} \left\| A(t_2) U_{i-1}^{(t_1 \rightarrow t_2)} x(t_1) - y(t_2) \right\|_2^2 + \frac{1}{2\lambda} \|x - x_{i-1}\|_2^2. \quad (2)$$

This alternating scheme is carried out for $I$ iterations. The motion estimation difficulty of GRAFT is alleviated gradually with the progress of this alternating scheme along with the image-quality improvement.

2.2 Motion Estimation Network

There exists various motion estimation approaches in the field of medical imaging. In this work, we select GRAFT [12] for motion estimation, which takes the full image sequence $x$ together with the target frame $x^{(t)}$ as inputs to predict the motion sequence $u^{(t)} = \left[ u^{(t \rightarrow 1)}, \ldots, u^{(t \rightarrow N)} \right]$ for all $N$ frames. GRAFT is suitable to be embedded into our framework due to its accuracy, speed and efficiency.
Accuracy An accurate motion estimation is the essential requirement for the unrolled MCMR framework. GRAFT contains a 4D correlation layer and iterative Gated Recurrent Unit (GRU) blocks to conduct precise motion predictions. It has shown superior results compared to conventional registration methods [12]. Furthermore, it is a group-wise estimation network considering all temporal frames and has a dedicated frameblock leveraging temporal redundancies to mitigate the impact of through-plane-motion. Because of this group-wise attribute, temporal coherence can be instilled during the training by appending temporal regularization in the loss function.

Speed Since the motion registration is invoked at each iteration step and for all pairs of motion frames, a fast motion estimation method is required. The reconstruction of \( N \) frames involves \( N^2 \) motion predictions. In this context, traditional registration-based methods such as [15,34] require hours to perform \( N^2 \) calculations, rendering them impractical for our framework. In contrast, GRAFT only requires a few seconds to estimate the motion of all pairs.

Efficiency For end-to-end learning over all iterative stages, the losses and gradients need to be accumulated. In the context that \( N^2 \) calculations (and for multiple iterations) are entailed, the training of large network architectures with more than 20 million trainable parameters [5,28] becomes infeasible due to their vast GPU memory footprint (>48GB). In contrast, GRAFT has only 5 million trainable parameters circumventing the GPU overcharge problems.

In order to use GRAFT for motion estimation and integrate it in our unrolled framework we define the following loss function to train GRAFT:

\[
L_i = \sum_{t_1=1}^{N} \sum_{t_2=1}^{N} \left\| \rho \left( \mathbf{u}_{i}^{(t_1 \rightarrow t_2)} x_{gt}^{(t_1)} - x_{gt}^{(t_2)} \right) \right\|_1 + \alpha \sum_{t_1=1}^{N} \sum_{t_2=1}^{N} \sum_{d \in \{x, y\}} \left\| \nabla_d u_{i}^{(t_1 \rightarrow t_2)} \right\|_1 + \beta \sum_{t_1=1}^{N} \left\| \nabla_t u_{i}^{(t_1)} \right\|_1. \tag{3}
\]

The first term ensures data fidelity during the training. We warp the ground-truth frame \( x_{gt}^{(t_1)} \) to the target ground-truth frame \( x_{gt}^{(t_2)} \) using the motion \( u_{i}^{(t_1 \rightarrow t_2)} \), which is predicted from GRAFT based on the previously reconstructed image \( x_{i-1} \). It should be noted that we apply the ground-truth frame in the loss function to train GRAFT, so that GRAFT is forced to learn and extract genuine correlations from the undersampled images and that any motion falsely originating from image artifacts are not rewarded. The Charbonnier function \( \rho(x) = (x^2 + 10^{-12})^{0.45} \) [33,25] is employed as the penalty function. Additional regularization terms for the spatial plane weighted by \( \alpha \) and along the temporal axis weighted by \( \beta \) are included in the loss to mitigate estimation singularities and to ensure motion coherence in spatial and temporal domain. This loss is calculated after every unrolled motion estimation step with progressive exponential weight decay \( L_{total} = \sum_{i=1}^{I} \gamma^{I-I} L_i \).
3 Experiments

Training was carried out on 35 subjects (a mix of patients and healthy subjects) of in-house acquired short-axis 2D CINE CMR, whereas testing was performed on 5 subjects. Data were acquired with 30/32/34 multiple receiver coils and 2D balanced steady-state free precession sequence on a 1.5T MR (Siemens Aera with TE=1.06 ms, TR=2.12 ms, resolution=1.9×1.9mm² with 8mm slice thickness, 8 breath-holds of 15s duration). Image sequence size varies from 176×132 (smallest) to 192×180 (largest) with 25 temporal cardiac phases. A stack of 12 slices along the long axis was collected, resulting in 304/15 image sequence (2D+t) for training/test, while the apical slices without clear cardiac contour were excluded.

The proposed framework with GRAFT as the embedded motion estimator was trained on an NVIDIA A40 GPU with AdamW [20] (batch size of 1, learning rate 1e-4 and weight decay of 1e-5). The number of unrolled iterations is set to 3 during training, but this can be flexibly adapted during inference and iterations are stopped when the peak signal to noise ratio (PSNR) converges (PSNR increment < 0.1). The trainable weights of GRAFT are shared during the iterative optimization. The hyperparameters α, β, γ and λ were set to 10, 10, 0.6 and 2 respectively. Three trainings and tests were conducted separately on retrospectively undersampled images with VISTA [2] of R = 8, 12 and 16 acceleration without any prior reconstruction or prior motion correction as conducted in [7,8,29]. During the training and testing, raw multi-coil k-space data were used for the reconstruction. We compare the proposed approach with non-motion compensated CG-SENSE (N-CG-SENSE) reconstruction and with non-iterative MCMR methods. In the latter case, the motion is pre-computed from GRAFT [12] and Elastix [15] and set to be constant during the whole optimization process.

4 Results and Discussion

Fig. 2. Quantitative and qualitative results of proposed MCMR method. A) Performance measured by peak signal to noise ratio (PSNR) for increasing unrolled iteration numbers over all test subjects and temporal frames for retrospective undersampling with R=8, 12 and 16. B) Qualitative image quality of the proposed MCMR at R=16 over the first three iterations (i=1,2,3) in comparison to the non-motion compensated CG-SENSE (N-CG-SENSE) at R=16 and the fully-sampled reference image (R=1).
Fig. 3. Qualitative reconstruction performance of CINE CMR in spatial (x-y) and spatio-temporal (y-t) plane of non-motion compensated CG-SENSE (N-CG-SENSE), proposed MCMR at third iteration, CG-SENSE with pre-estimated motion from GRAFT and Elastix, for different acceleration rates R=8, 12 and 16 in comparison to the fully-sampled (R=1) reference. The selected superior-inferior (y-axis) is marked with a blue line in the reference image. The obtained PSNR in comparison to the fully-sampled reference is shown at the bottom left. The highly accelerated cases for Elastix/GRAFT + CG-SENSE did not converge therefore results are not shown.

Fig. 2A reveals the relation between the estimation performance (indicated by averaged PSNR over all test subjects and temporal frames) and the unrolled iteration number of the proposed framework during inference for R=8, 12 and 16. The reconstruction accuracy is improved with increasing iteration numbers for all three cases. The reconstruction for R=12 and R=16 benefit most from our framework and obtain a significant performance lifting for the first three iterations. The advantage of unrolling the motion estimator in the joint MCMR optimization process is qualitatively illustrated in Fig. 2B, in which the MR image is highly undersampled with R=16. Not only is the image quality improved with the course of the iteration, but the motion estimator can also deliver more precise and detailed estimation by virtue of the higher quality image. A full qualitative analysis with R=8 and 12 is shown in Supplementary Fig. S1. Furthermore, the motion estimation based on unprocessed and artifact-affected images (from high undersampling rates) is challenging. We also trained GRAFT without the proposed unrolled optimization framework [12] to generalize for motion estimation from images with different undersampling rates. However, training started to fail for accelerations R=12 and beyond. In contrast, training difficulty is reduced if we embed GRAFT into the proposed joint optimization framework while the image quality is restored with the progress of the optimization. This makes GRAFT capable of estimating meaningful motion as showcased in Supplementary Fig. S1 for R=12 and even for R=16. Moreover, this mechanism also introduces a data augmentation process while more high-quality data is generated during training.

A comparison study of the proposed MCMR to N-CG-SENSE and CG-SENSE reconstruction with pre-computed motion based on GRAFT and Elastix is shown in Fig. 3. The training of GRAFT alone failed for R=12 and 16, whereas
Table 1. Quantitative analysis of reconstruction for accelerated CINE CMR (R=8, 12 and 16) using the proposed MCMR method, non-motion compensated CG-SENSE (N-CG-SENSE), GRAFT+CG-SENSE and Elastix+CG-SENSE. Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [35] are used to evaluate all test subjects. Their mean value, standard deviations are shown next to the respective methods execution times. The best results are marked in bold. The failed or inferior experiments are marked with 'N.A.'.

| Acc R | Methods          | SSIM     | PSNR    | Time (s) |
|-------|------------------|----------|---------|----------|
| 8     | Proposed MCMR    | 0.943 (0.018) | 36.26 (2.22) | 18.81s   |
|       | GRAFT [12] + CG-SENSE | 0.913 (0.019) | 34.93 (1.80) | 6.27s   |
|       | Elastix [15] + CG-SENSE | 0.645 (0.057) | 25.04 (2.10) | 4281s   |
|       | N-CG-SENSE       | 0.821 (0.038) | 30.80 (2.15) | 1.37s   |
| 12    | Proposed MCMR    | 0.932 (0.018) | 35.45 (2.00) | 18.81s   |
|       | GRAFT + CG-SENSE | N.A.     | N.A.    | 6.27s   |
|       | Elastix + CG-SENSE | 0.568 (0.072) | 23.51 (2.20) | 4281s   |
|       | N-CG-SENSE       | 0.637 (0.062) | 24.40 (2.39) | 1.37s   |
| 16    | Proposed MCMR    | 0.927 (0.019) | 34.78 (1.86) | 18.81s   |
|       | GRAFT + CG-SENSE | N.A.     | N.A.    | 6.27s   |
|       | Elastix + CG-SENSE | N.A.     | N.A.    | 4281s   |
|       | N-CG-SENSE       | 0.531 (0.08)  | 21.736 (2.45) | 1.37s   |

motion estimation with Elastix for R=16 was not conducted due to the poor quality results collected at R=12. The results are shown after the third iteration for the proposed MCMR, providing an optimal trade-off between time and performance. Our approach restores the undersampled MR sequence with high quality without artifacts in spatial and temporal domain. A quantitative analysis over all test subjects including reconstruction of all 25 temporal frames is shown in Table 1. Although the execution time of our approach is longer than the original GRAFT if the iteration is set > 1, an overall more precise reconstruction is achieved. Furthermore, the overall reconstruction is within acceptable clinical durations. The proposed MCMR enables to reconstruct an undersampled image with a much higher acceleration rate and offers the high flexibility to perform the reconstruction depending on different time/performance requirement.

The proposed MCMR framework also has some limitations: Currently, the proposed MCMR cannot guarantee that the estimated motion is diffeomorphic. The possibility to integrate the scaling and squaring layer [10] to ensure diffeomorphic motion estimation will be investigated in future work. Furthermore, we have not studied the interaction between our explicit motion correction framework with implicit motion correction (denoising) networks, which is also subject to future work. Moreover, we plan to evaluate the performance of our method on prospectively undersampled data in the future. Finally, the hyper-parameters applied in this work are estimated empirically and they might not be the optimal combination. Integration of the learnable hyper-parameters tuning [22] will be considered in our next step.
5 Conclusion

In this work, we proposed a deep learning-based MCMR framework and studied it in CINE CMR imaging. We explored the possibility to unroll the MCMR optimization with a motion estimation network while a dynamic joint optimization between the reconstruction and the motion estimation is carried out. Although this idea is still at a nascent stage, its potential in high quality reconstruction of highly accelerated data can be appreciated. The conducted experiments against baseline methods showcased a rapid, more robust and more precise reconstruction of our proposed framework.

6 Acknowledgements

This work was supported in part by the European Research Council (Grant Agreement no. 884622).

References

1. Aggarwal, H.K., Mani, M.P., Jacob, M.: Model based image reconstruction using deep learned priors (modl). In: IEEE International Symposium on Biomedical Imaging (ISBI). pp. 671–674 (2018)
2. Ahmad, R., Xue, H., Giri, S., et al.: Variable density incoherent spatiotemporal acquisition (vista) for highly accelerated cardiac mri. Magnetic Resonance in Medicine 74(5), 1266–1278 (2015)
3. Aviles-Rivero, A.I., Debroux, N., Williams, G., et al.: Compressed sensing plus motion (cs + m): A new perspective for improving undersampled mr image reconstruction. Medical Image Analysis 68, 101933 (2021)
4. Balakrishnan, G., Zhao, A., Sabuncu, M.R., et al.: An unsupervised learning model for deformable medical image registration. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9252–9260 (2018)
5. Balakrishnan, G., Zhao, A., Sabuncu, M.R., et al.: Voxelmorp: A learning framework for deformable medical image registration. IEEE Transactions on Medical Imaging 38(8), 1788–1800 (2019)
6. Batchelor, P., Atkinson, D., Irarrazaval, P., Hill, D., et al.: Matrix description of general motion correction applied to multishot images. Magnetic Resonance in Medicine 54, 1273–1280 (2005)
7. Bustin, A., Rashid, I., Cruz, G., et al.: 3d whole-heart isotropic sub-millimeter resolution coronary magnetic resonance angiography with non-rigid motion-compensated proet. Journal of Cardiovascular Magnetic Resonance 22(1) (2020)
8. Cruz, G., Hammernik, K., Kuestner, T., et al.: One-heartbeat cardiac cine imaging via jointly regularized non-rigid motion corrected reconstruction. In: Proc. International Society for Magnetic Resonance in Medicine (ISMRM). p. 0070 (2021)
9. Cruz, G., Atkinson, D., Hemmingsson, M., et al.: Highly efficient nonrigid motion-corrected 3d whole-heart coronary vessel wall imaging. Magnetic Resonance in Medicine 77(5), 1894–1908 (2017)
10. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning of probabilistic diffeomorphic registration for images and surfaces. Medical Image Analysis 57, 226–236 (2019)
11. Hammernik, K., Klatzer, T., Kobler, E., et al.: Learning a variational network for reconstruction of accelerated MRI data. Magnetic Resonance in Medicine 79(6), 3055–3071 (2018)
12. Hammernik, K., Pan, J., Rueckert, D., Küstner, T.: Motion-Guided Physics-Based Learning for Cardiac MRI Reconstruction. Asilomar Conference on Signals, Systems, and Computers (2021)
13. Huang, W., Ke, Z., Cui, Z.X., et al.: Deep low-rank plus sparse network for dynamic MRI imaging. Medical Image Analysis 73, 102190 (2021)
14. Jung, H., Sung, K., Nayak, K.S., et al.: k-t focuss: A general compressed sensing framework for high resolution dynamic MRI. Magnetic Resonance in Medicine 61(1), 103–116 (2009)
15. Klein, S., Staring, M., Murphy, K., et al.: Elastix: a toolbox for intensity-based medical image registration. IEEE Trans Med Imag 29(1), 196–205 (2009)
16. von Knobelsdorff-Brenkenhoff, F., Pilz, G., Schulz-Menger, J.: Representation of cardiovascular magnetic resonance in the aha/acc guidelines. Journal of Cardiovascular Magnetic Resonance 19(1), 1–21 (2017)
17. Küstner, T., Fuin, N., Hammernik, K., et al.: CINENet: deep learning-based 3D cardiac CINE MRI reconstruction with multi-coil complex-valued 4D spatio-temporal convolutions. Scientific Reports 10(1), 1–13 (2020)
18. Lee, D., Markl, M., Dall’Armellina, E., et al.: The growth and evolution of cardiovascular magnetic resonance: A 20-year history of the society for cardiovascular magnetic resonance (scmr) annual scientific sessions. Journal of Cardiovascular Magnetic Resonance 20(1) (Jan 2018)
19. Liu, F., Li, D., Jin, X., et al.: Dynamic cardiac MRI reconstruction using motion aligned locally low rank tensor (mallrt). Magnetic Resonance Imaging 66, 104–115 (2020)
20. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
21. Modat, M., Ridgway, G.R., Taylor, Z.A., et al.: Fast free-form deformation using graphics processing units. Computer Methods and Programs in Biomedicine 98(3), 278–284 (2010)
22. Mok, T.C., Chung, A.: Conditional deformable image registration with convolutional neural network. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 35–45. Springer (2021)
23. Odille, F., Vuissoz, P., Marie, P., Felblinger, J.: Generalized reconstruction by inversion of coupled systems (grics) applied to free-breathing MRI. Magnetic Resonance in Medicine 60, 146–157 (2008)
24. Odille, F., Menini, A., Escanyé, J.M., et al.: Joint reconstruction of multiple images and motion in MRI: Application to free-breathing myocardial T2 quantification. IEEE Transactions on Medical Imaging 35(1), 197–207 (2016)
25. Pan, J., Rueckert, D., Küstner, T., Hammernik, K.: Efficient image registration network for non-rigid cardiac motion estimation. In: Haq, N., Johnson, P., Maier, A., Würfl, T., Yoo, J. (eds.) Machine Learning for Medical Image Reconstruction. pp. 14–24 (2021)
26. Poddar, S., Jacob, M.: Dynamic MRI using smoothness regularization on manifolds (storm). IEEE Transactions on Medical Imaging 35(4), 1106–1115 (2016)
27. Pruessmann, K.P., Weiger, M., Börnert, P., Boesiger, P.: Advances in sensitivity encoding with arbitrary k-space trajectories. Magn Reson Med (46), 638–651 (2001)
28. Qi, H., Fuin, N., Cruz, G., et al.: Non-rigid respiratory motion estimation of whole-heart coronary MRI images using unsupervised deep learning. IEEE Transactions on Medical Imaging 40(1), 444–454 (2021)
29. Qi, H., Hajhosseiny, R., Cruz, G., et al.: End-to-end deep learning nonrigid motion-corrected reconstruction for highly accelerated free-breathing coronary mra. Magnetic Resonance in Medicine 86(1), 1983–1996 (2021)
30. Qin, C., Duan, J., Hammernik, K., et al.: Complementary time-frequency domain networks for dynamic parallel mr image reconstruction. Magnetic Resonance in Medicine 86(6), 3274–3291 (2021)
31. Sandino, C.M., Lai, P., Vasanawala, S.S., Cheng, J.Y.: Accelerating cardiac cine mri using a deep learning-based espirit reconstruction. Magnetic Resonance in Medicine 85(1) (2021)
32. Schmoderer, T., Aviles-Rivero, A.I., Corona, V., et al.: Learning optical flow for fast MRI reconstruction. Inverse Problems 37(9), 095007 (aug 2021)
33. Sun, D., Yang, X., Liu, M., Kautz, J.: Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 8934–8943 (2018)
34. Vercauteren, T., Pennec, X., Perchant, A., Ayache, N.: Diffeomorphic demons: Efficient non-parametric image registration. Neuro Image 45(1), S61–S72 (2009)
35. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process 13(4), 600–612 (2004)
Fig. S1. Reconstruction and motion estimation performance of the proposed MCMR method at R=8, 12 and 16 acceleration in the first three iterations (i=1,2,3) during inference. The non-motion compensated CG-SENSE (N-CG-SENSE) and the reference image (R=1) are shown as comparison.