Deep Learning-Based Motion Correction for Cardiovascular Magnetic Resonance Imaging

A Technical Report submitted to the Department of Biomedical Engineering

Presented to the Faculty of the School of Engineering and Applied Science
University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science, School of Engineering

Marina Awad

Spring 2023

On my honor as a University Student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments

Michael Salerno, MD, PhD, Department of Cardiovascular Medicine

Deep Learning-Based Motion Correction for Cardiovascular Magnetic Resonance Imaging

Marina M. Awad[a], Junyu Wang[b], Michael Salerno[b,1]

[a] Department of Biomedical Engineering, University of Virginia
[b] Department of Cardiovascular Medicine, Stanford University
[1] Correspondence: msalerno@stanford.edu

Abstract

Cardiovascular Magnetic Resonance (CMR) first-pass contrast-enhanced myocardial perfusion imaging has proven to be a promising noninvasive technique for evaluating patients with known or suspected coronary artery disease. It is performed during a single breath-hold (typically lasting sixty heartbeats for around 1 minute). However, patients are often not able to hold their breath for this period and involuntary motion of the diaphragm often occurs, especially for senior and pediatric patients. In the presence of cardiac and respiratory motion, CMR perfusion imaging techniques are susceptible to motion-induced artifacts, leading to poor image quality. The inter-frame motion artifacts make quantitative analysis for cardiac function evaluation difficult. Hence motion correction is an important processing step before robust quantification of myocardial perfusion analysis. The motion-compensated reconstruction of CMR perfusion imaging is performed off-line and is a time-consuming method. To address this limitation, we developed a deep learning-based framework for rapid motion correction of spiral first-pass myocardial perfusion imaging using a 2D U-Net that estimates the deformation field from a moving frame to a fixed frame. Additionally, we incorporated auxiliary myocardium mask information to the model to analyze its effects on motion correction performance. Because there are large contrast variations in the temporal series of CMR perfusion data that makes registration of perfusion images inherently difficult, we generated flattened contrast images of the fixed frames using robust principal component analysis (RPCA). Our proposed deep learning-based framework demonstrated faster motion correction than the traditional Advanced Normalization Tools (ANTs) toolbox and temporal smoothness was found to be statistically significant for the models that did not use auxiliary information.

Keywords: CMR perfusion, deep learning, motion correction, myocardial perfusion

Introduction

CMR Perfusion Imaging

About 20.1 million adults aged 20 and older have coronary artery disease (about 7.2%) [1]. First-pass contrast-enhanced myocardial perfusion imaging provides important diagnostic and prognostic information in coronary artery disease [2]. It is performed during a single breath-hold (typically lasting sixty heartbeats for around 1 minute) to limit movement of the heart within and through the imaging plane. However, patients are often not able to hold their breath for this period and involuntary motion of the diaphragm often occurs [3], especially for senior and pediatric patients. Hence, motion between frames must be corrected for quantitative analysis. The sources of cardiac motion are respiration, voluntary patient displacement, involuntary thoracic organ development, and the pumping action of the heart chambers [4].

Furthermore, for detection of coronary artery disease, a pharmacologically induced stress perfusion measurement is required to characterize myocardial perfusion defects. This is typically performed by the administration of adenosine to induce vasodilation. When the heart is under adenosine-induced stress, the ability of the patient to maintain a breath-hold for the duration of the first pass of contrast agent is further compromised.

High-resolution spiral perfusion imaging techniques, using a motion-compensated L1-SPIRiT reconstruction, are capable of whole-heart high-resolution perfusion imaging [5], but the motion-compensated reconstruction is performed off-line and is a time-consuming method, taking ~40 minutes per slice. Furthermore, in the presence of respiratory motion, these

techniques can suffer from significant degradation of image quality because of their sensitivity to respiratory motion-induced artifacts [6].

Small changes in the heart location can lead to the region of interest being contaminated by blood in the left ventricle cavity, resulting in potentially large differences in average myocardial signal intensity which affects quantification of myocardial blood flow. To develop a motion correction framework for spiral first-pass myocardial perfusion imaging that is both efficient and accurate, this project used deep learning. Attempts have been made using deep learning to learn motion correction. These methods have high image quality but have been targeted towards neurological cases with rigid head motion and milder motion displacements compared to CMR cases [7]. By developing a deep learning-based framework for accurate and rapid motion correction of CMR perfusion imaging, our technique aimed to reduce the effect of respiratory motion in the perfusion images without affecting the image quality. Furthermore, a rapid deep learning-based process for motion correction will assist clinicians in receiving faster results for pixel-wise CMR perfusion quantification because respiratory motion correction is an essential component to the workflow pipeline in the imaging process (Figure 1).

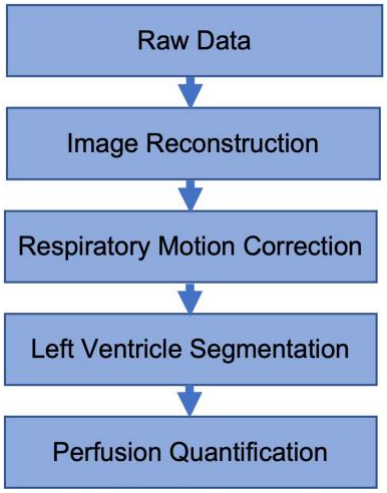

Fig. 1. Workflow diagram of CMR imaging process.

Current Motion Correction Methods

ANTs

The ANTs package extracts information from complex datasets that include imaging, such as CMR perfusion, and depends on the Insight ToolKit (ITK), a widely used medical image processing library to which ANTs developers contribute. It provides many utilities for image preprocessing and registration which have demonstrated excellent performance [8]. Its imaging programs can, for example, perform segmentation, estimate thickness, and perform motion correction for time-series data. However, ANTs is a time-consuming optimization-based method if accurate motion correction is desired. Users must decide which parameters to use such that the output is most optimal for their images. This can cause uncertainties, especially for new users.

VoxelMorph

VoxelMorph is an open-source unsupervised deep learning-based non-linear technique with the architecture of a U-Net that has demonstrated high accuracy in correcting motion and has demonstrated an ease in training methodology [9]. The study registered 3D MR brain scans and has achieved comparable accuracy to state-of-the-art ANTs registration, while taking orders-of-magnitude less time. On a CPU, VoxelMorph requires less than a minute while state of-the-art baselines take tens of minutes to over two hours [9]. An additional advantage of VoxelMorph over other learning-based methods is its end-to-end unsupervised framework. Hence, no ground truth is required, meaning no deformation fields are required during training, providing the network more freedom to estimate the deformation field during the training.

Innovation

Applications of motion correction for medical imaging, and particularly our focus for MRI, correct artifacts to generate good quality and reliable images for clinical interpretation. However, ANTs is time-consuming and VoxelMorph has targeted brain MR cases which have milder motion displacements relative to CMR.

The goal of this project was to deploy VoxelMorph as a backbone framework model for correcting motion of CMR perfusion imaging by applying cardiac cases, changing its structure, and adding preprocessing steps discussed later. By using deep learning instead of ANTs' optimization-method, we aimed to have the motion correction process be more time efficient so that clinicians can receive immediate clinical interpretation. More broadly, the research will assist physicians in their clinical duties by providing good quality images that are not disrupted by blurring artifacts which would affect a physician's evaluation of the image.

Materials and Methods

Data

The specific aims of our project were to develop a deep learning-based respiratory motion correction model (Aim 1) and validate and apply the proposed method on a

clinical setting (Aim 2). We had acquired 76 slices from 17 patients undergoing clinically ordered stress CMR studies with gadolinium (Gd)-based contrast agents on 3T scanners (SIEMENS Prisma/Skyra, Siemens Healthineers, Erlangen, Germany). Datasets were resized to 160 x 160 x 50 (frames). The spiral whole-heart perfusion images have a 1.25 mm in-plane resolution and 10 mm slice thickness [5]. The images were reconstructed using the L1-SPIRiT technique (Figure S1).

For Aim 1, we deployed the deep learning-based 2D U-Net motion correction model, VoxelMorph, to our dynamic perfusion image series. It is a network proposed to estimate the deformation field between image pairs. The code for the VoxelMorph model is freely available in Python library TensorFlow and it was run in Jupyter Notebook in Rivanna.

Data from 13 subjects was used for training for a total of 3,050 images from 61 slices. Data from another 4 subjects was used for validation for a total of 750 images from 15 slices. All data was normalized before being used in the model. Each subject case had between three to eight slices. Training and evaluation were conducted on a single GPU (NVIDIA Tesla A100).

Training Network

During the training process, image pairs in the perfusion dynamic series were randomly selected, and the network learns the deformation mapping from one frame to the other (moving frame *m* to fixed frame *f*) without a gold-standard image registration. The output of the network is the estimated deformation field between the moving and fixed frames, and the moved frame. Motion correction is conducted by applying the deformation field to the moving frame. The network structure for motion correction is demonstrated in Figure 2.

Since CMR perfusion imaging has dynamically varying contrast, we used the normalized cross correlation (NCC) loss function that is less sensitive to contrast variation instead of mean squared error. NCC computed local cross-correlations between fixed frames and moved frames. A higher NCC indicates an accurate deformation field result so that the moved frame is approximate to the fixed frame. Loss was monitored during training.

For the 2D U-Net component in Figure 2, a model with encoder layers of [16, 32, 32, 32] and decoder layers of [32, 32, 32, 32, 32, 16, 16] were used; 16 and 32 represent the number of kernels at each layer. These layers were selected based on what was proposed in Balakrishnan's VoxelMorph models. For the optimizer, we used ADAM. The model was trained for 500 epochs and 150 steps per epoch with a batch size of 5.

Input

A training data generator was created to yield moving image and fixed image pairs for the custom model from the 3,050 images. The fixed frame was selected to be a random frame from the slice that contains the moving frame. Additionally, we trained another model that leverages myocardium masks using manual contours of the endocardium and epicardium (Figure 3) which are available during training but not during testing. In a binary mask, each pixel is labeled either 0 or 1 with 1 representing our region

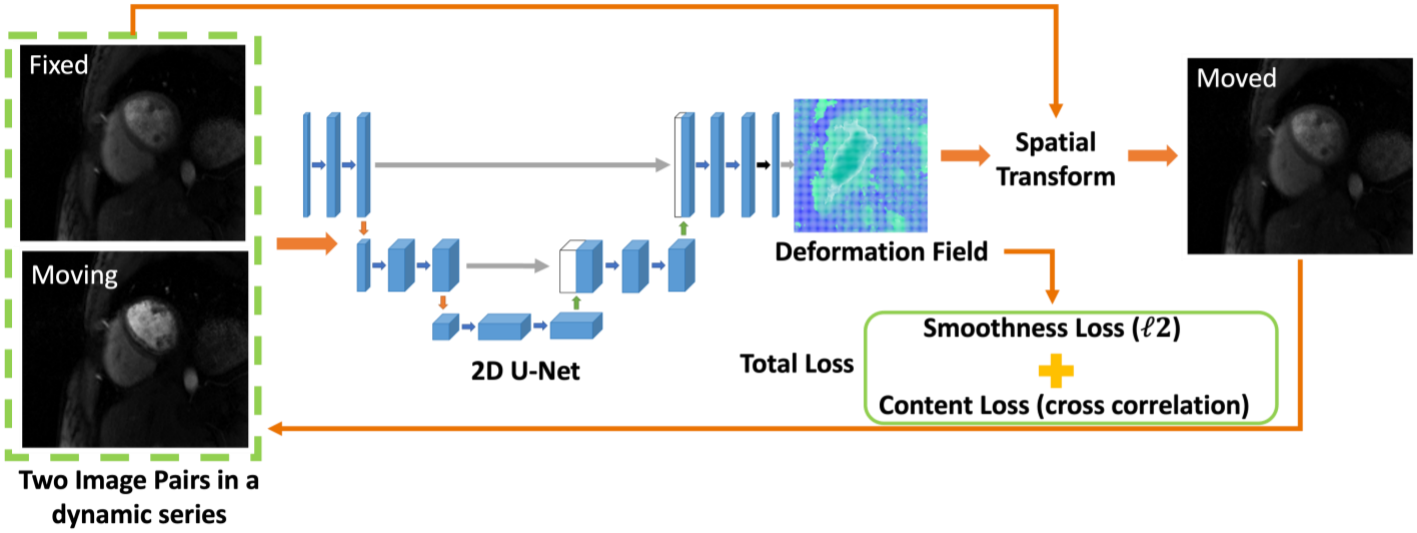

Fig. 2. The proposed image motion correction network for spiral perfusion imaging using deep learning.

of interest, the myocardium area. To incorporate these segmentations into the model, an additional loss function is implemented: Dice. If the deformation field by the model is accurate, then there should be an overlap in anatomical structures (such as of the myocardium mask) in the fixed frame and moved frame. A dice score of 0 means there is no overlap and a dice score of 1 means that there is perfect overlap in the anatomy.

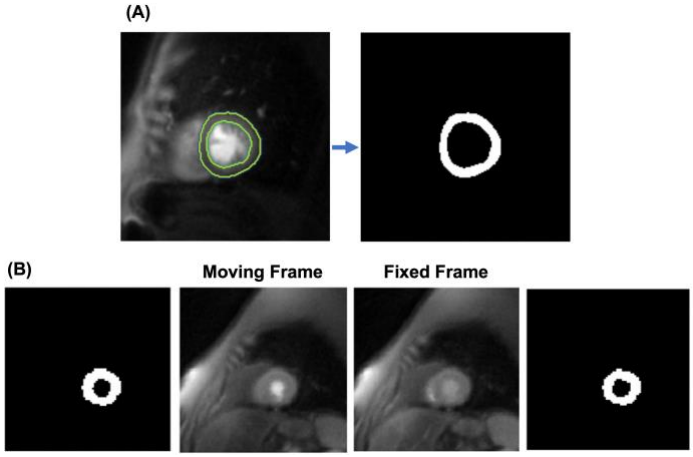

Fig. 3. (A) Example of conversion of manual contours of the endocardium (inner circle) and epicardium (outer circle) into a binary mask of the myocardium. (B) Example of inputs into the model when incorporating auxiliary myocardium mask information.

Contrast Removal

As mentioned, there are large contrast variations in the temporal series of CMR perfusion data which makes registration of perfusion images inherently difficult. There are rapidly changing signal intensities due to the arrival and wash-out of the contrast agent in the region of interest. Algorithms cannot differentiate if differences in signal intensity are caused by spatial motion artifacts or local contrast enhancement [10]. To alleviate this issue, we generated flattened contrast image series during pre-processing of the fixed frames using RPCA. The method separates the local signal enhancement from the baseline signal. Hence, the deformation field needed to remove respiratory motion can be calculated in the absence of the locally varying contrast enhancement, so the deformation field is then applied to the moving image to create the moved image.

When looking at a set of images in a slice, the data (M) can be viewed as a combination of a low-rank component (L) and a sparse component (S) [10]. This is formulated as:

$$M = L + S \qquad [1]$$

To optimally extract the low rank and sparse components, a parameter λ which balances the constraint on the rank of L and the sparsity of S is typically set as:

$$\lambda = 1/\sqrt{N_p} \qquad [2]$$

N_p is the number of pixels in an image, which in this case was set to 25,600 pixels. Hence our λ was optimally selected to be 0.00625. As λ increases, this leads to the low-rank components having higher rank and the sparse component being more sparse [10]. We used the low-rank component to represent the image with removal of contrast enhancement as it models the baseline signal. The sparse component itself models the contrast enhancement. Figure 4 displays an example of flattening contrast using RPCA by extracting the low-rank components of the image M.

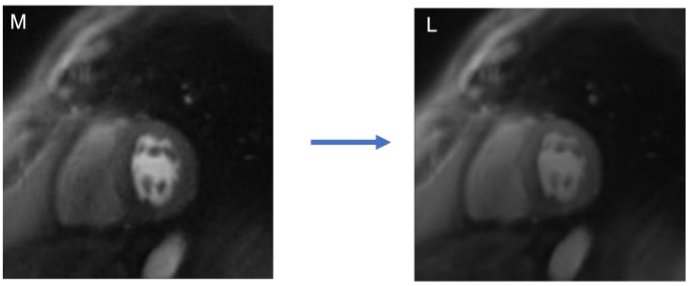

Fig. 4. Frame after removal of contrast using RPCA by separating image series M into its low-rank component L.

Testing Data Pairs

To test our deep learning-based model, a testing data generator was created to form moving and fixed image pairs. The fixed frame was the 25th frame in the slice containing the moving frame. The 25th frame was selected because that is the middle frame within a slice of 50 frames. Hence, it can be represented as a general representation of the slice for the moving frame to register to. This is important as the model registers the moving frame to the fixed frame by mapping it to the fixed space.

Implementation of ANTs

In Python, we used "BOLDAffine", an affine transformation, in the publicly available software package ANTs to do motion correction on our four CMR perfusion testing cases. All other parameters were kept to the default setting such as 0.2 for the gradient step size, 3 for the smoothing for the update field, and 0 for the smoothing for the total field. The fixed image was similarly also selected to be the 25th frame in the slice containing the moving frame as ANTs is also a pairwise registration.

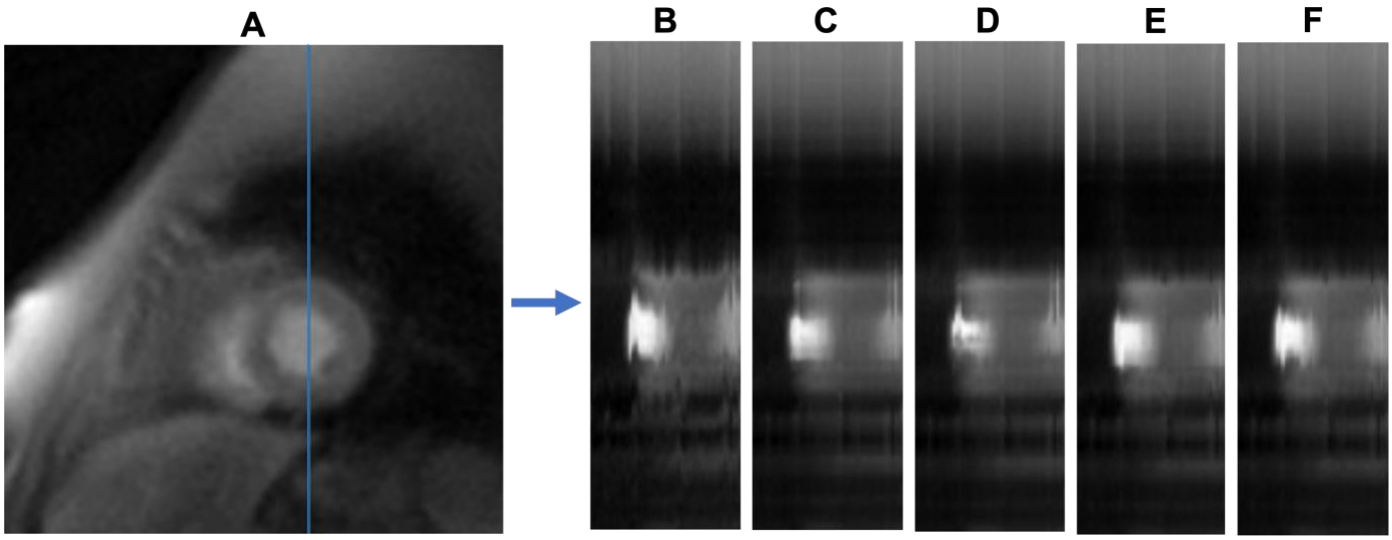

Fig. 5. (A) X-t profiles of a slice (50 Frames), denoted by the blue line. (B) No motion correction, (C) After DL MOCO, (D) After DL MOCO with RPCA, (E) After DL MOCO using Myocardium Mask, (F) After DL MOCO using Myocardium Mask with RPCA.

Results

X-t Profile

Using data from four subjects, we tested the different alterations of our deep learning-based model: with RPCA, using myocardium mask information, and using myocardium mask information with RPCA. To best visualize the results, the x-t profile of a slice (50 frames) was plotted, before motion correction and after motion correction (Figure 5). A blue line is drawn indicating the location we sought to observe, where the left ventricle cavity is present. Based on the x-t profile, our models reduce motion when compared to the no motion correction x-t profile (Figure 5B) where the up and down movements of the left ventricle are more apparent because of the breathing of the patient.

Temporal Smoothness

To quantify the correction of motion, the standard deviation (SD) of the second derivative of the voxel-wise time-intensity curves [10] was calculated for the 15 slices used for testing in MATLAB. Lower values indicate that the change in intensity between two successive images in the series is smooth and hence indicates a likely reduction in the amount of motion. From the boxplots in Figures 6 and S2, it is seen that all four of our deep learning-based models reduce motion as the SD of the second derivative of the intensities is lower.

Statistical testing was done by the Bonferroni-Holm Method to determine if there are any models that have a statistically significant difference in the SD of the second derivative of the intensities. It was found that the DL MOCO and DL MOCO with RPCA models have a statistically significant difference compared to no motion correction with p-values of 0.0357 and 0.0331 respectively. The models using myocardium masks had no significant difference ($p>0.05$).

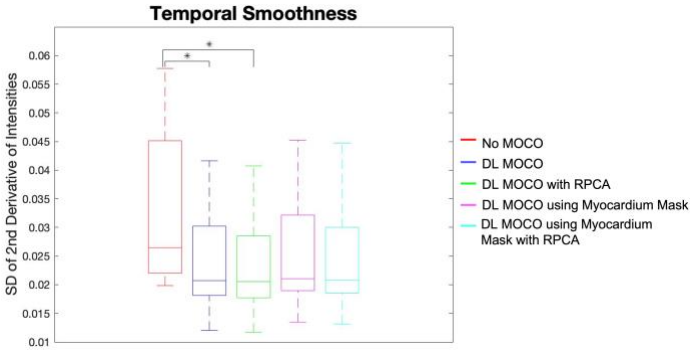

Fig. 6. Box plots indicating temporal smoothness. Statistical test is conducted using the Bonferroni-Holm Method and p-values of 0.0357 and 0.0331 for no MOCO compared to DL MOCO and DL MOCO with RPCA respectively are observed.

Dice Score

Additionally, we sought to compute the dice score overlap of the endocardium and myocardium regions of the manual segmentations on the fixed frames compared to warped segmentations made on the moved frames done by the model. For every pair of contours on the fixed frames and moved frames, a dice score was calculated for the endocardium and myocardium areas. The two dice scores were then averaged which was done for all 750 images of testing data. The dice score is not an indicator of smoothness but rather evaluates similarities between the segmentations. We were limited to using only myocardium

and endocardium masks because segmentations of other regions were not available. The average overall dice score was highest for the models using myocardium mask information during training (Figure 7, S3) with dice scores of 0.824 and 0.822 for DL MOCO using Myocardium Mask and DL MOCO using Myocardium Mask with RPCA respectively.

Based on a Tukey Honestly Significant Difference Test, it was found that there was a statistically significant difference in the average dice score from models using myocardium mask information when compared to the DL MOCO with RPCA. Specifically, the p-value for the pairwise comparison between DL MOCO with RPCA and DL MOCO using Myocardium Mask was 0.00469 and the p-value for the pairwise comparison between DL MOCO with RPCA and DL MOCO using Myocardium Mask with RPCA was 0.0115.

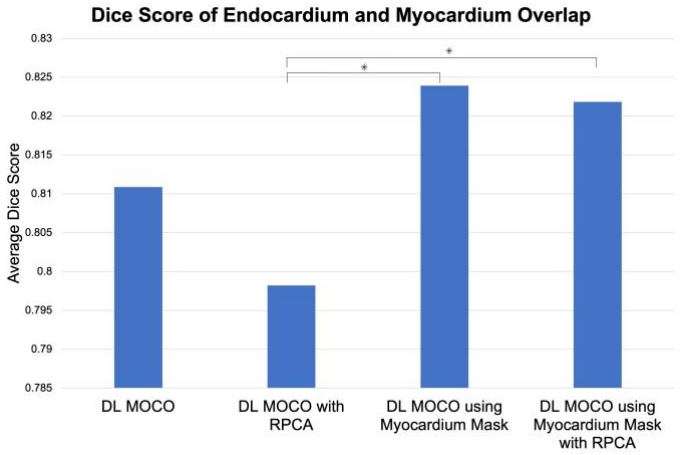

Fig. 7. Average dice score of endocardium and myocardium region, measuring segmentation overlap between fixed frames and moved frames. Statistical test is conducted using the Tukey Honestly Significant Difference Test and p-values of 0.00469 and 0.0115 for DL MOCO with RPCA compared to DL MOCO using Myocardium Mask and DL MOCO using Myocardium Mask with RPCA respectively are observed.

Time Efficiency

Time efficiency of our proposed deep learning-based motion correction technique was assessed by using a built-in Python function to measure execution time. A two-sample t-test was conducted to determine if there is a statistically significant difference between the time to register a frame using ANTs, a state-of-the-art medical image registration toolbox, and our technique. The average time using the deep learning approach, which was found to be independent of flattening contrast and adding myocardium mask information to the model, took 14 milliseconds and ANTs took 25 times more time (Figure 8). The test concluded that there is a statistically significant difference with a p-value of 4.52E-43.

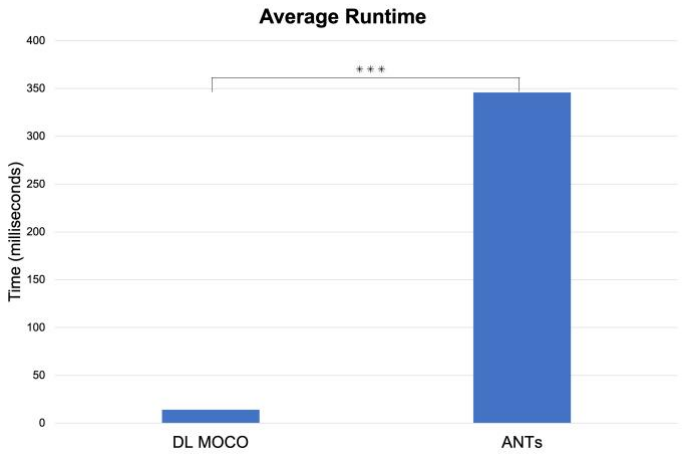

Fig. 8. Bar graph measuring average runtime for motion correction. A two-sample t-test was conducted which yielded a p-value of 4.52E-43.

Discussion

Based on the SD of the second derivative of the voxel-wise time-intensity curves, the deep learning-based models without the use of myocardium mask information reduced motion the most. The implementation of removing contrast of the fixed frame during testing did not significantly reduce motion. The additional segmentation information of the myocardium likely did not lead to a lower SD of the second derivative of the intensities because it provided an additional emphasis on the myocardium region. Implementing a model that uses multiple masks such as that of the right ventricle too can assist in increasing the temporal smoothness and learning network parameters.

With NCC as a loss function to quantify the dissimilarity between the intensities of two frames and the spatial regulation of the deformation in our models [9], it is likely that flattening contrast did not significantly affect motion correction results because NCC is insensitive to contrast variation. Future studies can analyze how using a MSE loss function for a model that uses flattened contrast images affects motion correction.

Overall, our deep learning-based technique is rapid in automatically reducing motion of CMR perfusion imaging. This is essential to be clinically translated and to assist in providing immediate feedback to clinicians after a scan. The results presented are a step closer to the elimination of motion and its related artifacts in CMR perfusion imaging so that clinicians can robustly quantify myocardial blood flow and read images with excellent image quality.

Future Work

Potential ways in increasing the reduction of motion in CMR perfusion images could include improving the input quality of the images. In this project, we used L1-SPIRiT for image reconstruction. However, alternative techniques such as DESIRE, a deep learning-based image reconstruction technique where complex-valued convolution is enforced [11], has also demonstrated good image quality. An advantage in using DESIRE is that its image reconstruction time is significantly shorter than L1-SPIRiT. Using DESIRE would help the overall pipeline process seen in Figure 1 occur more rapidly.

As mentioned previously, we were limited to creating masks from manual contours of the endocardium and epicardium as those were the manual contours available to us in our cases used for testing. However, further studies can be done to observe how utilizing more masks can assist the deep learning-based motion correction model in learning network parameters. This would require additional manual contours of regions such as the right ventricle and other structures.

In this project, RPCA was utilized to remove contrast of the fixed frames. However, alternative techniques can be used to flatten contrast such as Contrastive Unpaired Image-to-Image Translation (CUT) [12], a neural style transfer network. CUT uses an InfoNCE loss to maximize mutual information between corresponding input and output patches, while drawing upon other patches in the image as contrastive negatives. It can be used to flatten contrast so that frames with low contrast resemble contrast level of middle slice perfusion frames.

Furthermore, the performance of motion correction by ANTs heavily relies on the parameters inputted by the user. In future studies, more research can be done to analyze if there are more optimal parameters to be selected. We tested several transformations in the ANTs package and discovered "BOLDAffine" to be the most sufficient in our cases. It would be beneficial to also observe how changing other parameters can affect motion correction such as the choice of the fixed frame, gradient step sizes, and smoothing for the update field and total field.

End Matter

Author Contributions and Notes

M.M.A performed research. J.W. and M.S. advised the project. The authors declare no conflict of interest.

Acknowledgments

Alper Turgut, School of Medicine, University of Virginia

References

[1] C. W. Tsao et al., "Heart Disease and Stroke Statistics-2022 Update: A Report From the American Heart Association," Circulation, vol. 145, no. 8, pp. e153–e639, Feb. 2022, doi: 10.1161/CIR.0000000000001052.

[2] M. Gulati et al., "2021 AHA/ACC/ASE/CHEST/SAEM/SCCT/SCMR Guideline for the Evaluation and Diagnosis of Chest Pain: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines," Circulation, vol. 144, no. 22, pp. e368–e454, Nov. 2021, doi: 10.1161/CIR.0000000000001029.

[3] B. Pontré et al., "An Open Benchmark Challenge for Motion Correction of Myocardial Perfusion MRI," IEEE J Biomed Health Inform, vol. 21, no. 5, pp. 1315–1326, Sep. 2017, doi: 10.1109/JBHI.2016.2597145.

[4] M. Benovoy, M. Jacobs, F. Cheriet, N. Dahdah, A. E. Arai, and L.-Y. Hsu, "A Robust Universal Nonrigid Motion Correction Framework for First-Pass Cardiac Magnetic Resonance Perfusion Imaging," J Magn Reson Imaging, vol. 46, no. 4, pp. 1060–1072, Oct. 2017, doi: 10.1002/jmri.25659.

[5] J. Wang et al., "High spatial resolution spiral first-pass myocardial perfusion imaging with whole-heart coverage at 3 T," Magnetic Resonance in Medicine, vol. 86, no. 2, pp. 648–662, 2021, doi: 10.1002/mrm.28701.

[6] R. Zhou et al., "Simple motion correction strategy reduces respiratory-induced motion artifacts for k-t accelerated and compressed-sensing cardiovascular magnetic resonance perfusion imaging," Journal of Cardiovascular Magnetic Resonance, vol. 20, no. 1, p. 6, Feb. 2018, doi: 10.1186/s12968-018-0427-1.

[7] T. Küstner, K. Armanious, J. Yang, B. Yang, F. Schick, and S. Gatidis, "Retrospective correction of motion-affected MR images using deep learning frameworks," Magnetic Resonance in Medicine, vol. 82, no. 4, pp. 1527–1540, 2019, doi: 10.1002/mrm.27783.

[8] O. Camara, M. Pop, K. Rhode, M. Sermesant, N. Smith, and A. Young, Eds., Statistical Atlases and Computational Models of the Heart: First International Workshop, STACOM 2010, and Cardiac Electrophysiological Simulation Challenge, CESC 2010, Held in Conjunction with MICCAI 2010, Beijing, China, September 20, 2010. Proceedings, vol.

6364. in Lecture Notes in Computer Science, vol. 6364. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010. doi: 10.1007/978-3-642-15835-3.

[9] G. Balakrishnan, A. Zhao, M. R. Sabuncu, J. Guttag, and A. V. Dalca, "VoxelMorph: A Learning Framework for Deformable Medical Image Registration," IEEE Trans. Med. Imaging, vol. 38, no. 8, pp. 1788–1800, Aug. 2019, doi: 10.1109/TMI.2019.2897538.

[10] C. M. Scannell, A. D. M. Villa, J. Lee, M. Breeuwer, and A. Chiribiri, "Robust Non-Rigid Motion Compensation of Free-Breathing Myocardial Perfusion MRI Data," IEEE Transactions on Medical Imaging, vol. 38, no. 8, pp. 1812–1820, Aug. 2019, doi: 10.1109/TMI.2019.2897044.

[11] J. Wang, D. S. Weller, C. M. Kramer, and M. Salerno, "DEep learning-based rapid Spiral Image REconstruction (DESIRE) for high-resolution spiral first-pass myocardial perfusion imaging," NMR Biomed, vol. 35, no. 5, p. e4661, May 2022, doi: 10.1002/nbm.4661.

[12] T. Park, A. A. Efros, R. Zhang, and J.-Y. Zhu, "Contrastive Learning for Unpaired Image-to-Image Translation." arXiv, Aug. 20, 2020. Accessed: Nov. 10, 2022. [Online]. Available: http://arxiv.org/abs/2007.15651

Supplementary Material

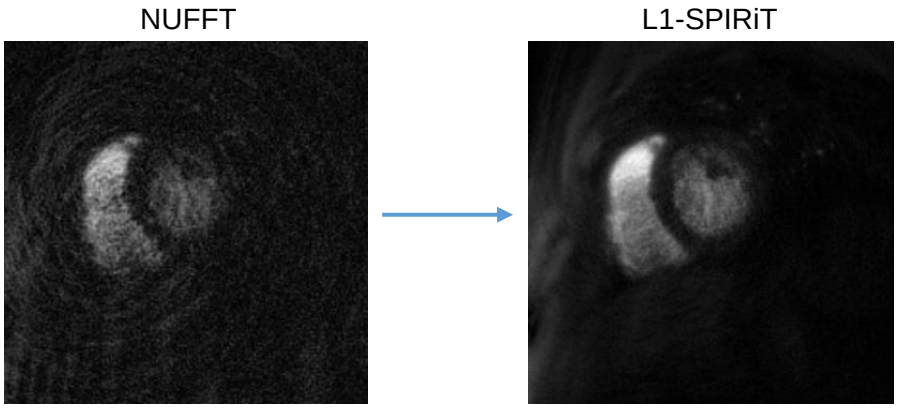

Fig. S1. Example of L1-SPIRiT image reconstruction on a CMR perfusion image.

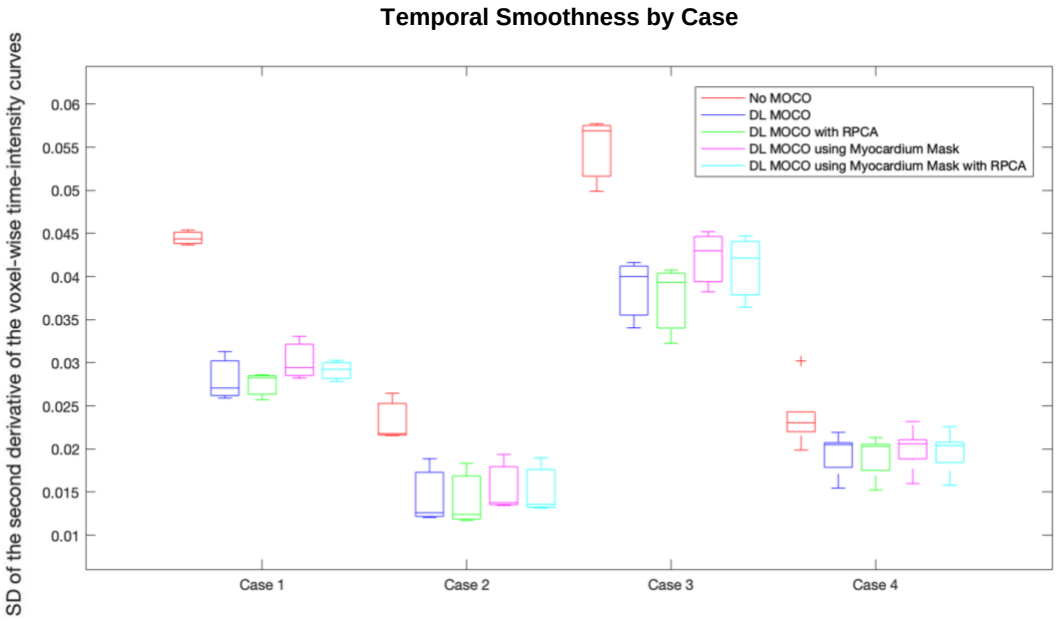

Fig. S2. Box plots indicating temporal smoothness for each testing case, calculated by the SD of the second derivative of the voxel-wise intensity curves.

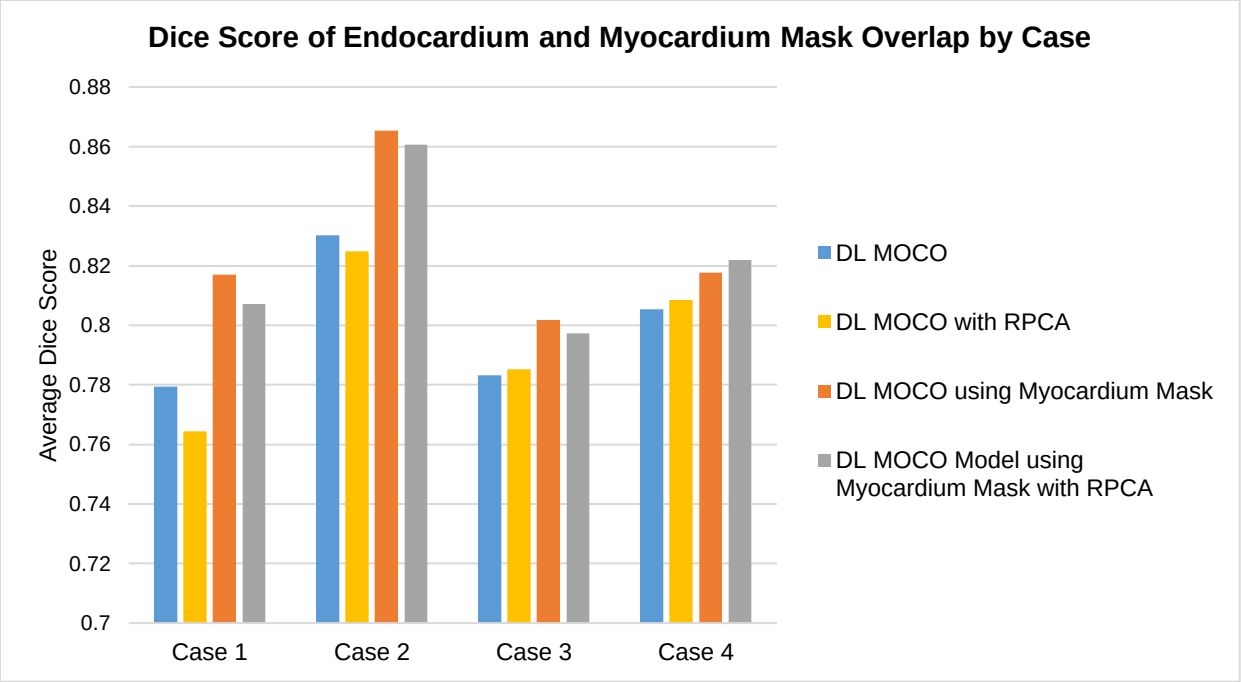

Fig. S3. Average dice score by case of the endocardium and myocardium region, measuring segmentation overlap between fixed frames and moved frames.