Journal Pre-proof

Dynamic MRI Reconstruction with End-to-end Motion-Guided Network

Qiaoying Huang, Yikun Xian, Dong Yang, Hui Qu, Jingru Yi, Pengxiang Wu, Dimitris N. Metaxas

PII: S1361-8415(20)30265-6
DOI: https://doi.org/10.1016/j.media.2020.101901
Reference: MEDIMA 101901

To appear in: *Medical Image Analysis*

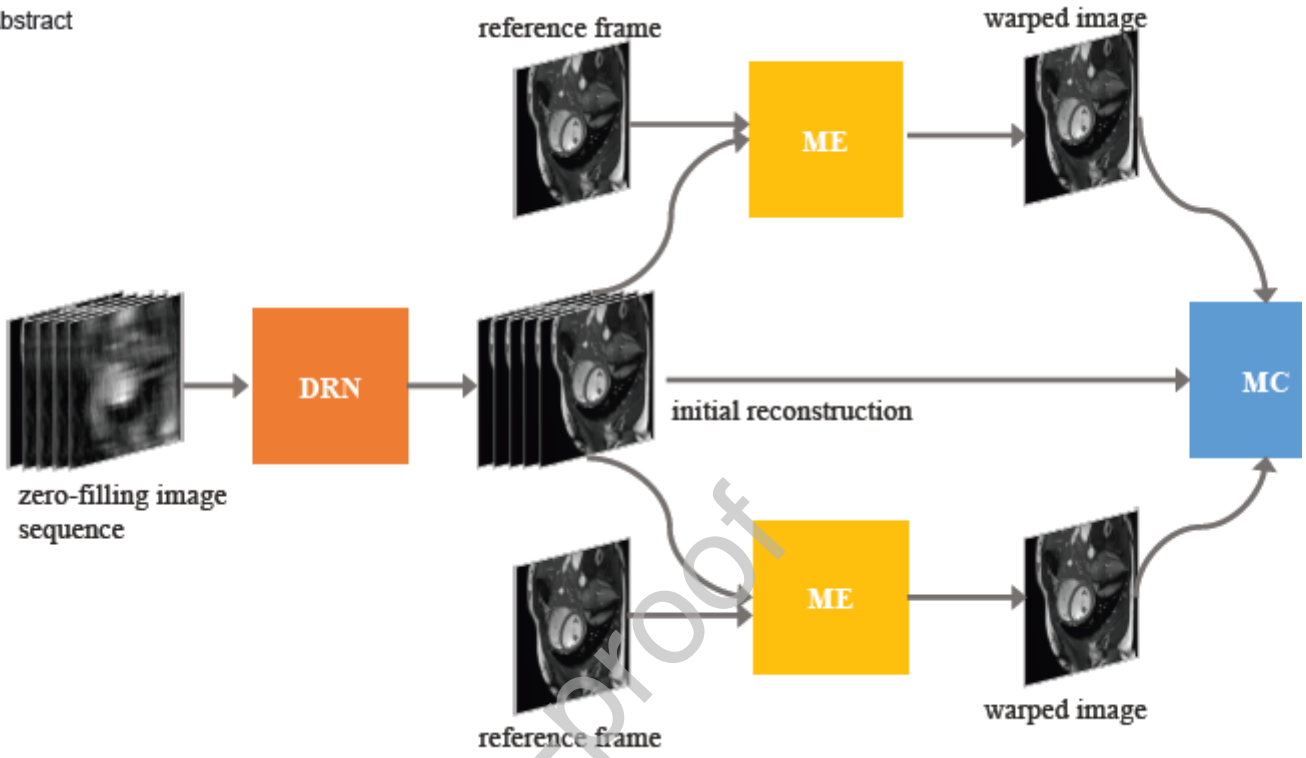
Received date: 9 September 2019
Revised date: 23 October 2020
Accepted date: 9 November 2020

Please cite this article as: Qiaoying Huang, Yikun Xian, Dong Yang, Hui Qu, Jingru Yi, Pengxiang Wu, Dimitris N. Metaxas, Dynamic MRI Reconstruction with End-to-end Motion-Guided Network, *Medical Image Analysis* (2020), doi: https://doi.org/10.1016/j.media.2020.101901

Highlights

- A recurrent neural network for spatial-temporal data reconstruction
- A network for motion estimation with a novel loss capturing long-term motion information
- Two versions of motion-guided deep neural networks for dynamic MRI reconstruction

*Graphical Abstract



Medical Image Analysis (2020)

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

Dynamic MRI Reconstruction with End-to-end Motion-Guided Network

Qiaoying Huang[a,*], Yikun Xian[a], Dong Yang[b], Hui Qu[a], Jingru Yi[a], Pengxiang Wu[a], Dimitris N. Metaxas[a]

[a]*Department of Computer Science,Rutgers University, Piscataway, NJ 08854, USA*
[b]*NVIDIA, Bethesda, MD 20814, USA*

ARTICLE INFO

Article history:
Received xxx
Received in final form xxx
Accepted xxx
Available online xxx

Communicated by xxx

Keywords: Dynamic MRI Reconstruction, Motion Estimation, Motion Compensation

ABSTRACT

Temporal correlation in dynamic magnetic resonance imaging (MRI), such as cardiac MRI, is informative and important to understand motion mechanisms of body regions. Modeling such information into the MRI reconstruction process produces temporally coherent image sequence and reduces imaging artifacts and blurring. However, existing deep learning based approaches neglect motion information during the reconstruction procedure, while traditional motion-guided methods are hindered by heuristic parameter tuning and long inference time. We propose a novel dynamic MRI reconstruction approach called MODRN and an end-to-end improved version called MODRN(e2e), both of which enhance the reconstruction quality by infusing motion information into the modeling process with deep neural networks. The central idea is to decompose the motion-guided optimization problem of dynamic MRI reconstruction into three components: Dynamic Reconstruction Network, Motion Estimation and Motion Compensation. Extensive experiments have demonstrated the effectiveness of our proposed approach compared to other state-of-the-art approaches.

1. Introduction

Magnetic Resonance Imaging (MRI) is a widely-used medical imaging modality in routine clinical practice, such as disease diagnosis and treatment monitoring. One traditional limitation of MRI is slow acquisition, and therefore, raw data collected from MRI scan is usually under-sampled in frequency domain (so-called *k*-space). Here comes a natural problem: how can we effectively and efficiently reconstruct a high-quality image from under-sampled *k*-space data? This problem is referred to as MRI reconstruction, and involves 2D reconstruction and dynamic reconstruction (*i.e.* , 2D plus temporal axis).

This paper specifically focuses on dynamic MRI that is critical in practical applications, e.g., cardiovascular and pulmonary. However, reconstruction of high spatiotemporal resolution from under-sampled MRI *k*-space data is very challenging due to the lack of a gold standard in clinical practice and the strong dependence on heuristic parameter tuning (Yang et al., 2016). Many works have been proposed to tackle these challenges. Some of them argued that motion plays a leading role in dynamic reconstruction because correlation and redundancy exist along the temporal dimension, such as cardiac moves relatively periodically against the static background (Jung et al., 2010; Jung and Ye, 2010).

In particular, one line of research lies in traditional Compressed Sensing (CS) approaches that have dominated dynamic reconstruction in the past few years. Some studies have successfully incorporated the physical motion into the CS schemes to improve reconstruction performance by refining the results after the image reconstruction step (Jung et al., 2009; Cruz et al., 2016). These methods calculated displacement motion fields or

*Corresponding author.
e-mail: qh55@cs.rutgers.edu (Qiaoying Huang), yx150@cs.rutgers.edu (Yikun Xian), don.yang.mech@gmail.com (Dong Yang), hui.qu@cs.rutgers.edu (Hui Qu), jy486@cs.rutgers.edu (Jingru Yi), pw241@cs.rutgers.edu (Pengxiang Wu), dnm@cs.rutgers.edu (Dimitris N. Metaxas)

optical flow to estimate motion between image pairs. In terms of the form of motion constraint, many algorithms were proposed, among which, Horn-Schunck (Horn and Schunck, 1981) and Lucas-Kanade (Lucas et al., 1981) are widely adopted. However, two major drawbacks of these methods are that they require heuristic parameter tuning and long reconstruction time. Additionally, motion estimation alone is a very difficult task because inaccurate results are easily incurred by large or fast object movement. Therefore, there is still much room for improvement in dynamic MRI reconstruction.

Recent advances in deep learning technique have sparked the new research interests in MRI reconstruction. Deep convolutional neural networks have been employed to learn mapping directly from k-space data to fully-sampled reconstructed image, and are considered as a promising alternative for MRI reconstruction (Zhu et al., 2018). Compared to CS-based methods, deep learning approaches are faster in inference and learn the implicit prior automatically based on the training data (Sun et al., 2016; Schlemper et al., 2017, 2018; Lønning et al., 2018; Qin et al., 2018b; Huang et al., 2019b). However, most of them were designed only for 2D reconstruction problem, but not applicable to dynamic MRI reconstruction, which still remains largely unsolved yet critical in clinical scenarios. These state-of-the-art methods tend to generate blurry or temporal inconsistent results because they did not take motion information into consideration.

To the end, we develop a deep neural model called Motion-guided Dynamic Reconstruction Network (MODRN) that explicitly injects motion information into the reconstruction process via motion estimation and motion compensation (ME/MC) to improve the reconstruction quality of spatiotemporal imaging. Different from traditional motion estimation algorithms which may fail in low resolution and weak contrast, we employ the unsupervised deep learning model with optical flow estimation (Ren et al., 2017; Meister et al., 2017), which is more robust and accurate in different scenarios. To the best of our knowledge, this is the first work that embeds motion information into deep neural network for dynamic MRI reconstruction. The contribution of this work are four aspects. Firstly, inspired by the optimization procedures of model-based dynamic reconstruction, we derive a recurrent neural network, which simultaneously links the relationship of data over time and iterations. Secondly, we introduce an unsupervised deep learning based motion estimation method to learn the motion between the reference image and the reconstructed image by using the combination of forward, backward and neighboring loss. Thirdly, we present a motion compensation network for refining the reconstructed images with the guide of the learned motion. Finally, we propose an end-to-end framework that combines all the subnetworks to further improve the performance. The source code is available at

This manuscript is an extension of our previous work (Huang et al., 2019a). It inherits the main model but further improve it by end-to-end training. We extend the manuscript by adding related work, ablation studies and comparison experiments, also with minor change of the methodology section.

2. Related Work

Motion information has been exploited in various computer vision problems, such as optical flow estimation (Caballero et al., 2017; Makansi et al., 2017), video super resolution (SR) (Makansi et al., 2017) and enhancement (Bao et al., 2018) and motion debluring (Jin et al., 2019). Since information along the temporal dimension is highly related, estimating accurate motion is beneficial for the model to share feature and improve performance. In medical image analysis, motion information has been successfully applied to the problems including dynamic segmentation (Qin et al., 2018a), classification (Zheng et al., 2019) and tracking (Yi et al., 2019). However, motion information has seldom been considered to incorporate into dynamic reconstruction problem solved by deep learning methods. In the following section, we first review 2D reconstruction methods as well as dynamic reconstruction approaches, and then analyze how motion information plays an important role in conventional computer vision problems.

2.1. 2D MRI Reconstruction

Deep neural network (DNN) is an effective technique due to its powerful ability of feature learning and representation. To name a few, Sun et al. (2016) first proposed a new network called ADMM-Net based on the Alternating Direction Method of Multipliers (ADMM) framework and showed that DNN achieved promising performance in 2D MRI reconstruction with less computation time compared with traditional CS-based reconstruction methods. Similarly, Huang et al. (2019c) proposed a FR-Net according to the fast iterative shrinkage-thresholding (FISTA) algorithm and demonstrated promising results in 2D cardiac reconstruction. Schlemper et al. (2017) regarded image reconstruction as a de-aliasing problem and proposed a deep neural network with cascading convolutional neural networks (CNN) and also proposed a differentiable data consistency (DC) layer to incorporate the data fidelity. Besides, Hyun et al. (2018) proposed a variant of U-Net (Ronneberger et al., 2015) for 2D MRI reconstruction. Huang et al. (2019b) further improved the reconstruction result by using cascaded U-Net and channel-wise attention techniques (Hu et al., 2018). Furthermore, generative adversarial network (GAN) is employed by Yang et al. (2018) who proposed a variant called DAGAN to generate a high-resolution and noise-free image using mean square error loss and perceptual loss. Quan et al. (2018) proposed to train the generator and discriminator with cyclic-consistency loss and they further improved reconstruction result with a two-fold chained network. Other researches exploit different neural network structures, for example, Hammernik et al. (2018) formulated the reconstruction process with a variational network that combines the mathematical structure of variational models and they demonstrated promising results on multi-coil knee data. More straightforward, Tezcan et al. (2018) learned the distribution of the MR image patches as prior information by a variational autoencoder (VAE) model to help reconstruction.

All the aforementioned methods have proved that deep learning is efficient and effective in 2D reconstruction. However,

only limited works apply deep learning into dynamic MRI reconstruction, hence the applicability of deep learning to this problem still need to be explored.

2.2. Dynamic MRI Reconstruction

So far, only a few works apply deep neural network into dynamic reconstruction. For example, Schlemper et al. (2018) adapted their 2D reconstruction model to dynamic reconstruction and also introduced a data sharing (DS) layer to regularize the temporal correlation between different frames. Qin et al. (2018b) modeled the iterative procedures of the reconstruction problem by a recurrent network with two different types of recurrent units that implies information over time and iteration. The CS-based methods also exploited the correlations between k-space and time. For example, Lingala et al. (2011) proposed k-t SLR that reconstructs under-sampled data with a low rank and sparsity regularization. Jung et al. (2009) proposed k-t FOCUSS that tries to improve the performance by formalizing the reconstruction problem with a sparsity regularization in the x-f domain.

However, none of the above approaches have incorporated motion information. The most recent work is Seegoolam et al. (2019), where motion information is used to improve dynamic MRI reconstruction. They try to estimate the motion in each cascade from the intermediate reconstructed image, which may contain noise and artifacts that influence the learning process. k-t FOCUSS+ME/MC (Jung et al., 2009; Jung and Ye, 2010) used interframes motion estimation and compensation with a fixed reference frame during the image recovery process for cardiac MRI data. Inspired by the k-t FOCUSS+ME/MC method, we utilize multiple fully-sampled frames as reference and learn the motion between the reference frames and the reconstructed images to compensate the detail of the reconstructed image. Similar ideas can be found in SR and texture transfer approaches (Caballero et al., 2017; Zhang et al., 2019) where they introduced additional images to assist the learning process by taking advantage of the similarity nature of referenced images. They argued that the more similar the target image and the reference image are, the easier it is to learn the texture information from the referenced image. This assumption is matched with cardiac reconstruction scenarios, since images are highly related along the temporal dimension. Therefore, we propose an end-to-end trainable Motion-guided Dynamic Reconstruction Network model that employs motion estimation and compensation to further improve the quality of the reconstructed image.

3. Problem Formulation

We start with the formulation of dynamic MRI reconstruction problem with motion estimation.

Formally, given a sequence of under-sampled k-space data $\{y_t\}_{t\in[\tau]}$ of τ frames in a complete cardiac cycle, the dynamic MRI reconstruction problem is to predict reconstructed images $\{z_t\}_{t\in[\tau]}$ from $\{y_t\}$, which can be formalized as an optimization problem:

$$\underset{\{z_t\}}{\operatorname{argmin}}\ \mathcal{L}_{rec}(\{z_t\}), \tag{1}$$

where

$$\mathcal{L}_{rec}(\{z_t\}) = \sum_{t=1}^{\tau} \frac{1}{2}\|F_u(z_t) - y_t\|_2^2 + \lambda R(z_t). \tag{2}$$

The first fidelity term $\|F_u(z_t) - y_t\|_2^2$ is used to guarantee data consistency by restricting the reconstructed image z_t to be close to the input measurement y_t. $F_u(\cdot)$ is an operator that transforms image-domain z_t into Fourier domain followed by undersampling. The second term $R(\cdot)$ is a regularization function that depends on the prior knowledge of the input $\{y_t\}$. Common choices include sparsity in transformed domain (Lingala et al., 2011), total variation (TV) penalties (Knoll et al., 2012) and low-rank property (Trzasko et al., 2011). λ is a weighting factor.

In order to capture anatomical motion in the dynamic MRI acquisition, it is natural to incorporate motion estimation/motion compensation (ME/MC) technique in the reconstruction process (Jung et al., 2010). Specifically, based on the brightness constancy assumption (Horn and Schunck, 1981), for the reconstructed image sequence $z = \{z_t\}$, if a voxel at position (x, y, t) with intensity $z(x, y, t)$ has small movement $(\Delta x, \Delta y, \Delta t)$ with respect to the next frame, it satisfies that

$$\begin{aligned} z(x, y, t) &= z(x + \Delta x, y + \Delta y, t + \Delta t) \\ &\approx z(x, y, t) + \frac{\partial z}{\partial x}\Delta x + \frac{\partial z}{\partial y}\Delta y + \frac{\partial z}{\partial t}\Delta t. \end{aligned} \tag{3}$$

Denote the gradient of the space-time image z at position (x, y) by $\nabla z = \left(\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}\right)$, and the estimated displacement motion fields or optical flow of z by $\mathbf{v} = \left(\frac{\Delta x}{\Delta t}, \frac{\Delta y}{\Delta t}\right)$. Eq. (3) can be simplified as

$$0 = \nabla z^\top \mathbf{v} + \frac{\partial z}{\partial t}, \tag{4}$$

which is underdetermined with variable $\mathbf{v}$. Therefore, we add an additional l_1 regularization on the predicted $\mathbf{v}$ and derive the optical flow constraint as below:

$$\mathcal{L}_{me}(\mathbf{v}) = \left\|\nabla z^\top \mathbf{v} + \frac{\partial z}{\partial t}\right\|_1 + \delta\|\mathbf{v}\|_1. \tag{5}$$

This motion estimation constraint can be incorporated into the objective function (2) to jointly solve both reconstruction and motion estimation problems:

$$\underset{\{z_t, \mathbf{v}\}}{\operatorname{argmin}}\ \mathcal{L}_{rec}(\{z_t\}) + \eta\mathcal{L}_{me}(\mathbf{v}). \tag{6}$$

The solution to problem (6) is non-trivial and traditional CS-based algorithms are usually computationally expensive and require long running time for hyper-parameter tuning. Recent advances in deep learning provide an alternative way for efficient MRI reconstruction, but very few works focus on the dynamic reconstruction problem and they only target for the simpler problem (2) without considering motion information. Therefore, it is essential to explicitly incorporate motion information into deep neural model to solve the joint problem (6).

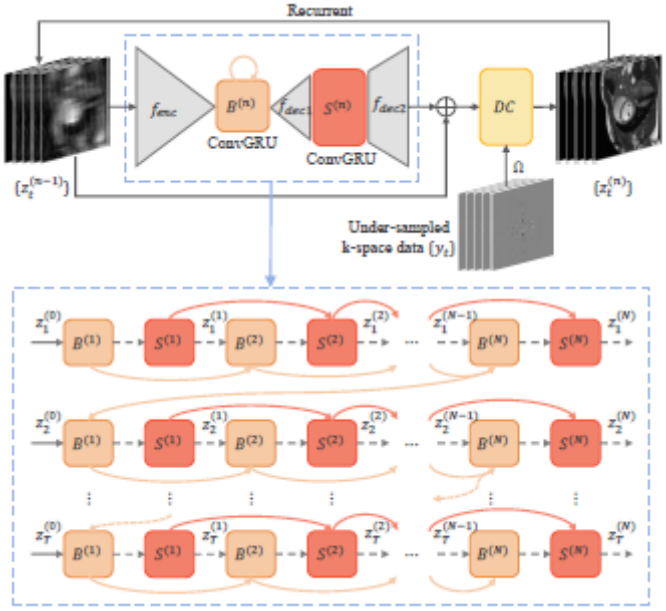


Fig. 1. Network architecture of DRN. It takes as input the zero-filling images (inverse Fourier transform of under-sampled k-space data) and outputs reconstructed images. The reconstruction procedure has N iterative steps. The blue box illustrates the workflow of ConvGRU units B and S.

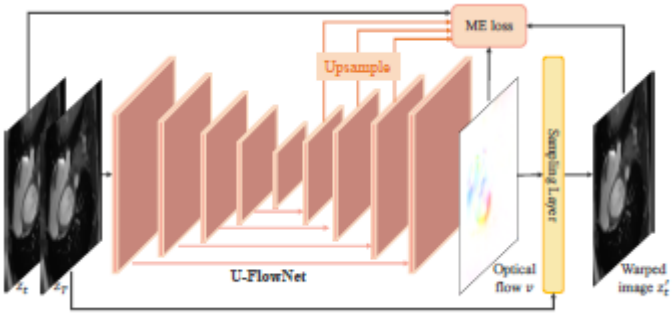


Fig. 2. Motion estimation (ME) component for unsupervised optical flow learning. It takes as input two frames to learn the multi-scale optical flow and is optimized by the ME loss.

4. Methodology

In this section, we introduce the details of our proposed method called Motion-guided Dynamic Reconstruction Network (MODRN), as well as an end-to-end version called MODRN(e2e).

4.1. Motion-guided Dynamic Reconstruction Network

Our method dissects the motion-guided dynamic reconstruction problem into three closely-connected parts: (i) Dynamic Reconstruction Network (DRN) for estimating initial reconstructed image from Eq. (2), (ii) Motion Estimation (ME) component for generating motion information through Eq. (5), and (iii) Motion Compensation (MC) component for refining the initial reconstructed image in Eq. (2) by the learned motion in Eq. (5).

4.1.1. Dynamic Reconstruction

Instead of directly solving Eq. (2), we adopt an iterative process through Dynamic Reconstruction Network (DRN) to approximate reconstructed images $\{z_t\}$. Formally, given under-sampled k-space data $\{y_t\}_{t\in[\tau]}$ with sampled mask Ω, DRN learns to reconstruct images in a total number of N iterations: $z_t^{(n)} = \mathrm{DRN}\left(z_t^{(n-1)}, y_t, \Omega\,; \Theta_{\mathrm{rec}}\right), n = 1, 2, \ldots, N$, where Θ_{rec} denotes the learnable parameters of DRN. Given the t^{th} frame, in the n^{th} iteration, the workflow of DRN is:

$$\begin{cases} x_{bt}^{(n)}, b_t^{(n)} = B\left(f_{enc}\left(z_t^{(n-1)}\right), b_t^{(n-1)}\right), \\ x_{st}^{(n)}, s_t^{(n)} = S\left(f_{dec1}\left(x_{bt}^{(n)}\right), s_t^{(n-1)}\right), \\ z_t^{(n)} = DC\left(f_{dec2}(x_{st}^{(n)}) + z_t^{(n-1)}, y_t, \Omega\right), \end{cases} \tag{7}$$

where $z_t^{(0)}$ is a zero-filling image and $z_t^{(n)}$ is the reconstructed image of y_t after the n-th iteration. B and S are two ConvGRU units (Ballas et al., 2015) that respectively output features $x_{bt}^{(n)}$ and $x_{st}^{(n)}$ together with hidden states $b_t^{(n)}$ and $s_t^{(n)}$. f_{enc} and f_{dec1}, f_{dec2} are convolutional encoder and decoders in the U-Net (Ronneberger et al., 2015), which is used as the backbone of the DRN to capture course-to-fine features of reconstructed images. Eq. (7) is visualized in Fig. 1 for better understanding. One benefit here is that regularization function $R(\cdot)$ in Eq. (2) is now built upon the convolutional network for automated feature learning and hence avoid the requirements of prior knowledge on the selection of R. $DC(\cdot)$ is the differentiable DC layer (Schlemper et al., 2017) that takes the same effect as the data consistency term $\|F_u(z_t) - y_t\|_2^2$ in Eq. (2) to force the reconstructed image to be consistent with the input data. It fills the reconstructed image z_t with the original values of input data y_t in the Fourier domain by the sampled mask Ω.

More importantly, in order to capture dynamic information of image sequence during each iteration, we introduce two kinds of ConvGRU units, *i.e.* B and S, inspired by the work (Qin et al., 2018b)). B considers both temporal and iteration information, while S only includes iteration information. Specifically, both B and S are used to improve the quality of image z_t over N iterations, as Eq. (7) shown. Besides, B additionally captures temporal information between neighboring frames z_{t-1} and z_t, which is implemented by $x_{bt}^{(1)}, b_t^{(1)} = B\left(f_{enc}\left(z_t^{(0)}\right), b_{t-1}^{(N)}\right)$, where hidden state $b_t^{(0)}$ is initialized as $b_{t-1}^{(N)}$, as the lower part of Fig. 1 shown. Finally, we impose l_1 loss on the reconstructed images $\{z_t^{(N)}\}$ with respect to the ground truth $\{z_t\}$ for penalizing, i.e. $\ell_{\mathrm{rec}}(\Theta_{\mathrm{rec}}) = \sum_{t=1}^{\tau} \left\|z_t^{(N)} - z_t\right\|_1$.

4.1.2. Motion Estimation

In analogy to Eq. (5), the Motion Estimation (ME) component takes as input the sequence of reconstructed images $\{z_t\}_{t\in[\tau]}$ and learns to predict displacement motion fields as well as the corresponding warped images. As shown in Fig. 2, our proposed ME component embraces two parts. One is called U-FlowNet for motion field estimation, which has a UNet structure (Ronneberger et al., 2015). The other is a differentiable sampling layer based on Spatial Transformer Network (STN) (Jaderberg et al., 2015), which endows convolutional network with the ability to warp the spatial deformation between images.

Traditional optimization algorithms for motion estimation depend on a strong assumption that the brightness of two frames should be consistent and the movement of the foreground object is small, while our method does not succumb to any assumption and hence is more applicable in practical dynamic MRI reconstruction.

The performance of ME is heavily affected by noisy input, therefore ME can be pre-trained with any two fully sampled images, without loss of generality, z_i and z_j $(1 \le i \le j \le \tau)$. The image pair is first fed to the U-FlowNet, which produces two-channel displacement v_{ij} along the x and y directions. Then, the sampling layer warps z_i towards z_j using v_{ij} and yields a warped image denoted by z'_j through differentiable bilinear interpolation. That is $z'_j, v_{ij} = \mathrm{ME}(z_i, z_j; \Theta_{\mathrm{me}})$, where Θ_{me} denotes the learnable parameters of ME component. This leads to a natural definition of motion estimation loss ℓ_{me} between z_i and z_j:

$$\ell_{\mathrm{me}}(\Theta_{\mathrm{me}}; z_i, z_j) = \left\| z'_j - z_j \right\|_1 + \beta \left\| v_{ij} \right\|_1 + \gamma \left\| v_{ij} \right\|_{TV}. \quad (8)$$

The first term is an image reconstruction loss used to keep the majority of high-frequency parts on images. Two additional regularization terms reinforce constraints on the motion field v_{ij}, where l_1 regularization is to suppress unreal large magnitude of displacement and total-variation (TV) regularization is to make the displacement locally smooth.

It is natural to pretrain ME using every two neighboring images z_t, z_{t+1} with loss $\ell_{me}(\Theta_{\mathrm{me}}; z_t, z_{t+1})$. However, the loss only enforces temporal consistency between consecutive frames, but there is no guarantee for long-term coherence. Therefore, we consider to train it with three sets of ME losses to capture long-term motion information. In particular, let $T \in [2, \tau - 1]$ be a self-defined period and z_1 be the start frame of the period and z_T be the end frame, τ is the sequence length. For example, if $T = \tau/2$ and z_1 is End-Diastole frame, a period represents a systole phase, and if $T = \tau$ and z_1 is End-Diastole frame, it means a whole cardiac cycle. In order to capture all motion pattern within this period, we define the combined ME loss as follows.

$$\begin{aligned} \ell_{me}(\Theta_{\mathrm{me}}; \{z_t\}) = & \sum_{t=2}^{T-1} \ell_{me}(\Theta_{\mathrm{me}}; z_1, z_t) + \sum_{t=2}^{T-1} \ell_{me}(\Theta_{\mathrm{me}}; z_t, z_T) \\ & + \sum_{t=2}^{T-2} \ell_{me}(\Theta_{\mathrm{me}}; z_t, z_{t+1}), \end{aligned} \quad (9)$$

where three terms on the right-hand-side are respectively called forward ME loss, backward ME loss and neighboring ME loss as illustrated in Fig. 3. Note that within the systole or diastole period, a larger T tends to capture longer term motion information while the downside is that it becomes harder for the neural network to model very long sequential information.

4.1.3. Motion Compensation

Motion Compensation (MC) component is used to refine reconstructed images $\{z_t\}_{t\in[\tau]}$ through motion information and to generate motion compensated image $\{c_t\}_{t\in[\tau]}$. Inspired by the work (Jung et al., 2010), in the MC stage, we also add k additional fully-sampled reference frames at positions $0, T, 2T, \ldots, kT$ $(kT \le \tau)$ to learn more accurate displacement

Fig. 3. Motion Estimation (ME) loss includes forward, backward and neighboring losses that incorporate long-term motion information.

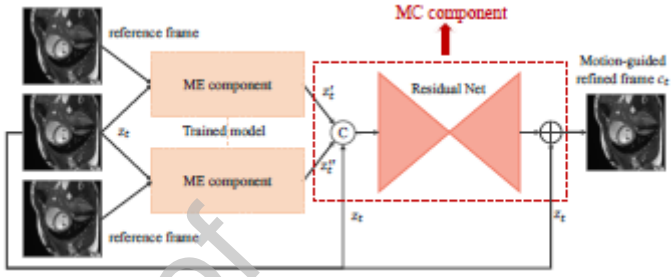


Fig. 4. Illustration of MC component that takes use of a reconstructed image z_t and two corresponding warped frames z'_t and z''_t from the ME component to generate a motion-guided refined image c_t.

motion fields, where T is the aforementioned self-defined period. As shown in Fig. 4, the MC component takes as input a reconstructed image z_t together with two warped images z'_t and z''_t and outputs the corresponding motion-guided refined image c_t. The warped images are generated from the reconstructed image z_t with two closest reference frames through the ME component. z'_t and z''_t respectively represent forward and backward motion information, which are then concatenated and fed to a residual network to generate residual information. Finally, the reconstructed image z_t is summed up with the residual information to generate the motion-guided refined image c_t. The process of MC component can be formalized as $c_t = \mathrm{MC}(\Theta_{\mathrm{mc}}; z_t, z'_t, z''_t)$, where Θ_{mc} denotes learnable parameter and is optimized via l_1 loss with respect to the ground truth image, i.e. $\ell(\Theta_{\mathrm{mc}}) = \sum_{t=1}^{\tau} \|c_t - z_t\|_1$.

4.1.4. MODRN Training Procedure

The training procedure of MODRN is as follows. First, we pretrain the ME component with fully-sampled images using the loss in Eq. 9 to acquire good prediction ability on the motion between different frames. Second, we train the DRN individually to get initial reconstructed images. Third, we train the MC component with the initial reconstructed images from DRN and the corresponding warped images from ME. The final reconstructed images can then be generated from the MC component. Note that the above three training steps are executed separately.

4.2. End-to-end MODRN

To further improve the performance of the MODRN, we also propose an end-to-end model called MODRN(e2e) by combining DRN, ME and MC into a whole differentiable network. As illustrated in Fig. 5, zero-filling image sequence (the Fourier transform of under-sampled k-space sequence y_t) is first fed to

DRN	U-FlowNet	Residual Net
Conv-7-32	Conv-3-64	Conv-3-64
Conv-3-64	Conv-3- 128	Conv-3-128
Conv-3-128	Conv-3-256	Conv-3-256
ConvGru-3-128	Conv-3-512	Conv-3-512
Deconv-3-64	Conv-3-1024	Deconv-3-256
ConvGru-3-128	Deconv-3-512	Deconv-3-128
Deconv-7-2	Deconv-3-256	Deconv-3-64
	Deconv-3-128	Conv-3-2
	Deconv-3-64	
	Conv-3-2	

Table 1. Detailed Network Architectures. All networks are based on UNet structure, which have skip connections to connect the corresponding layers of encoder and decoder. The operation in each grid represents *[operation]-[kernel size]-[features channel number].*

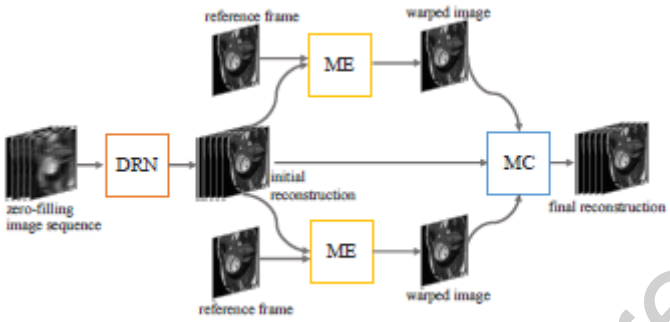


Fig. 5. Illustration of the end-to-end MODRN model that combines DRN, ME and MC into a single differentiable network.

DRN to obtain initial reconstructed images. Then, each reconstructed image together with two reference frames are passed to ME component to generate warped images, which are later used to produce final reconstructed images through MC component. Training in such end-to-end manner is beneficial for allowing gradients from MC component to be backpropagated to ME components and DRN. In other words, these two components are not only optimized for motion estimation and initial reconstruction, but for directly moving towards generating final motion-guided frames.

5. Model Analysis

In this section, we conduct ablation studies of our models as follows. We first train the ME with various losses and compare it with traditional motion estimation methods in section 5.1; Then, we experiment with the influence of hyper-parameters of MODRN in Section 5.2; Finally, we evaluate MODRN and its variants to demonstrate the effectiveness of using motion information in section 5.3.

Evaluation Dataset: We experiment with a short-axis (SAX) cardiac dataset composed of 21 normal people and 3 dyssynchrony disease patients. Each subject contains around 12 SAX planes and each plane includes 24 phases (2D images) that form a whole cardiac cycle. The myocardium of each subject is manually labeled by two Cardiologists. The accuracy of the annotation is validated by another expert. The image resolution is normalized to $1.25mm$ and image size is cropped to 152×152 pixels. In order to simulate k-space data, we first use Fourier transform to transform the magnitude images into the Fourier domain to get complex-valued k-space data. Then we adopt the same Cartesian under-sampling method as introduced in (Jung et al., 2007), which assumes that sampled mask Ω follows a zero-mean Gaussian distribution and keeps 8 center spatial frequencies. We consider two different settings on the dataset respectively with under-sampling rates of 20% (or acceleration rate 5×) and 12.5% (8×). For convenience, we refer to these two cases as *Rate* 5× and *Rate* 8×. We randomly sample 13 normal person and 1 patient as training data, 4 normal person and 1 patient as validation data and the rest as test data.

γ \ β	0.01		0.1		1		10	
	Dice↑	HD↓	Dice↑	HD↓	Dice↑	HD↓	Dice↑	HD↓
0.5	0.8637	1.8748	0.8641	1.8776	0.8668	1.8011	0.8698	1.8877
5	0.8783	1.9016	0.8779	1.8894	0.8779	1.8608	**0.8825**	**1.8021**
50	0.8485	1.8922	0.8485	1.8922	0.8510	1.9189	0.8492	1.8970

Table 2. Results of grid search on validation set with different γ and β. Dice and HD scores are calculated on all frames and on the myocardium region.

Method	Dice↑	HD↓
Reference	0.8155	2.2030
Lucas-Kanade	0.8165	2.1215
ME-neighbor	0.8192	2.0838
ME-combined	**0.8230**	**1.9820**

Table 3. Motion estimation results on test set compare to other approaches. Dice and HD scores are calculated on all time frames and on the myocardium region.

Implementation Details:

The intensity of all training data are normalized between 0 and 1. The detailed architecture of each component are listed in Table 1. We implement all the deep learning models with PyTorch and train them on NVIDIA 1080Ti. All models are trained under the same setting: Adam optimizer is adopted with an initialized learning rate of 5×10^{-4} and decreasing rate of 0.5 for every 20 epochs. Batch size is set to be 1. All models are trained with 60 epochs. Note that our MODRN model is trained in a pipelined framework, as described in Section 4.1.4, while our MODRN(e2e) model is trained from scatch in an end-to-end manner.

Evaluation Metrics: We report the Normalized Root Mean Square Error (NRMSE) reconstruction, Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) scores on the magnitude images as performance metrics.

5.1. Study of ME Loss

In this section, we study the ME of MODRN that is individually pretrained with fully-sampled images. For the combined ME loss defined in Eq. 9, we fix $T = 12$, which means we calculate the motion between the End-Distole (ED) frame and the End-Systole (ES) frame. First, we tune the hyper-parameters γ and β in Eq. 8. We run a grid search on the validation set to find the optimal γ and β in terms of Dice's score and Hausdorff Distance (HD). The results are reported in Table 2, which is the

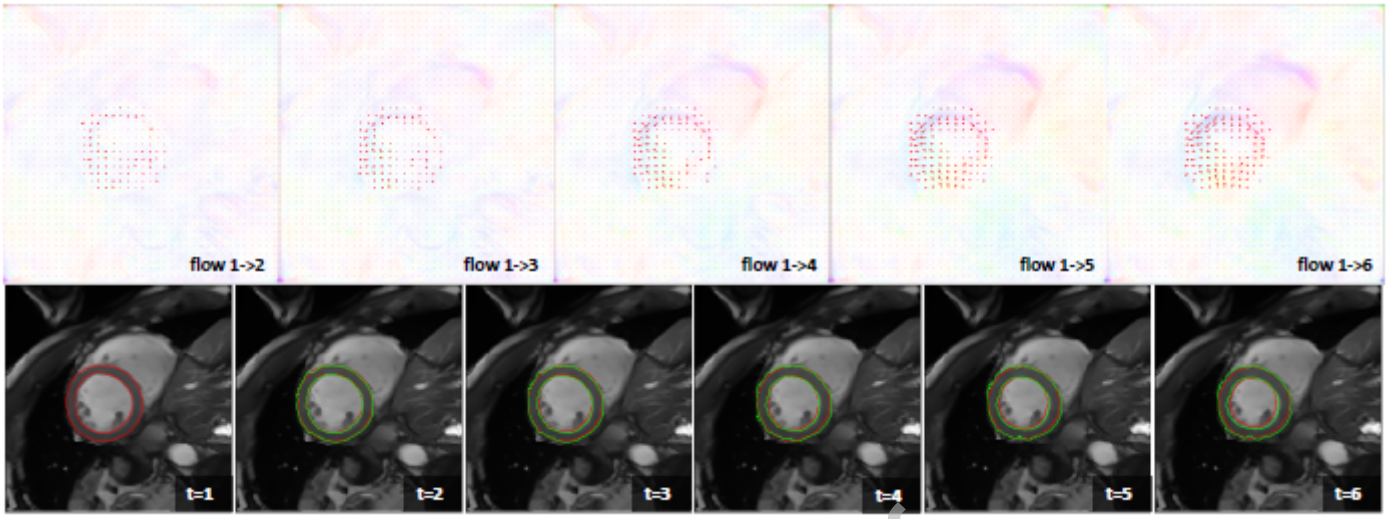


Fig. 6. The first row represents optical flow between image $t = 1$ and image $t = T$. The arrows denotes large displacement and direction (all displacement are multiplied with a constant for visualization). The second row represents warped image and warped myocardium contour from image $t = 1$ using flow $1 \rightarrow T$. The red contour is the ground truth myocardium and the green contour is the warped myocardium.

average score of forward, backward and neighboring motion. We find that the best performance of Dice and HD is achieved when $\gamma = 5$ and $\beta = 10$. We adopt this combination of γ and β in the following experiments.

Then, we evaluate the effectiveness of our combined ME loss in terms of the motion estimation results compared to other motion estimation approaches. Two baseline methods are considered here. *Reference* method directly calculates metrics using the segmentation of the target phase and the reference phase. Since it is impractical to obtain the ground truth of optical flow from cardiac MRI, we compute the overlapped area of the myocardium between the targeting image and the warped image. The other baseline is *Lucas-Kanade* (Lucas et al., 1981), which is a widely used optical flow estimation method. For our ME loss, we also include two variants. *ME-neighbor* is the ME network trained with only neighboring loss (last term in Eq. 9), and *ME-combined* is the one trained with the combined ME loss defined in Eq. 9. The average results are reported in Table 3. We observe that our ME-combined achieves the best Dice of 0.8230 and HD of 1.9820 among all methods. It indicates that by using the ME-combined loss that the ME captures more accurate motion information.

In addition, we qualitatively study the learned motion from ME by visualizing the results from the best performed ME-combined loss. Taking the first six images in the systole phase, the row one of Fig. 6 represents the forward optical flow. The flow marked by $1 \rightarrow t$ ($t \in [2, 6]$) means the flow is computed from the 1st image to the t-th image. The second row represents the warped image with the corresponding warped myocardium contour from image 1 the t-th image by using the flow $1 \rightarrow t$. The motion estimation is challenging, we observe that the performance degrades when the gap increases between the reference image and the target image, but our model still performs well even with a long length.

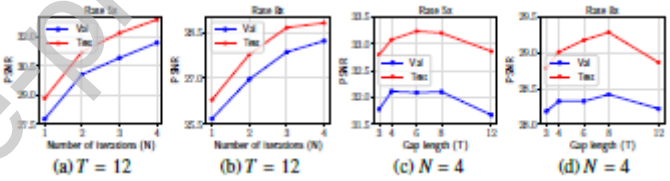


Fig. 7. Average PSNR (dB) values vary with the number of iterations N and length T on both validation and test data with $5\times$ and $8\times$ acceleration factors. (a) and (b) are the results of $T = 12, N = 1, 2, 3, 4$. (c) and (d) are $T = 3, 4, 6, 8, 12$ and $N = 4$.

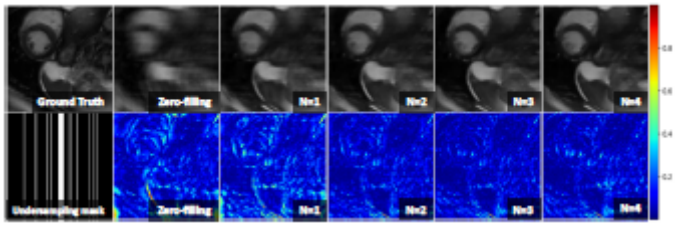


Fig. 8. Visualization of reconstructed images and errors with the number of iterations ($N = 1, 2, 3, 4$) in the case of Rate $8\times$.

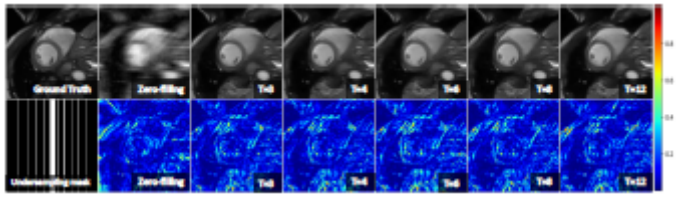


Fig. 9. Visualization of reconstructed images and errors with different length $T = 3, 4, 6, 8, 12$ in the case of Rate $8\times$.

5.2. *Hyper-parameters of MODRN*

In this section, we fixed the pretrained ME component with the optimal β and γ, and study the hyper-parameters of

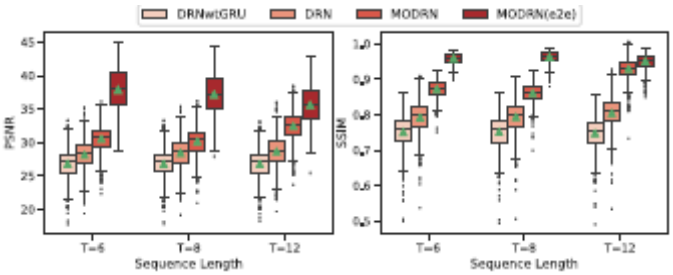


Fig. 10. Boxplots of PSNR and SSIM values for models trained on *Rate* 8× data. Hues indicate the model DRNwtGRU, DRN, MODRN and MODRN(e2e). The green triangle is the mean value.

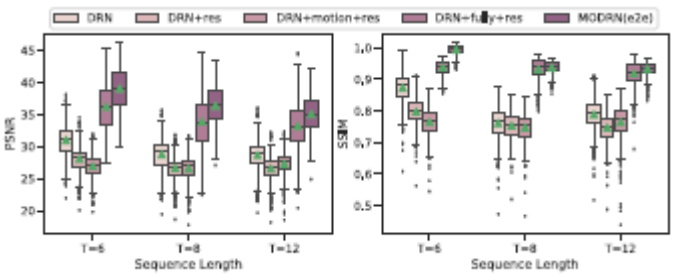


Fig. 11. Boxplots of PSNR and SSIM values for models trained on *Rate* 8× data. Hues indicate the model DRN, DRN+residual, DRN+motion+residual, DRN+fully+residual and MODRN(e2e). The green triangle is the mean value.

MODRN: the number of iterations N in DRN and the gap length of every two reference frames T in MC. We design two experiments to evaluate how these two variables affect the performance. Due to hardware limitation, the largest number of iterations is set to be $N = 4$ and the longest length is set to be $T = 12$.

In the first experiment, we train the DRN with fixed $T = 12$, which means two reference frames are at position of ED frame and ES frame. We report PSNR of validation set (Val) and test set (Test) with different values of $N = 1, 2, 3, 4$ in the two cases of Rate 5× and Rate 8×. As shown in Fig. 7 (a) and (b), the performance improves significantly as N becomes larger. We also visualize the reconstructed images and errors in Fig. 8. It qualitatively shows that the reconstruction error decreases when N becomes higher.

Next, we fix $N = 4$ and take the different values of $T = 3, 4, 6, 8, 12$, the divisors of the whole sequence length 24. We report both cases of Rate 5× and 8× in Fig. 7 (c) and (d). When T is small, the improvement should be large due to the similarity between the reconstructed image and the reference image. However, the exception occurs in cases $T = 3$ and $T = 4$. One reason could be that the model is overfitting in these cases. The model iterates 4 times ($N = 4$) and more reference images are provided to make it easier to learn to fit the training set, while it fails to fit additional data reliably. We observe that the best performance is achieved when $T = 8$ in both cases and the performance decreases if T becomes larger. This is when T becomes lager, it's challenging to capture accurate motion information. We also visualize the reconstructed images and errors in Fig. 9, which shows that the quality of reconstruction images become comparably good when $T \geq 6$.

5.3. Ablation Study of MODRN(e2e)

To fully evaluate the proposed method, in this section, we perform a 3-fold cross-validation in the following experiments. In each fold, the training data consists of 14 normal cases and 2 abnormal case, and the remaining cases are used as test data. In all, 24 cases have been cross validated.

5.3.1. Effect of End-to-end Training

We first show the importance of incorporating motion information into the reconstruction process and training with end-to-end paradigm for the dynamic reconstruction problem. Specifically, we consider four variants of our models: *DRN* (DRN alone in Section 4.1.1), *DRNwtGRU* (DRN without GRU hidden unit), *MODRN* (regular version in 4.1.4) and *MODRN(e2e)* (end-to-end version in Section 4.2). The results of PSNR and SSIM in the case of Rate 8× are reported in Fig. 10. We observe two interesting trends. First, models with explicit motion modeling parts (MODRN and MODRN(e2e)) achieve better performance than the ones without motion (DRN and DRNwtGRU). This boost of performance mainly attributes to the motion information exploited by ME/MC components. We also find that DRN outperforms DRNwtGRU, which indicates the importance of utilizing the dynamic sequence of image. Second, end-to-end training paradigm by MODRN(e2e) makes further improvement. One reason is that motion information captured by ME/MC can be directly backpropagated to DRN so that it can learn to make good predictions towards final motion-guided reconstructions.

We further qualitatively evaluate the results of motion estimation by MODRN and MODRN(e2e). As shown in Fig. 12, we present the forward flows of the first eight images of the diastole phase. We adopt the same color scheme as (Ilg et al., 2017; Meister et al., 2017) in the first and third rows. Since MODRN(e2e) learns more precise motion than MODRN, the plotted motion looks different. Our goal is to predict the motion from the 1^{st} frame to the $T^{th}(T = 2 \cdots 8)$ frame. Both MODRN and MODRN(e2e) are fed as input with the same 1^{st} frame and the 2^{nd} frame slightly differs from the 1^{st} one, so the initial motion between $t = 1$ and $t = 2$ is easy to learn. That is why both MODRN and MODRN(e2e) perform well in the case of $t = 2$, as indicated in the second and the fourth row where the green contour (the warped contour learned by the corresponding motion) is almost overlapped with the red contour (the ground truth contour). However, when t becomes larger, it is harder to predict precise motion, because the t^{th} frame varies a lot with respect to the 1^{th} frame, and the accumulative prediction error at each step may make the model less accurate. Fig. 12 shows that with the help of the end-to-end training, our MODRN(e2e) is still able to capture motion information well, especially that in the regions of myocardium contours pointed by the white arrows in the last three frames. We also observe that the motion of MODRN(e2e) model are smoother than the MODRNmodel, which validates the effectiveness of the end-to-end training paradigm. One explanation of the success of MODRN(e2e) is that the gradients can be backprop-

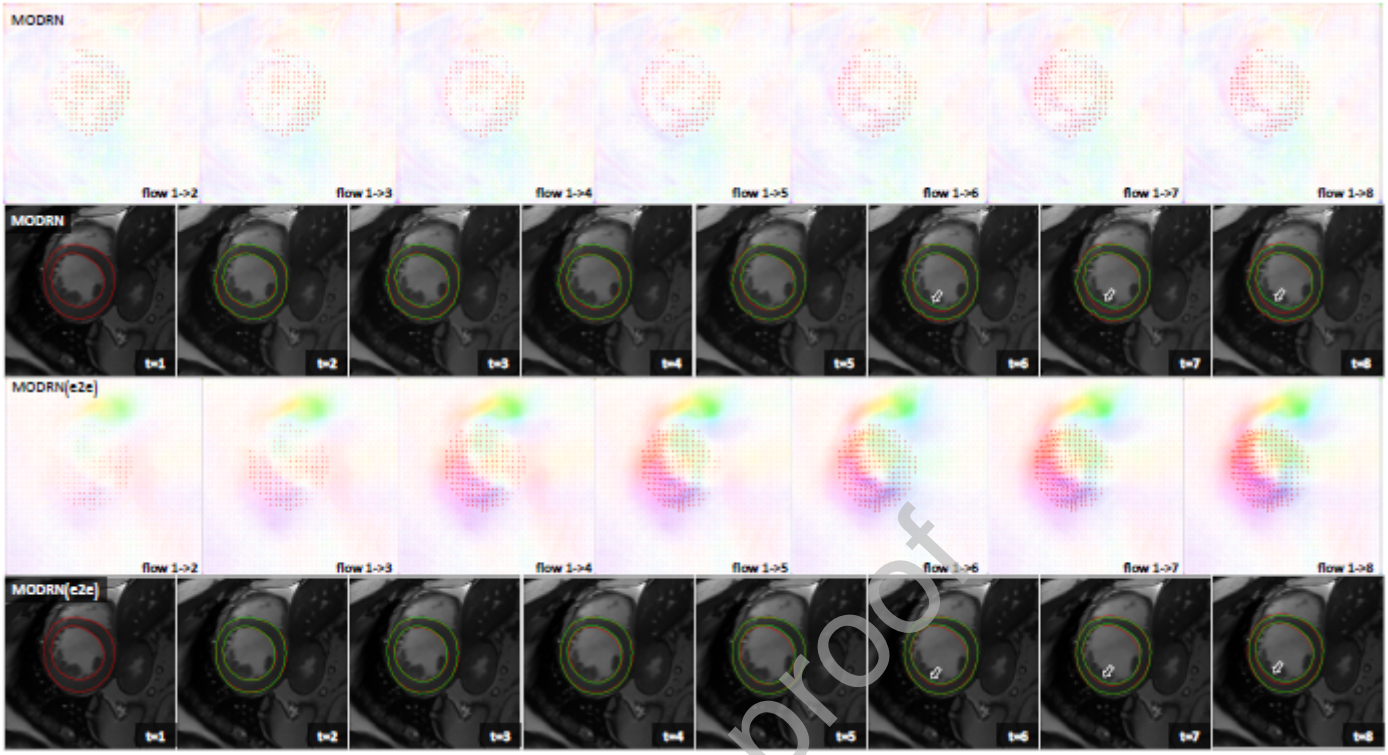


Fig. 12. The first and the third rows represent optical flow between image $t = 1$ and image T learned from MODRN and MODRN(e2e) model. The red arrows denote large displacement and direction (all displacement are multiplied with a constant for visualization). The second and forth rows represent warped image and warped myocardium contour from image $t = 1$ using flow $1 \rightarrow T$. The red contour is the ground truth myocardium and the green contour is the warped myocardium. The contours highlighted by white arrows in image $t = 6, 7, 8$ are MODRN(e2e) performs better than MODRN.

agated through the whole network to avoid optimization goes into a local minimum. While for the MODRN that is trained in a pipelined framework, for which the current component takes input as the output of the previous component, easily leading to errors accumulation.

5.3.2. *Effect of Components of MODRN(e2e)*

In the last experiment, we evaluate the significance of each component in MODRN(e2e). Specifically, we compare the *MODRN(e2e)* with following variants: *DRN* network (Fig. 1), *DRN+res* (DRN network plus residual network that is a part of MC of Fig. 4), *DRN+motion+res* (DRN network plus ME and residual network), *DRN+fully+res* (DRN network plus fully-sampled reference frames and residual network). To clarify, the MODRN(e2e) consists of all the components: DRN, ME, MC and fully-sampled reference frames. As we can see the results presented in Fig 11, models with part of components solely cannot outperform the MODRN(e2e) approach. It shows that each component of our model contributes to the performance of dynamic reconstruction under different settings.

For example, *DRN+fully+res* significantly outperforms the models without incorporating fully-sampled reference frames. The performance gain is due to the similarity between the reconstructed image and the reference image. However, the benefit of fully-sampled references is more than this. They can be leveraged to learn better motion fields via the motion estimation components to further boost reconstruction performance, as MODRN(e2e) achieves even better performance than *DRN+fully+res*. Since as T increases, the similarity between the reconstructed image and the reference image will decrease, thus MODRN(e2e) utilizes motion to warp the reference frame to the corresponding reconstructed image to reduce the dissimilarity. Last but not least, the introduction of fully-sampled references can largely lead to performance gain, and it is feasible to acquire them in practical scenarios. For example, we can set a relatively large T and acquire two fully-sampled frames at position 0 and T, corresponding to ED and ES phases and then use motion information to overcome the long-term gap between the fully-sampled frames and undersampled frames. The idea of reference frames have been commonly used in dynamic MRI Jung and Ye (2010). For instance, one dynamic imaging method called RIGR (Reduced-encoding Imaging by Generalized-series Reconstruction) Liang and Lauterbur (1994) also employs two fully encoded reference frames. It is easy to obtain fully sampled reference frames in many practical dynamic MR applications, such as fMRI and perfusion imaging.

Another finding is that the performance of *DRN+res* and *DRN+motion+res* is worse than that of *DRN*. For *DRN*, *DRN+res* and *DRN+motion+res*, none of them is guided by the fully-sampled reference frames. The estimated motion or residual tends to be easily influenced by the artifacts and noise in the reconstructed images. Therefore, the "contaminated" motion or residual learned from additional components will worsen the reconstruction quality instead.

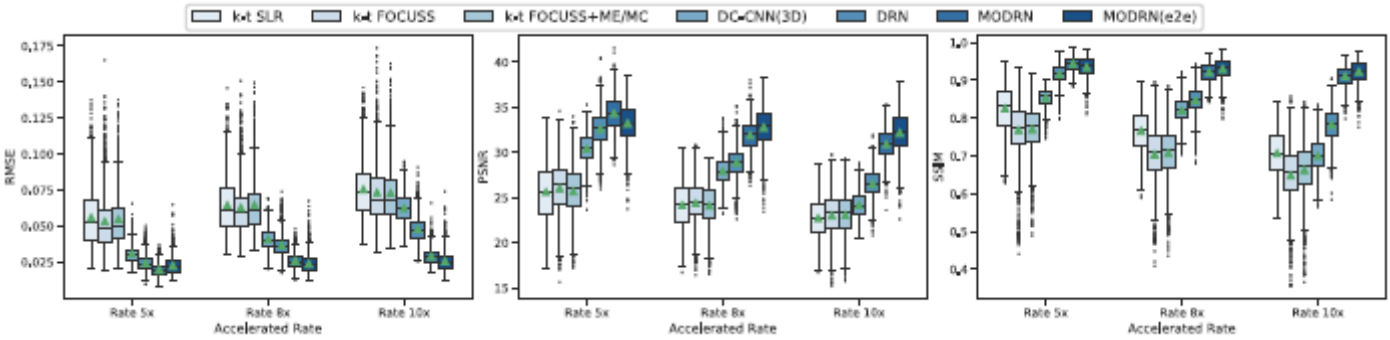
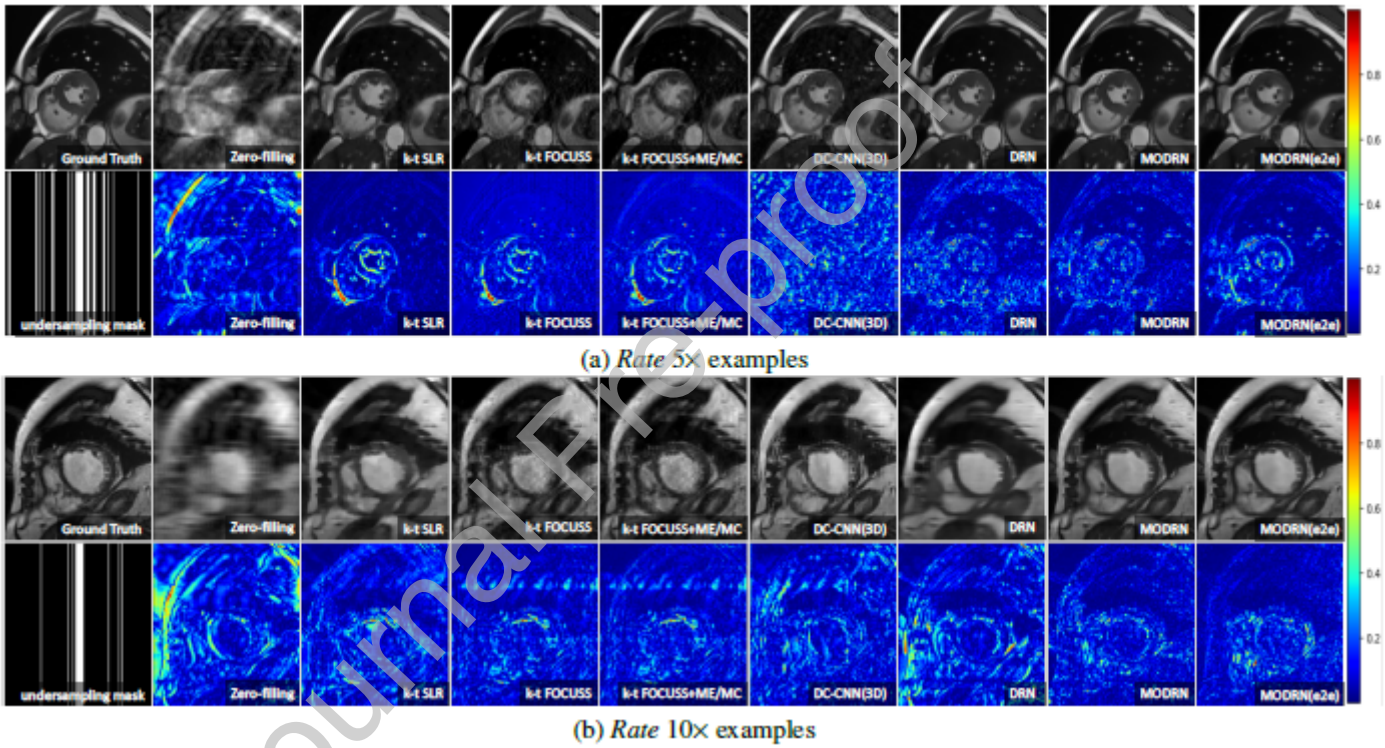


Fig. 13. Boxplots of NRMSE, PSNR and SSIM values for models trained on *Rate* 5×, *Rate* 8× and *Rate* 10× data. Hues indicate the model k-t SLR, k-t FOCUSS, k-t FOCUSS+ME/MC, DC-CNN(3D), MODRN and MODRN(e2e). The green triangle is the mean value.



(a) *Rate* 5× examples

(b) *Rate* 10× examples

Fig. 14. Visualization of reconstructed images and errors.

6. Application to another dataset and Comparisons to State-of-the-Art

To validate the generalizability of our approach, we evaluate our model on a public dataset: ACDC dataset [1](Bernard et al., 2018). It covers several well-defined pathologies: (1) myocardial infarction - MINF (2) dilated cardiomyopathy - DCM (3) hypertrophic cardiomyopathy - HCM (4) abnormal right ventricle - ARV. The dataset composes of 100 subjects (all from different patients) that are divided into 5 evenly distributed groups (4 pathological groups and 1 healthy subject group). The benefit of the ACDC dataset is that it provides rich motion patterns, which is suitable to assess our motion-guided model. We use a subset of the ACDC dataset containing 9600 2D slices. The image resolution is normalized to 1.25*mm* and the image size is cropped to 144 × 144 pixels. Since there is no public available dataset for dynamic MRI reconstruction, we adopt the same *k*-space data simulation method as the dataset in Section 5. We first transform the magnitude images into the Fourier domain to get the complex-valued *k*-space data. Then we under sampled the complex-valued data in the Fourier domain to get the undersampled *k*-space data. Three different settings are considered with respect to different acceleration rates of 5×, 8× and 10×. We set the largest number of iterations $N = 3$ and length $T = 10$ due to the GPU limitations. We apply the same γ and β according to previous experiment. We split the dataset into training (60 subjects), validation (20 subjects) and test (20 subjects) set and ensure each of them contains both pathologies and healthy cases.

[1]https://www.creatis.insa-lyon.fr/Challenge/acdc/

We consider three variants of our models: *DRN*, *MODRN* and *MODRN(e2e)*, as well as four state-of-the-art baselines including *k-t SLR* (Lingala et al., 2011), *k-t FOCUSS* (Jung et al., 2009), *k-t FOCUSS+ME/MC* (Jung et al., 2010) and *DC-CNN (3D)* (Schlemper et al., 2018). The first three baselines are traditional CS-based methods and only k-t FOCUSS+ME/MC includes ME/MC procedures. The last one is also a deep learning-based method that explores spatio-temporal information using 3D convolution and data sharing layer. The metrics to evaluate all approaches include NRMSE, PSNR and SSIM.

The results of all methods are reported in Fig. 13. We observe that our three methods DRN, MODRN and MODRN(e2e) consistently outperform four baselines in all *Rate* 5×, *Rate* 8× and *Rate* 10× cases. The CS-based methods have larger variance compared with deep learning-based methods since it is hard to find a heuristic parameter setting that achieves best results in all the cases. In particular, MODRN(e2e) achieves the best performance in terms of all metrics and it mainly attributes to the motion information exploited by ME/MC components and end-to-end training. In contrast, the performance of the models without ME/MC are largely influenced by high accelerated rate data. For example, in the case of *Rate* 8× and *Rate* 10×, the gap of all three metrics between DRN and MODRN model is large, implying the significance of using motion information.

We also visualize the reconstructed images and error with respect to ground truth of all methods in Fig. 14. It is obvious that all CS-based methods have streaking artifacts and larger reconstruction error while our MODRN models eliminate the most blurring artifacts and recover more high-frequency details.

7. Conclusion

We present a novel deep learning-based approach called MODRN as well as an end-to-end version named MODRN(e2e) for motion-guided dynamic MRI reconstruction problem. It is featured by a Dynamic Reconstruction Network (DRN) for preliminary image reconstruction from under-sampled *k*-space data and a motion estimation (ME) component to predict the motion of image sequence, which is further exploited by a motion compensation (MC) component to refine the motion-guided reconstructed images. We extensively evaluate our approach on two dynamic cardiac datasets in different settings. The experimental results show the effectiveness of MODRN compared to state-of-the-art methods and prove the significance of motion information from ME/MC components.

References

Ballas, N., Yao, L., Pal, C., Courville, A., 2015. Delving deeper into convolutional networks for learning video representations. arXiv preprint arXiv:1511.06432 .

Bao, W., Lai, W.S., Zhang, X., Gao, Z., Yang, M.H., 2018. Memc-net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement. arXiv preprint arXiv:1810.08768 .

Bernard, O., Lalande, A., Zotti, C., Cervenansky, F., Yang, X., Heng, P.A., Cetin, I., Lekadir, K., Camara, O., Ballester, M.A.G., et al., 2018. Deep learning techniques for automatic mri cardiac multi-structures segmentation and diagnosis: Is the problem solved? IEEE transactions on medical imaging 37, 2514–2525.

Caballero, J., Ledig, C., Aitken, A., Acosta, A., Totz, J., Wang, Z., Shi, W., 2017. Real-time video super-resolution with spatio-temporal networks and motion compensation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4778–4787.

Cruz, G., Atkinson, D., Buerger, C., Schaeffter, T., Prieto, C., 2016. Accelerated motion corrected three-dimensional abdominal mri using total variation regularized sense reconstruction. Magnetic resonance in medicine 75, 1484–1498.

Hammernik, K., Klatzer, T., Kobler, E., Recht, M.P., Sodickson, D.K., Pock, T., Knoll, F., 2018. Learning a variational network for reconstruction of accelerated mri data. Magnetic resonance in medicine 79, 3055–3071.

Horn, B.K., Schunck, B.G., 1981. Determining optical flow. Artificial intelligence 17, 185–203.

Hu, J., Shen, L., Sun, G., 2018. Squeeze-and-excitation networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 7132–7141.

Huang, Q., Yang, D., Qu, H., Yi, J., Wu, P., Metaxas, D., 2019a. Dynamic mri reconstruction with motion-guided network, in: Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning, PMLR, London, United Kingdom. pp. 275–284.

Huang, Q., Yang, D., Wu, P., Qu, H., Yi, J., Metaxas, D., 2019b. Mri reconstruction via cascaded channel-wise attention network, in: 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), IEEE. pp. 1622–1626.

Huang, Q., Yang, D., Yi, J., Axel, L., Metaxas, D., 2019c. Fr-net: Joint reconstruction and segmentation in compressed sensing cardiac mri, in: International Conference on Functional Imaging and Modeling of the Heart, Springer. pp. 352–360.

Hyun, C.M., Kim, H.P., Lee, S.M., Lee, S., Seo, J.K., 2018. Deep learning for undersampled mri reconstruction. Physics in Medicine & Biology 63, 135007.

Ilg, E., Mayer, N., Saikia, T., Keuper, M., Dosovitskiy, A., Brox, T., 2017. Flownet 2.0: Evolution of optical flow estimation with deep networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 2462–2470.

Jaderberg, M., Simonyan, K., Zisserman, A., et al., 2015. Spatial transformer networks, in: Advances in neural information processing systems, pp. 2017–2025.

Jin, M., Hu, Z., Favaro, P., 2019. Learning to extract flawless slow motion from blurry videos, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 8112–8121.

Jung, H., Park, J., Yoo, J., Ye, J.C., 2010. Radial k-t focuss for high-resolution cardiac cine mri. Magnetic Resonance in Medicine 63, 68–78.

Jung, H., Sung, K., Nayak, K.S., Kim, E.Y., Ye, J.C., 2009. k-t focuss: a general compressed sensing framework for high resolution dynamic mri. Magnetic resonance in medicine 61, 103–116.

Jung, H., Ye, J.C., 2010. Motion estimated and compensated compressed sensing dynamic magnetic resonance imaging: What we can learn from video compression techniques. International Journal of Imaging Systems and Technology 20, 81–98.

Jung, H., Ye, J.C., Kim, E.Y., 2007. Improved k–t blast and k–t sense using focuss. Physics in Medicine & Biology 52, 3201.

Knoll, F., Clason, C., Bredies, K., Uecker, M., Stollberger, R., 2012. Parallel imaging with nonlinear reconstruction using variational penalties. Magnetic resonance in medicine 67, 34–41.

Liang, Z.P., Lauterbur, P.C., 1994. An efficient method for dynamic magnetic resonance imaging. IEEE Transactions on medical imaging 13, 677–686.

Lingala, S.G., Hu, Y., DiBella, E., Jacob, M., 2011. Accelerated dynamic mri exploiting sparsity and low-rank structure: kt slr. IEEE transactions on medical imaging 30, 1042–1054.

Lønning, K., Putzky, P., Caan, M.W., Welling, M., 2018. Recurrent inference machines for accelerated mri reconstruction .

Lucas, B.D., Kanade, T., et al., 1981. An iterative image registration technique with an application to stereo vision .

Makansi, O., Ilg, E., Brox, T., 2017. End-to-end learning of video super-resolution with motion compensation, in: German conference on pattern recognition, Springer. pp. 203–214.

Meister, S., Hur, J., Roth, S., 2017. Unflow: Unsupervised learning of optical flow with a bidirectional census loss. arXiv preprint arXiv:1711.07837 .

Qin, C., Bai, W., Schlemper, J., Petersen, S.E., Piechnik, S.K., Neubauer, S., Rueckert, D., 2018a. Joint learning of motion estimation and segmentation for cardiac mr image sequences, in: International Conference on Medical

Image Computing and Computer-Assisted Intervention, Springer. pp. 472–480.

Qin, C., Hajnal, J.V., Rueckert, D., Schlemper, J., Caballero, J., Price, A.N., 2018b. Convolutional recurrent neural networks for dynamic mr image reconstruction. IEEE transactions on medical imaging .

Quan, T.M., Nguyen-Duc, T., Jeong, W.K., 2018. Compressed sensing mri reconstruction using a generative adversarial network with a cyclic loss. IEEE transactions on medical imaging 37, 1488–1497.

Ren, Z., Yan, J., Ni, B., Liu, B., Yang, X., Zha, H., 2017. Unsupervised deep learning for optical flow estimation., in: AAAI, p. 7.

Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.

Schlemper, J., Caballero, J., Hajnal, J.V., Price, A., Rueckert, D., 2017. A deep cascade of convolutional neural networks for mr image reconstruction, in: International Conference on Information Processing in Medical Imaging, Springer. pp. 647–658.

Schlemper, J., Caballero, J., Hajnal, J.V., Price, A.N., Rueckert, D., 2018. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. IEEE transactions on Medical Imaging 37, 491–503.

Seegoolam, G., Schlemper, J., Qin, C., Price, A., Hajnal, J., Rueckert, D., 2019. Exploiting motion for deep learning reconstruction of extremely-undersampled dynamic mri, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 704–712.

Sun, J., Li, H., Xu, Z., et al., 2016. Deep admm-net for compressive sensing mri, in: Advances in Neural Information Processing Systems, pp. 10–18.

Tezcan, K.C., Baumgartner, C.F., Luechinger, R., Pruessmann, K.P., Konukoglu, E., 2018. Mr image reconstruction using deep density priors. IEEE transactions on medical imaging .

Trzasko, J., Manduca, A., Borisch, E., 2011. Local versus global low-rank promotion in dynamic mri series reconstruction, in: Proc. Int. Symp. Magn. Reson. Med, p. 4371.

Yang, A.C.Y., Kretzler, M., Sudarski, S., Gulani, V., Seiberlich, N., 2016. Sparse reconstruction techniques in mri: methods, applications, and challenges to clinical adoption. Investigative radiology 51, 349.

Yang, G., Yu, S., Dong, H., Slabaugh, G., Dragotti, P.L., Ye, X., Liu, F., Arridge, S., Keegan, J., Guo, Y., et al., 2018. Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction. IEEE transactions on medical imaging 37, 1310–1321.

Yi, J., Wu, P., Huang, Q., Qu, H., Hoeppner, D.J., Metaxas, D.N., 2019. Online neural cell tracking using blob-seed segmentation and optical flow, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 0–0.

Zhang, Z., Wang, Z., Lin, Z., Qi, H., 2019. Image super-resolution by neural texture transfer, in: arXiv:1903.00834v1.

Zheng, Q., Delingette, H., Ayache, N., 2019. Explainable cardiac pathology classification on cine mri with motion characterization by semi-supervised learning of apparent flow. Medical Image Analysis 56, 80 – 95.

Zhu, B., Liu, J.Z., Cauley, S.F., Rosen, B.R., Rosen, M.S., 2018. Image reconstruction by domain-transform manifold learning. Nature 555, 487.

Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Conflict of Interest Statement

I declare that there is no conflict of interest.