High-quality segmentation of low quality cardiac MR images using k-space artefact correction

Ilkay Oksuz[1] ILKAY.OKSUZ@KCL.AC.UK
James Clough[1] JAMES.CLOUGH@KCL.AC.UK
Wenjia Bai[2] W.BAI@IMPERIAL.AC.UK
Bram Ruijsink[1] JACOBUS.RUIJSINK@KCL.AC.UK
Esther Puyol-Antón[1] ESTHER.PUYOLANTON@KCL.AC.UK
Gastao Cruz[1] GASTAO.CRUZ@KCL.AC.UK
Claudia Prieto[1] CLAUDIA.PRIETO@KCL.AC.UK
Andrew P. King[1] ANDREW.KING@KCL.AC.UK
Julia A. Schnabel[1] JULIA.SCHNABEL@KCL.AC.UK
[1] *School of Biomedical Engineering & Imaging Sciences , King's College London, UK*
[2] *Data Science Institute and Department of Medicine, Imperial College London, UK*

Abstract

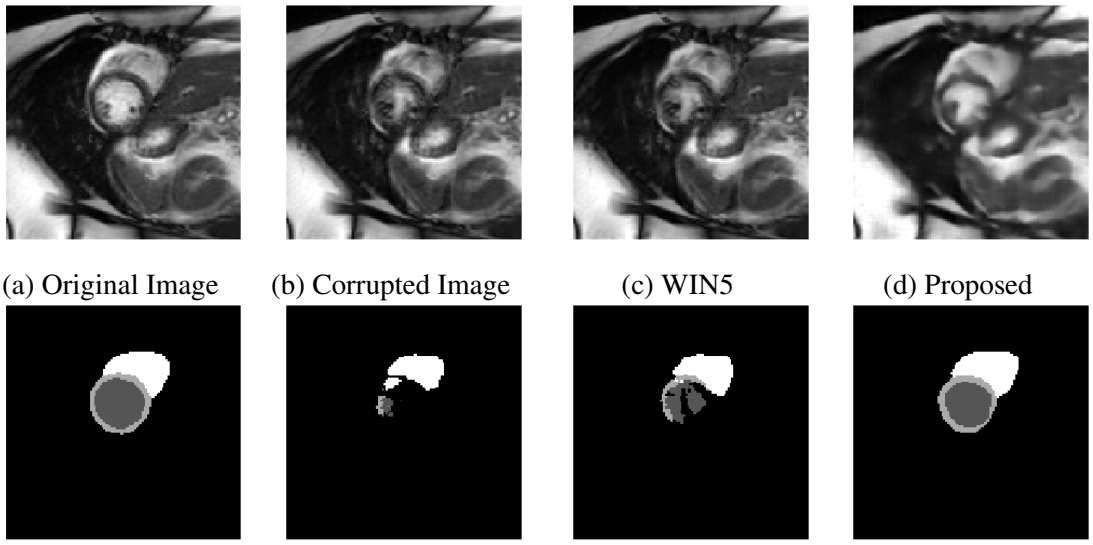
Deep learning methods have shown great success in segmenting the anatomical and pathological structures in medical images. This success is closely bounded with the quality of the images in the dataset that are being segmented. A commonly overlooked issue in the medical image analysis community is the vast amount of clinical images that have severe image artefacts. In this paper, we discuss the implications of image artefacts on cardiac MR segmentation and compare a variety of approaches for motion artefact correction with our proposed method Automap-GAN. Our method is based on the recently developed Automap reconstruction method, which directly reconstructs high quality MR images from k-space using deep learning. We propose to use a loss function that combines mean square error with structural similarity index to robustly segment poor-quality images. We train the reconstruction network to automatically correct for motion-related artefacts using synthetically corrupted CMR k-space data and uncorrected reconstructed images. In the experiments, we apply the proposed method to correct for motion artefacts on a large dataset of 1,400 subjects to improve image quality. The improvement of image quality is quantitatively assessed using segmentation accuracy as a metric. The segmentation is improved from 0.63 to 0.72 dice overlap after artefact correction. We quantitatively compare our method with a variety of techniques for recovering image quality to showcase the influence on segmentation. In addition, we qualitatively evaluate the proposed technique using k-space data containing real motion artefacts.

Keywords: Cardiac MR Segmentation, Image Quality, Image Artefacts, Image Artefact Correction, Deep Learning, UK Biobank, Automap

1. Introduction

Image segmentation is an extensively investigated problem in medical imaging, for which deep learning methods have demonstrated considerable success (Litjens et al., 2017). In general, neural network architectures are trained using controlled databases, but their performance does not always translate to clinical data in practice. Variability of image acquisition protocols, the presence of

pathology (Shao et al., 2018) and image artefacts (Oksuz et al., 2018b) can all cause low segmentation accuracy. In particular for cardiac magnetic resonance (CMR) images, which can contain a variety of different imaging artefacts (Ferreira et al., 2013), it is often difficult to segment structures and extract cardiac indices using trained neural networks on data 'in the wild', which hinders the translation of deep learning models into clinical practice. In this work, we first highlight the low segmentation accuracy when the original image is of low quality. We then propose a pipeline to improve the image quality, which translates into improved accuracy in the subsequent segmentation task. We compare a wide range of methods that can address the problem of image quality in CMR imaging. For poor quality CMR images, ground truth segmentations are usually not available, and traditionally, these low quality images are excluded from further analysis, leading to the need to recall patients to reacquire images. In the context of cohort studies, excluding individual subjects diminishes the research value of the evaluation.



(a) Original Image (b) Corrupted Image (c) WIN5 (d) Proposed

(e) Original Mask (f) Corrupted Mask (g) WIN5 Mask (h) Proposed Mask

Figure 1: Image quality is a prerequisite for high segmentation accuracy. The segmentation accuracy of the state-of-the-art network (Bai et al., 2018) drops with degradation of image quality (e,f). Our proposed method is able to generate high quality images (d), which improves the segmentation (h) compared to a state-of-the-art image denoising technique (Liu and Fang, 2017) (c,g)

Common motion artefacts in CMR manifest as temporal and/or spatial blurring of the images, which makes subsequent processing difficult (Ferreira et al., 2013). Examples of a good quality image and a synthetically motion-corrupted image are shown in Fig.1a and 1b for a short-axis view CINE CMR scan. Applying a state-of-the-art CMR segmentation algorithm (Bai et al., 2018) on these images, one can see that the method is capable of segmenting the left ventricle, myocardium and right ventricle with considerable success when the original image is of good quality (Fig.1e). However, when the image quality is reduced using a synthetic corruption strategy proposed in (Oksuz et al., 2018b), the segmentation performance diminishes (Fig.1f). To improve the image quality

we use an end-to-end image artefact correction algorithm that uses k-space input. Our algorithm outputs high quality images (Fig.1d) and correspondingly improved segmentations (Fig.1h) compared to state-of-the-art image denoising techniques (Liu and Fang, 2017) (Fig.1c and g).

Our approach is based on automatically correcting for artefacts during the reconstruction process based on our previous work (Oksuz et al., 2018a). We use a deep neural network for correcting artefacts and evaluate our method on a synthetic dataset of 1400 2D+time CMR images from the UK Biobank (Petersen et al., 2015). We also evaluate the performance on real artefact cases to demonstrate the performance of our method. There are three major contributions of this work. First, we illustrate the implications of low image quality on CMR segmentation in an extensive study. Second we propose to use motion artefact correction directly from k-space using a novel loss function. Finally, we provide a thorough investigation on the influence of using different artefact correction mechanisms on CMR segmentation.

2. Background

Deep learning techniques have been utilized for segmentation problems with high success (Litjens et al., 2017). However the influence of variability in acquisition protocols, pathology and image artefacts is an often overlooked problem. In a recent work Shao et al. (2018) showed the shortcomings of deep learning in the presence of pathology for brain MRI segmentation with an emphasis on the selection of training data.

Estimating high quality images from corrupted (or under-sampled) k-space has been a more investigated subject in the literature (Han and Ye, 2018). The problem can be addressed either in the k-space domain or the image domain. One choice is to correct the k-space before applying the inverse Fourier transform (IFT) as proposed by Han and Ye (2018). A more common approach is to use the IFT on k-space and learn a mapping between the corrupted reconstructed images and good quality images. To this end a variety of image denoising techniques can be utilized such as autoencoders (Xie et al., 2012), residual learning networks (Zhang et al., 2017) or wide networks (Liu and Fang, 2017).

In the context of CMR artefact correction, early works focused on changes in acquisition schemes (Saremi et al., 2008) and analytical methods for motion artefact reduction (Kim et al., 2008). For automatic correction of the CMR, Lötjönen et al. (2005) used short-axis and long-axis images to optimize the locations of the slices using mutual information as a similarity measure. More recently deep learning methods have been utilized to accelerate MR image acquisition by using undersampling in k-space. Schlemper et al. (2018a) proposed to use a deep cascaded network to generate high quality images. In a more recent work the authors proposed to use deep latent representations for myocardial segmentation from low quality images (Schlemper et al., 2018b). Hauptmann et al. (2019) proposed to use a residual U-net to reduce aliasing artefacts due to undersampling with the purpose of accelerating image acquisition. Our work differs from these works, which rely on a relatively small number of accurate k-space profiles, since the k-space of motion corrupted images does not necessarily have accurate profiles, or if it does we do not know *a priori* which ones they are. To the best of our knowledge our work is the first work that has investigated the influence of motion artefacts and correction strategies on CMR segmentation.

3. Methods

The proposed framework of using a deep neural network for motion artefact correction on k-space data is based on a generative-adversarial network setup. Our aim is to train a successful generator to reconstruct good quality images from motion artefact corrupted k-space data similar to (Oksuz et al., 2018a) but with a novel loss function.

3.1. Network Architecture

The proposed Automap-GAN algorithm follows an adversarial setup and consists of a generator and a discriminator. The generator network follows a similar architecture to (Zhu et al., 2018), which was originally developed for image reconstruction using domain specific information. In our case we additionally use a discriminator to increase the robustness and realism of the reconstructed images. The input to the network is a complex n-by-n k-space matrix, which we concatenate into a $(2 \times n \times n)$-by-1 vector. We then use two fully connected layers: FC1 with $2 \times n \times n$ neurons and FC2 with $n \times n$ neurons. The output from FC2 is reshaped and two convolutional layers with 64 filters and 5×5 filter size are used. After that a deconvolutional layer with 64 filters of size 7×7 is applied and finally a 1×1 layer is used to aggregate the results into an image.

The discriminator takes a generated image or a real image as input and uses two convolutional layers and a final dense layer for classification. The final output of the discriminator is a decision as to whether the generated image looks real or fake. By using outputs of the generator (artefact corrected images) and the real images from the dataset the discriminator is trained to distinguish between the artefact corrected images and high quality images. The loss function for the model uses a combination of mean squared error and structural similarity index between the predicted and real images as detailed in Section 3.2.

3.2. Loss function

We use a combination of two loss functions following the idea proposed in (Zhao et al., 2017). The mean squared error (MSE) loss is defined as:

$$L_{\text{MSE}} = \frac{1}{N_p} \sum_{p=0}^{N_p} (I_x(p) - I_y(p))^2$$

where p denotes each pixel and N_p denotes the total number of pixels in images I_x and I_y.

Alongside this measure, we also computed the structural similarity index (SSIM) (Wang et al., 2004). SSIM has been shown to provide sensitivity to structural information and texture. The SSIM between two images is defined as follows for any image regions x and y:

$$\text{SSIM}(p) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}$$

where μ_x and μ_y are the average intensities for regions x and y, σ_x and σ_y are variance values for regions x and y, σ_{xy} is the covariance of regions x and y and c_1 and c_2 are constant values for stabilizing the denominator. The SSIM loss for every pixel p defined as:

$$L_{\text{SSIM}} = \frac{1}{N_p} \sum_{p=0}^{N_p} 1 - \text{SSIM}(p)$$

and the combined loss is defined as:

$$L_{\text{total}} = \alpha L_{\text{MSE}} + (1-\alpha) L_{\text{SSIM}}$$

3.3. Implementation Details

The parameters of the convolutional and fully-connected layers were initialized randomly from a zero-mean Gaussian distribution and trained until no substantial progress was observed in the training loss. In this study, we use the RMSprop optimizer to minimize the loss. One important aspect during training is the activity regularizer, which is used after the deconvolutional layer. In our implementation, we first trained without this regularizer, finding that including it early in training led to the loss being trapped in poor local minima. Once training converged without the regularizer, it was then added, which led to the generation of sharper looking images.

The training was stopped early if no significant improvement was observed. An improvement was considered significant if the relative increase in performance was at least 0.5% over 20 epochs. To better generalize the model we applied data augmentation by rotating images in increments of 90 degrees. We also found that the success of our implementation was highly sensitive to the choice of learning rate, which we set to be 2×10^{-5}. The α parameter balancing the MSE and SSIM losses was set at 0.6 for all experiments.

During training, a batch-size of 30 2D k-space datasets was used. We used the Keras Framework with Tensorflow backend for implementation and training the network took around 3 days on a NVIDIA Quadro 6000P GPU. Segmentation of a corrected single image sequence took around 2 seconds once the network was trained.

4. Experimental Results

We evaluated our algorithm on a subset of the UK Biobank dataset consisting of 148 good quality CINE CMR acquisitions. 50 temporal frames from each subject at mid-ventricular level were used to generate synthetic motion artefacts. We used 5400 2D images for training, 600 for validation and 1400 images for testing and these three sets were not over-lapping. The data were chosen to be free of other types of image quality issues such as missing axial slices and were visually verified by an expert cardiologist. The details of the acquisition protocol of the UK Biobank dataset can be found in (Petersen et al., 2015). All images were preprocessed to extract a 128×128 pixel region and were segmented using the pre-trained CNN proposed in (Bai et al., 2018).

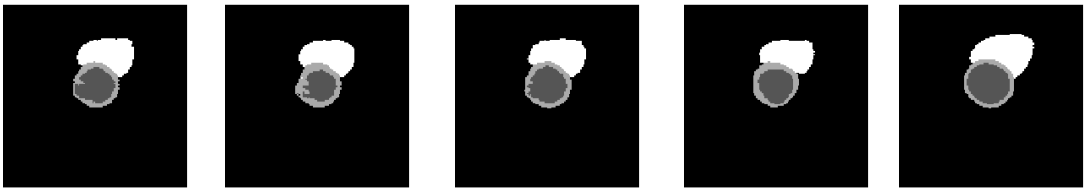
4.1. K-space corruption for synthetic data

We generated k-space corrupted data in order to simulate motion artefacts. We followed a Cartesian sampling strategy for k-space corruption to generate synthetic but realistic motion artefacts (Oksuz et al., 2018b). We first transformed each 2D short axis sequence to the Fourier domain and changed 1 in 3 Cartesian sampling lines to the corresponding lines from other cardiac phases to mimic motion artefacts. We added a random frame offset when replacing the lines. In this way the original good quality images from the training set were used to generate corresponding CMR artefact images. This is a realistic approach as the motion artefacts that occur from mis-triggering often arise from similar misplacement of k-space lines.

4.2. Results on synthetic dataset

Methods of comparison: We compared our algorithm with a variety of artefact correction strategies to produce robust segmentation results with the state-of-the-art segmentation network (Bai et al., 2018). The methods of comparison cover all possible combinations of motion artefact correction in k-space and the image domain. For image-to-image to artefact removal (i.e. post-reconstruction) we used a convolutional autoencoder (Xie et al., 2012) (CAE), a deep network based on residual learning (DNCNN) and a wide network with larger receptive fields and more channels in each layer as proposed in (Liu and Fang, 2017) (WIN5). For k-space to k-space artefact correction, we adopted the algorithm of Han and Ye (2018), which was originally proposed for accelerating MR acquisition. All of the methods used for comparison were trained with the proposed MSE/SSIM loss function. In addition to these methods we also used a variant of our algorithm, which only used MSE loss, to evaluate the influence of the combined loss function.

Qualitative evaluation: We show example segmentations achieved by different correction strategies together with the ground truth image in Figure 2. The improved image segmentation quality can be visualized for our proposed method in comparison to image-to-image denoising techniques.



(a) CAE (b) DNCNN (c) WIN5 (d) Proposed (e) Original

Figure 2: Synthetic dataset segmentation results after applying different artefact correction methods for an example case. CAE (a), DNCNN (b), WIN5 (c), proposed method (d) and ground truth segmentation (e). The proposed method is able to generate the best segmentation masks.

Quantitative evaluation: To quantitatively evaluate performance we used three metrics of left ventricular myocardium segmentation accuracy. We first computed the Dice overlap measure, which is defined between two regions A and B as:

$$D(A,B) = \frac{2\|A \cap B\|}{\|A\| \cup \|B\|}.$$

We also computed the mean contour distance (MCD) and Hausdorff distance (HD) to evaluate the mean and the maximum distances respectively between the segmentation contours C_A and C_B. These are defined as,

$$MCD(A,B) = \frac{1}{2C_A} \sum_{p \in C_A} d(p,C_B) + \frac{1}{2C_B} \sum_{p \in C_B} d(p,C_A)$$

$$HD(C_A,C_B) = \max\{\max_{a \in C_A} \min_{b \in C_B} d(a,b), \min_{b \in C_B} \max_{a \in C_A} d(a,b)\}$$

Table 1: Mean Dice, MCD and HD results of segmentation after motion artefact correction with various methods using the network output as ground truth.

	Dice	MCD	HD
Corrupted	0.635 ± 0.027	2.583 ± 1.078	6.213 ± 2.712
K-space (Han and Ye, 2018)	0.645 ± 0.038	2.214 ± 1.205	5.971 ± 2.587
CAE (Xie et al., 2012)	0.653 ± 0.041	2.109 ± 1.011	5.602 ± 2.210
DNCNN (Zhang et al., 2017)	0.662 ± 0.034	2.071 ± 0.832	5.582 ± 2.001
WIN5 (Liu and Fang, 2017)	0.681 ± 0.018	1.966 ± 0.818	5.404 ± 1.818
Proposed-MSE (Oksuz et al., 2018a)	0.679 ± 0.015	1.932 ± 0.918	4.968 ± 1.718
Proposed-Combined Loss	$\mathbf{0.722 \pm 0.014}$	$\mathbf{1.829 \pm 0.914}$	$\mathbf{4.811 \pm 1.521}$

where d represents the distance between points $a \in C_A$ and $b \in C_B$. Note that we report results for the myocardial region only for brevity. Similar results were observed for the right and left ventricle cavities.

Table 1 shows the Dice segmentation accuracies for the segmentation masks produced by the segmentation network using the output of each image artefact correction algorithm. For these experiments the ground truth was the myocardium segmentation that was generated using the network on the original uncorrupted data. The proposed adversarial Automap technique is capable of correcting motion artefacts and enabling high segmentation accuracy compared to the other techniques in both mid-ventricular slices and over all slices. We can also see that the image-to-image denoising techniques have enabled better segmentation performance compared to k-space based artefact correction. The combined MSE and SSIM loss improves segmentation performance compared to the MSE loss alone.

In Table 2, we report the Dice, MCD and HD scores, but this time using as ground truth the expert annotated masks at end-systole and end-diastole phases. Again, the adversarial Automap technique was capable of correcting motion artefacts and enabling high accuracy of segmentation compared to the other techniques. In general, segmentation performance was a bit lower compared to Table 1. This is due to the fact that the ground truth segmentations have a mean Dice score of 0.912 with the segmentations from the original images (i.e. produced by (Bai et al., 2018)) as illustrated in Table 2.

4.3. Qualitative results on real motion artefact case

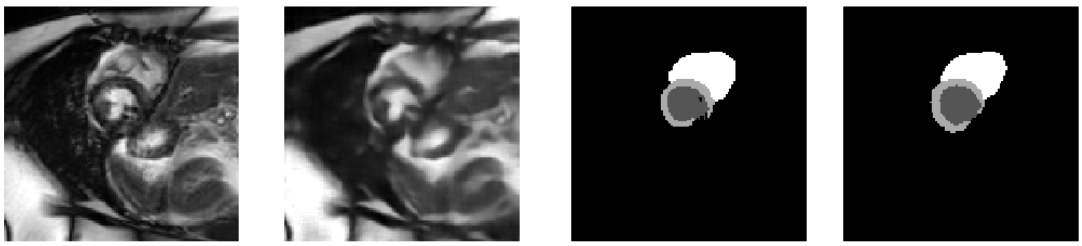
To illustrate the performance of our technique on artefact correction, we applied it to a dataset from the UK Biobank containing real mis-triggering artefacts (i.e. not synthetically corrupted). This dataset could not be segmented by the expert cardiologists due to severe motion artefacts. The visual segmentation and image quality results are illustrated in Figure 3, which shows improved image quality and CMR segmentation with the network especially in the left ventricular blood pool.

5. Discussion and Conclusion

In this paper, we have evaluated the influence of a variety of image artefact correction mechanisms on CMR segmentation accuracy. We have proposed an architecture with a combined MSE and

Table 2: Mean Dice, MCD and HD results of segmentation after motion artefact correction with various methods using the expert annotated mask as ground truth.

	Dice	MCD	HD
Original Images (Bai et al., 2018)	0.912 ± 0.027	0.876 ± 0.497	2.583 ± 0.625
Corrupted	0.590 ± 0.053	2.691 ± 1.184	6.587 ± 2.018
K-space (Han and Ye, 2018)	0.613 ± 0.032	2.304 ± 1.004	6.021 ± 1.907
CAE (Xie et al., 2012)	0.641 ± 0.040	2.184 ± 1.118	5.819 ± 1.820
DNCNN (Zhang et al., 2017)	0.657 ± 0.047	2.182 ± 0.972	5.712 ± 1.576
WIN5 (Liu and Fang, 2017)	0.676 ± 0.038	2.032 ± 0.808	5.481 ± 1.471
Proposed-MSE (Oksuz et al., 2018a)	0.671 ± 0.037	2.018 ± 0.701	5.261 ± 1.487
Proposed-Combined Loss	$\mathbf{0.696 \pm 0.027}$	$\mathbf{1.958 \pm 0.674}$	$\mathbf{5.161 \pm 1.107}$



(a) Artefact Image (b) Proposed (c) Artefact mask (d) Proposed mask

Figure 3: Example of a mis-triggering artefact and its segmentation from the UK Biobank dataset. Artefact image (a), corrected image with proposed method (b), segmentation of artefact image (c) and segmentation after proposed correction (d). The proposed method is able to correct the motion artefacts and improve segmentation especially inside the left ventricular blood pool. (The ground truth segmentation does not exist for this dataset.)

SSIM loss to achieve high segmentation accuracy. We have shown that low image quality has a deteriorating influence on image segmentation pipelines, which should be taken into account. Our method has great potential to address the problem of low segmentation accuracy on clinical data, which is likely to contain more motion corruption compared to the UK Biobank data. One limitation is the high memory requirements of our technique to correct motion artefacts.

Future work will involve end-to-end image artefact correction and segmentation mechanisms to improve the image quality and segmentation accuracy simultaneously. Moreover, we will investigate alternative implementations of our algorithm to reduce the memory requirements.

In conclusion, our work opens up the path for a reconsideration of common assumptions about how well segmentation accuracy transfers to clinical data 'in the wild'.

Acknowledgments

This work was supported by an EPSRC programme Grant (EP/P001009/1) and the Wellcome EPSRC Centre for Medical Engineering at the School of Biomedical Engineering and Imaging Sci-

ences, King's College London (WT 203148/Z/16/Z). This research has been conducted using the UK Biobank Resource under Application Number 17806. The GPU used in this research was generously donated by the NVIDIA Corporation.

References

Wenjia Bai, Matthew Sinclair, Giacomo Tarroni, Ozan Oktay, Martin Rajchl, Ghislain Vaillant, Aaron M. Lee, Nay Aung, Elena Lukaschuk, Mihir M. Sanghvi, Filip Zemrak, Kenneth Fung, Jose Miguel Paiva, Valentina Carapella, Young Jin Kim, Hideaki Suzuki, Bernhard Kainz, Paul M. Matthews, Steffen E. Petersen, Stefan K. Piechnik, Stefan Neubauer, Ben Glocker, and Daniel Rueckert. Automated cardiovascular magnetic resonance image analysis with fully convolutional networks. *Journal of Cardiovascular Magnetic Resonance*, 20(1):65, Sep 2018.

Pedro F Ferreira, Peter D Gatehouse, Raad H Mohiaddin, and David N Firmin. Cardiovascular magnetic resonance artefacts. *Journal of Cardiovascular Magnetic Resonance*, 15(1):41, 2013.

Yoseob Han and Jong Chul Ye. k-space deep learning for accelerated mri. *arXiv preprint arXiv:1805.03779*, 2018.

Andreas Hauptmann, Simon Arridge, Felix Lucka, Vivek Muthurangu, and Jennifer A Steeden. Real-time cardiovascular mr with spatio-temporal artifact suppression using deep learning–proof of concept in congenital heart disease. *Magnetic resonance in medicine*, 81(2):1143–1156, 2019.

Yoon-Chul Kim, Jon-Fredrik Nielsen, and Krishna S Nayak. Automatic correction of echo-planar imaging (epi) ghosting artifacts in real-time interactive cardiac mri using sensitivity encoding. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 27(1):239–245, 2008.

Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen A.W.M. van der Laak, Bram van Ginneken, and Clara I. Sánchez. A survey on deep learning in medical image analysis. *Medical Image Analysis*, 42:60 – 88, 2017.

Peng Liu and Ruogu Fang. Wide inference network for image denoising via learning pixel-distribution prior. *arXiv preprint arXiv:1707.05414*, 2017.

Jyrki Lötjönen, Mika Pollari, Sari Kivistö, and Kirsi Lauerma. Correction of motion artifacts from cardiac cine magnetic resonance images1. *Academic radiology*, 12(10):1273–1284, 2005.

Ilkay Oksuz, James Clough, Aurelien Bustin, Gastao Cruz, Claudia Prieto, Rene Botnar, Daniel Rueckert, Julia A. Schnabel, and Andrew P. King. Cardiac mr motion artefact correction from k-space using deep learning-based reconstruction. In Florian Knoll, Andreas Maier, and Daniel Rueckert, editors, *Machine Learning for Medical Image Reconstruction*, pages 21–29, Cham, 2018a. Springer International Publishing. ISBN 978-3-030-00129-2.

Ilkay Oksuz, Bram Ruijsink, Esther Puyol-Antón, Aurelien Bustin, Gastao Cruz, Claudia Prieto, Daniel Rueckert, Julia A. Schnabel, and Andrew P. King. Deep learning using k-space based data augmentation for automated cardiac mr motion artefact detection. In Alejandro F. Frangi,

Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI 2018*, pages 250–258, Cham, 2018b. Springer International Publishing. ISBN 978-3-030-00928-1.

Steffen E Petersen, Paul M Matthews, Jane M Francis, Matthew D Robson, Filip Zemrak, Redha Boubertakh, Alistair A Young, Sarah Hudson, Peter Weale, Steve Garratt, et al. Uk biobank's cardiovascular magnetic resonance protocol. *Journal of cardiovascular magnetic resonance*, 18 (1):8, 2015.

Farhood Saremi, John D Grizzard, and Raymond J Kim. Optimizing cardiac mr imaging: practical remedies for artifacts. *Radiographics*, 28(4):1161–1187, 2008.

Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic mr image reconstruction. *IEEE transactions on Medical Imaging*, 37(2):491–503, 2018a.

Jo Schlemper, Ozan Oktay, Wenjia Bai, Daniel C Castro, Jinming Duan, Chen Qin, Jo V Hajnal, and Daniel Rueckert. Cardiac mr segmentation from undersampled k-space using deep latent representation learning. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 259–267. Springer, 2018b.

Muhan Shao, Shuo Han, Aaron Carass, Xiang Li, Ari M Blitz, Jerry L Prince, and Lotta M Ellingsen. Shortcomings of ventricle segmentation using deep convolutional networks. In *Understanding and Interpreting Machine Learning in Medical Image Computing Applications*, pages 79–86. Springer, 2018.

Zhou Wang, Alan C Bovik, Hamid R Sheikh, Eero P Simoncelli, et al. Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4):600–612, 2004.

Junyuan Xie, Linli Xu, and Enhong Chen. Image denoising and inpainting with deep neural networks. In *Advances in neural information processing systems*, pages 341–349, 2012.

Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26 (7):3142–3155, 2017.

Hang Zhao, Orazio Gallo, Iuri Frosio, and Jan Kautz. Loss functions for image restoration with neural networks. *IEEE Transactions on Computational Imaging*, 3(1):47–57, 2017.

Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487, 2018.