

SEGMENTATION INFORMED DEEP LEARNING ALGORITHMS FOR CARDIAC MRI RECONSTRUCTION

A THESIS SUBMITTED TO
THE GRADUATE SCHOOL OF ENGINEERING AND SCIENCE
OF BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF
MASTER OF SCIENCE
IN
ELECTRICAL AND ELECTRONICS ENGINEERING

By
Mert Acar
August 2023

SEGMENTATION INFORMED DEEP LEARNING ALGORITHMS
FOR CARDIAC MRI RECONSTRUCTION

By Mert Acar

August 2023

We certify that we have read this thesis and that in our opinion it is fully adequate,
in scope and in quality, as a thesis for the degree of Master of Science.

Tolga Çukur(Advisor)

Mustafa Alper Selver

Emine Ülkü Sarıtaş Çukur

İlkay Öksüz

Approved for the Graduate School of Engineering and Science:

Orhan Arıkan
Director of the Graduate School

ABSTRACT

SEGMENTATION INFORMED DEEP LEARNING ALGORITHMS FOR CARDIAC MRI RECONSTRUCTION

Mert Acar

M.S. in Electrical and Electronics Engineering

Advisor: Tolga Çukur

August 2023

Deep learning methods have produced impressive results in accelerated magnetic resonance imaging (MRI) reconstruction from under-sampled k-space acquisitions. However, existing MRI reconstruction models are commonly trained with loss functions that uniformly weigh contributions from separate voxels across the field-of-view (FOV), without attributing focus on relatively important regions within the FOV. Furthermore common frameworks for model training rely on availability of large sets of fully-sampled MRI data to construct a ground-truth for the network output. This heavy reliance is undesirable as it is challenging to collect such large datasets in many applications, and even impossible for high spatiotemporal-resolution protocols. In this thesis, we first introduce a self-supervised learning methodology for dynamic cardiac MRI that trains the network to reconstruct acquisitions in the absence of fully-sampled data. We then introduce a segmentation-aware reconstruction framework which implicitly guides the reconstruction process around an ROI with the segmentation error signal. Lastly, we introduce RATNet, a reconstruction framework augmented with attention capabilities which explicitly carries spatial information into the reconstruction process to focus around regions of interest. Self-supervision reduces the excessive demand on fully-sampled data whereas the segmentation-aware reconstruction framework backpropagates the spatial information signal in to the reconstruction network. Lastly, RATNet incorporates the attention layers into reconstruction which are sensitive to focusing information supplied by the spatial information network. We demonstrate recovering fully-sampled images from under-sampled acquisitions in cardiac MRI and show their state-of-the-art performance in medical image reconstruction.

Keywords: MRI Reconstruction.

ÖZET

KARDİYAK MRG REKONSTRÜKSİYONU İÇİN BÖLÜTLEME BİLGİSİYLE DESTEKLENEN DERİN ÖĞRENME ALGORİTMALARI

Mert Acar

Elektrik ve Elektronik Mühendisliği, Yüksek Lisans

Tez Danışmanı: Tolga Çukur

Ağustos 2023

Derin öğrenme yöntemleri, az örneklenmiş sinyal uzayından hızlandırılmış manyetik rezonans görüntüleme (MRG) rekonstrüksiyonunda etkileyici sonuçlar üretmiştir. Ancak, mevcut MRG rekonstrüksiyon modelleri, alan içindeki ayrı ayrı voksellerden katkıları homojen bir şekilde değerlendiren hata fonksiyonlarıyla yaygın olarak eğitilir ve alan içindeki göreceli olarak önemli bölgelere odaklanmazlar. Ayrıca, model eğitimi için yaygın olarak kullanılan ortak sistemler, ağ çıktısı için bir referans oluşturmak için tam örneklenmiş büyük MRG veri kümesinin bulunabilirliğine dayanır. Bu bağımlılık, birçok uygulamada büyük veri kümelerini toplamanın zor olması nedeniyle istenmeyen bir durumdur ve yüksek uzamsal-zamansal çözünürlük protokolleri için çoğu zaman imkansızdır. Bu tezde, öncelikle dinamik kalp MRG için öz denetimli öğrenme yöntemini tanıtarak ağı, tam örneklenmiş veri olmadan rekonstrüksiyon yapacak şekilde eğitiyoruz. Daha sonra, rekonstrüksiyon sürecini bölümlleme hata sinyaliyle tanı için önemli bölge etrafında yönlendiren bir rekonstrüksiyon sistemi tanıtıyoruz. Son olarak, odaklanma bölgelerine yönelik olarak mekansal bilgiyi açıkça taşıyan bir dikkat yeteneği ile güçlendirilmiş bir rekonstrüksiyon sistemi olan RAT-Net'i tanıtıyoruz. Öz denetime dayalı öğrenme, tam örneklenmiş verilere olan fazla talebi azaltırken, bölümlleme bilinçli rekonstrüksiyon sistemi, mekansal bilgi sinyalini rekonstrüksiyon ağına geri yayar. Son olarak, RATNet, rekonstrüksiyonda mekansal bilgi ağı tarafından sağlanan odaklanma bilgisine duyarlı olan dikkat katmanlarını içerir. Sonuç olarak kardiyak MRG'de az örneklenmiş alımlardan tam örneklenmiş görüntülerin geri kazanılmasını gösteriyor ve tıbbi görüntü rekonstrüksiyonunda en üstün performanslarını sergiliyoruz.

Anahtar sözcükler: MRG Rekonstrüksiyonu.

Acknowledgement

The work presented in this thesis has been produced benefiting from the 2232 International Fellowship for Outstanding Researchers Program of TUBITAK (Project No: 118C353). However, the entire responsibility of the publication belongs to the owner of the thesis. The financial support received from TUBITAK does not mean that the content of the publication is approved in a scientific sense by TUBITAK.

Contents

1	Introduction	1
2	Self-Supervised Dynamic MRI Reconstruction	6
2.1	Theory	7
2.1.1	Dynamic MRI Reconstruction	7
2.1.2	Self-Supervised Learning	9
2.2	Methods	10
2.3	Experimental Results	13
2.4	Conclusion	15
3	Segmentation-Aware MRI Reconstruction	16
3.1	Methods	17
3.1.1	Proposed Framework	17
3.1.2	Stabilization	17
3.1.3	Model Architectures	19

3.1.4	Implementation Details	22
3.2	Experimental Results	22
3.3	Conclusion	24
4	Region-of-Interest-Attentive MRI Reconstruction	25
4.1	Related Work	27
4.2	Theory	29
4.2.1	Accelerated MRI Reconstruction	30
4.2.2	Reconstruction Network	31
4.2.3	Spatial Attention Module	32
4.2.4	Spatial Loss Network	32
4.2.5	Learning Procedures	33
4.3	Methods	36
4.3.1	Network Architecture	36
4.3.2	Methods of Comparison	39
4.3.3	Datasets	41
4.3.4	Evaulation Metrics	42
4.4	Experimental Results	43
4.4.1	Single-Coil Reconstruction	43
4.4.2	Multi-Coil Reconstruction	44

4.4.3	Ablation Study	44
4.5	Discussion	45
4.6	Conclusion	47

5	Conclusion	48
---	------------	----



List of Figures

- 2.1 The self-supervised training strategy adapted to an unrolled network architecture for dynamic MRI reconstructions. 10
- 2.2 Representative reconstructions from competing architectures at (a) 4-fold, (b) 8-fold acceleration. The first column displays Fourier reconstructions of fully-sampled and undersampled data. In remaining columns, the second row shows error maps for each architecture. The prefix 'SS' denotes a model trained via self supervision. 14
- 3.1 Segmentation-aware reconstruction framework. During training the network is trained end-to-end with stabilization. During inference an image is recollected from only the reconstruction network, along with an auxiliary segmentation map. Stabilization technique is illustrated by the sample inputs on the left across epochs 18
- 3.2 Representative reconstructions from competing techniques at (a) 4-fold, (b) 8-fold acceleration. The second row displays the localized reconstruction around the heart. The area on which the local performance calculations are taken is indicated with the red rectangle. First column displays Fourier reconstructions of fully-sampled data. In remaining columns, reconstruction with no help from the segmentation maps, joint reconstruction without stabilization and joint reconstruction with stabilization is shown respectively. . . . 20

4.1	RATNet model with reconstruction network and loss function architecture. 1 depicts the regularization block architecture with the attention module addition. 2 details the data consistency scheme.	26
4.2	Reconstructions of representative brain magnitude acquisitions from TCGA-LGG dataset at 4-fold and 8-fold acceleration rates are shown for self-attention based methods (SATNet, SOGAN), focusing based methods (ROIRecNet, ROICSNet), proposed RATNet along with reference image. Zoom-in display windows are added to aid visualization of performance differences around diagnostically relevant sub-regions.	29
4.3	Reconstructions of representative coil-combined cardiac acquisitions at 4-fold and 8-fold acceleration rates are shown for the classical CS based LORAKS method, self-attention based methods (SATNet, SOGAN), focusing based methods (ROIRecNet, ROICSNet), proposed RATNet along with reference image. Zoom-in display windows are added to aid visualization of performance differences around diagnostically relevant sub-regions.	33
4.4	Attention maps in RATNet for three different test samples from OCMR dataset in coil-combined setup. Map activations are taken from the attention module of the last regularization block.	35

List of Tables

2.1	Performance comparisons (MSE, PSNR and SSIM) on dynamic cardiac data with x4 acceleration rate across experiment setups. MSE is scaled with 10^3 . Self-supervised schemes have 'SS' prefix while static reconstruction settings are indicated by '2D' suffix. . .	11
2.2	Performance comparisons (MSE, PSNR and SSIM) on dynamic cardiac data with x8 acceleration rate across experiment setups. MSE is scaled with 10^3 . Self-supervised schemes have 'SS' prefix while static reconstruction settings are indicated by '2D' suffix. . .	12
3.1	Performance comparisons of various baselines on cardiac MRI data across x4 and x8 acceleration rates. (S) suffix signifies the stabilization technique. "F" stands for focused measurements that are taken over the area of diagnostic interest. Stabilized training aids Multi-Decoder Unet and UNet $\rightarrow$ UNet in focused metrics indicating improved reconstruction quality for cardiac cavity.	21

3.2	Performance comparisons (MSE, PSNR and SSIM) on cardiac MRI data with x4 and x8 acceleration rates across experiment setups. MSE is scaled with 10^3 . (S) suffix indicates the stabilization technique. "F" stands for focused measurements that are taken over the area of diagnostic interest. Stabilized training improves the focused metrics indicating improved reconstruction quality for cardiac cavity.	24
4.1	Reconstruction performance of competing methods on coil-combined OCMR and TCGA-LGG datasets at 4-fold acceleration. "F" prefix stands for the focused measurements taken around the diagnostically relevant sub-regions of the image.	36
4.2	Reconstruction performance of competing methods on coil-combined OCMR and TCGA-LGG datasets at 8-fold acceleration. "F" prefix stands for the focused measurements taken around the diagnostically relevant sub-regions of the image.	38
4.3	Reconstruction performance of competing methods on multi-coil OCMR dataset at 4-fold and 8-fold accelerations. "F" prefix stands for the focused measurements taken around the diagnostically relevant sub-regions of the image.	42
4.4	Ablation study on proposed method. "A" stands for spatial attention modules and "SL" stands for the use of spatial loss function. Illustrates the effect of each sub-component on the reconstruction quality. Performance metrics around the diagnostically relevant sub-regions of the image are indicated via the prefix "F" for "focused".	45

Chapter 1

Introduction

Based on the acquisition of the Fourier domain representation of proton density within the body through the application of appropriate RF pulses and spatially-dependent magnetic fields, magnetic resonance imaging (MRI) has become a highly popular imaging technique. An analysis from [1] reveals growth patterns in the number of MRI scanners and MRI scans per inhabitant in selected OECD countries, demonstrating the increasing popularity of MRI. According to estimations, there were approximately 50,000 MRI scanners worldwide by 2018, and around 5,000 new ones are sold every year [1]. This popularity is driven by the numerous advantages of MRI compared to other imaging techniques.

MRI provides exceptional image quality, increased diagnostic information through multiple contrasts, and superior soft-tissue resolution compared to CT scans, all while avoiding ionizing radiation [2]. The ability to inspect tissues under multiple different contrasts on the same device is particularly beneficial for radio-diagnostic purposes [3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14].

However, a significant challenge in MRI lies in the prolonged scan time [15, 16, 17]. Multidimensional k-space acquisition through free induction decay or echo signals results in extended scan durations. A routine imaging times for various examinations range from 20 to 60 minutes [18]. Reducing this duration

would make examinations more cost-effective, increase accessibility for more individuals, decrease motion artifacts caused by patient movement, allow imaging of young children without sedation or general anesthesia, accommodate the limited ability of some elderly individuals to remain completely still, and enhance patient satisfaction [18].

Nevertheless, a challenge arises due to the relationship between signal-to-noise ratio (SNR) and contrast-to-noise ratio (CNR) in imaging, which generally follows a proportional relationship to the square root of time [18]. When imaging time is shortened without a well-designed acceleration method, artifacts occur, and image quality deteriorates due to reasons such as Nyquist criterion violation. Consequently, ongoing research focuses on MRI acceleration techniques that aim to reduce multi-contrast MRI scan times without significantly compromising image quality [19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 28, 31].

One avenue of research in this area is parallel imaging. Abbreviated as pMRI, parallel imaging techniques utilize imaging from multiple receiver coils to exploit the distinct spatial sensitivities of each coil. This reduces aliasing artifacts resulting from undersampling. After the introduction of SMASH in 1997, a multitude of parallel imaging methods have been proposed. These methods fall into categories such as image domain-based methods (e.g., SENSE and PILS), k-space-based methods (e.g., SMASH, AUTO-SMASH, and GRAPPA), and hybrid techniques (e.g., SPACE RIP and Generalized SENSE) [32]. These techniques generally utilize coil sensitivity information acquired through a prescan, autocalibration during the experiment, or a combination of both.

The SENSE (sensitivity encoding) method, the most commonly used pMRI technique [32], operates by undoing the signal superposition caused by aliasing through solving a linear system of equations using coil sensitivity information. More details can be found in [33]. The GRAPPA (generalized autocalibrating partially parallel acquisitions) method involves partial k-space acquisition with a fully sampled autocalibration signal (ACS) region. It generates individual coil images by estimating missing k-space data for each coil using a kernel constructed from the ACS. These coil images are then combined [34]. The SPACE RIP

(sensitivity profiles from an array of coils for encoding and reconstruction in parallel) method discretizes the multi-coil imaging equation using known coil sensitivities. It transforms the equation into a linear system with a large matrix to be inverted and solved [35].

Another research direction in MRI acceleration is compressed sensing (CS). Compressed sensing capitalizes on the sparsity of MRI signals in certain transform domains like wavelet transform or total variation. It achieves this through subsampling of k-space to generate incoherent, noise-like aliasing artifacts. The approach involves solving an optimization problem, such as the one shown below [36]:

$$\begin{aligned} &\text{minimize } \|\Psi m\|_1 \\ &\text{s.t. } \|F_S m - y\|_2 < \epsilon \end{aligned} \tag{1.1}$$

Here, m represents the reconstructed image, Ψ denotes the linear operator for a transform expected to yield a sparse image representation, y represents measured k-space data, and ϵ controls the consistency between acquired measurements and reconstruction. Other terms are occasionally added for improved performance. Successful implementation of CS requires meeting three conditions: the image should have a sparse representation in the transform domain, artifacts from undersampling should be incoherent (often achieved through random or non-Cartesian sampling trajectories), and a non-linear reconstruction algorithm balancing sparsity in the transform domain and consistency with undersampled k-space data must be utilized. Despite its adoption and potential, CS has limitations. In 2D sequences, achieving incoherence can be challenging [37]. Additionally, blocky or unnatural reconstructions may arise due to sparsifying transforms being insufficient to capture the complexity of biological anatomies [38]. The iterative solution required for CS optimization can also lead to undesirably long reconstruction times [38]. Lastly, the tuning of hyperparameters in CS presents a challenge, with the risk of excessively smooth or unnatural images on one end and residual artifacts from undersampling on the other.

Another direction of research pursued in this topic is deep learning (DL) approaches for image reconstruction in MRI, which have shown its rapid developments in recent years [39, 40, 41, 42, 43]. Reconstruction refers to the task of mapping undersampled acquisitions to high-quality images that are as consistent as possible with corresponding fully-sampled acquisitions. Many recent methods for MRI reconstruction using learning-based techniques leverage unrolled network architectures exemplified by works such as [44, 45, 46, 39]. Note that there are a number of other approaches that use models without physics-driven data-consistency modules or generative models as their backbones [47, 48, 49, 50, 51].

Unrolling-based approaches entail the construction of network architectures based on the iterative steps of an optimization algorithm. This algorithm aims to iteratively refine the image reconstruction, and the learned network parameters incorporate both the reconstruction model and algorithmic steps through training.

Conversely, approaches that are not reliant on unrolling commonly leverage standard network architectures, often designed for tasks other than MRI reconstruction [43, 52, 53]. These networks are adapted to learn the mapping from input to output, occasionally integrating domain-specific MRI knowledge into the learning process [54, 55, 56, 57].

This thesis firstly introduces a self-supervised reconstruction framework for dynamic MRI sequences which include a temporal axis in addition to spatial axes. For dynamic MRI, recent studies have mostly expressed the reconstruction as a static 2D or 3D problem where frames are independently handled as separate instances [58], [44]. The unrolled architecture is trained end-to-end using a large database of fully-sampled MRI data along with corresponding undersampled acquisitions. Supervised learning has produced impressive results to date, however sizable dynamic MRI datasets may not be available in many applications.

Secondly, this thesis introduces a segmentation-aware MRI reconstruction framework to guide the reconstruction effort around an region of interest (ROI) using corresponding segmentation information. DL models are trained to recover

high-quality images consistent with fully-sampled acquisitions given as input undersampled acquisitions. Training is typically performed with a global loss function expressed over the entire image. However, such global loss functions are often dominated by diagnostically-irrelevant background tissues.

Finally, this thesis introduces a ROI-attentive MRI reconstruction pipeline which builds upon our previous work. DL models are trained to recover high-quality images consistent with fully-sampled acquisitions given as input undersampled acquisitions. Training is typically performed with a global loss function expressed over the entire image. However, such global loss functions are often dominated by diagnostically-irrelevant background tissues.

Chapter 2

Self-Supervised Dynamic MRI Reconstruction

The work described in this chapter has been published as *Acar, M., Çukur, T., Öksüz, İ. (2021). Self-supervised Dynamic MRI Reconstruction. In: Haq, N., Johnson, P., Maier, A., Würfl, T., Yoo, J. (eds) Machine Learning for Medical Image Reconstruction. MLMIR 2021. Lecture Notes in Computer Science, vol 12964. Springer, Cham. https://doi.org/10.1007/978-3-030-88552-6_4*

The excessive demand for high-quality ground-truth data in supervised learning has sparked interest in self-supervision methods for MRI reconstruction. Promising results have been reported in recent studies for static MRI applications with deep network trained without any access to fully-sampled data [59], [60], [61]. However, the utility of self-supervision approaches in dynamic cardiac MRI currently remains unknown to the best of our knowledge. The main motivation for the current study is to devise self-supervision strategies in training of deep reconstruction models for dynamic cardiac MRI, given the severely limited number and scope of public datasets on cardiac MRI. Inspired by a recent self-supervision approach for static MRI, here we introduce a self-supervised deep model for dynamic MRI that learns to self-recover subsets of data in readily undersampled acquisitions [59]. To do this, data are split into two nonoverlapping

segments in k-space, where the first set is used to enforce data consistency while the second is used to learn network weights.

Supervised deep models of greater complexity are commonly observed to elevate task performance due to their representational capacity [62]. In contrast, we hypothesize that elevating complexity in self-supervised models can instead constrain model performance due to deficiencies in training data. To test this hypothesis, here we implement the proposed self-supervised learning strategy on recent state-of-the-art deep models for dynamic MRI, with varying degrees of complexity. We consider a convolutional recurrent neural network (CRNN) model of low complexity [45], a dynamic reconstruction network (DRN) model of intermediate complexity [62], and a motion-guided DRN (MODRN) model [62] of high complexity. To examine benefits of exploiting temporal correlations, we comparatively demonstrate these models against frame-by-frame reconstructions using a decoupled version of DRN, and a cascade network (CascadeNet) model [44]. Experiments were conducted on a public cardiac MRI dataset [63]. Both supervised and self-supervised variant models were trained. Our results clearly indicate that more compact models can offer increased reliability against performance loss in self-supervised settings compared to complex models.

2.1 Theory

2.1.1 Dynamic MRI Reconstruction

Dynamic MRI can be accelerated via undersampling across the phase-encoding dimension. Let the temporal sequence of fully-sampled, complex MR images is denoted as $\{\mathbf{x}_t\}_{t \in \tau} \in \mathbb{C}^N$ where each 2D frame is cast into a column vector across spatial dimensions of length $N = N_x N_y$ and τ is the number of frames. Reconstruction can be achieved by recovering $\{\mathbf{x}_t\}_{t \in \tau}$ from an set of undersampled k-space measurements $\{\mathbf{y}_t\}_{t \in \tau} \in \mathbb{C}^K (K \ll N)$ such that:

$$\mathbf{y}_t = \mathbf{F}_u \mathbf{x}_t + \mathbf{e} \quad (2.1)$$

where $\mathbf{F}_u$ is the partial Fourier encoding operator and $\mathbf{e} \in \mathbb{C}^K$ denotes measurement noise. As undersampled acquisitions violate the Nyquist condition, Eq. 2.1 is underdetermined and therefore its solutions benefits from prior information. Prior knowledge can be incorporated as a regularization term:

$$\mathcal{L}_{rec}(\{\mathbf{x}_t\}) = \sum_{t=1}^{\tau} (\mathcal{R}(\mathbf{x}_t) + \lambda \|\mathbf{F}_u \mathbf{x}_t - \mathbf{y}_t\|_2^2) \quad (2.2)$$

In Eq. 2.2 $\mathcal{R}$ stands for the regularization term imposed on $\mathbf{x}$, and $\lambda \in \mathbb{R}$ is the relative weighting of the data fidelity term against the regularizer. In DL-based reconstruction, regularization is achieved via mapping through a convolutional neural network (CNN) to map undersampled data to fully-sampled images:

$$\mathcal{L}_{rec}(\{\mathbf{x}_t\}) = \sum_{t=1}^{\tau} (\|\mathbf{x}_t - \text{cnn}(\mathbf{x}_{t_u}|\theta)\|_2^2 + \lambda \|\mathbf{F}_u \mathbf{x}_t - \mathbf{y}_t\|_2^2) \quad (2.3)$$

where $\text{cnn}(\cdot|\theta)$ is the CNN mapping characterized by the parameter vector θ , $\mathbf{x}_{t_u} = \mathbf{F}_u^H \mathbf{y}_t$. Note that $\mathbf{F}_u^H$ denotes the Hermitian of the Fourier operator.

The quality of the reconstructed image is affected heavily by the internal structure of the function $\text{cnn}(\cdot|\theta)$, which represents the architecture of the underlying network. As a common design choice across all of the experimented architectures, the regularization is done in cascaded CNN iterations interleaved with Data Consistency (DC) modules proposed in [44], which enforces the data fidelity of the already-sampled k-space points denoted with the set Ω . While the solution uses the predicted values from CNN iterations for the unknown k-space samples (i.e. $i \notin \Omega$), for the already-sampled entries (i.e. $i \in \Omega$) we take the noise-weighted linear combination of the predictions and the acquired samples. For the interpolated k-space output from the CNN layer $\{\mathbf{m}_t\}_{cnn} = \mathbf{F}\{\mathbf{x}_t\}_{cnn} = \mathbf{F}_{cnn}(\mathbf{x}_{t_u}|\theta)$:

$$\text{DC}(\{\mathbf{m}_t\}_{cnn}) = \begin{cases} \{\mathbf{m}_t\}_{cnn}(i) & \text{if } i \notin \Omega \\ \frac{\{\mathbf{m}_t\}_{cnn}(i) + \mu \{\mathbf{m}_t\}_0(i)}{1 + \mu} & \text{if } i \in \Omega \end{cases} \quad (2.4)$$

where $\{\mathbf{m}_t\}_0$ is the zero-filled k-space obtained from the zero-filled complex-valued image $\{\mathbf{m}_t\}_0 = \mathbf{F}\mathbf{F}_u^H \mathbf{y}_t$. Here the parameter μ is inversely proportional to the noise power in the acquisitions. It can be seen that in the limit $\mu \rightarrow \infty$ the data-consistency operator converges to strict data fidelity [44].

2.1.2 Self-Supervised Learning

As discussed previously, fully-sampled acquisitions in dynamic cardiac imaging are often difficult to collect due to motion, signal decay, and long scan times. To address associated challenges in training neural networks, we introduce a self-supervised learning strategy for dynamic MRI reconstructions inspired by [59]. In this strategy, a subset of k-space data in undersampled MR acquisitions is masked out, and the network is trained to predict the masked samples from non-masked samples. Thus, the network weights are optimized to minimize the prediction error for masked samples, in an effort to prevent reliance on fully-sampled ground truth. Assuming the superset of acquired k-space samples is Ω , this superset is split into two non-overlapping subsets as follows:

$$\Omega = \Theta \cup \Lambda \quad (2.5)$$

The set Θ is used in the data consistency operation during training to enforce fidelity whereas the set Λ is used to define the loss function in k-space:

$$\min_{\theta} \mathcal{L}(\mathbf{m}_{\Lambda}, \mathbf{m}_{cnn_{\Lambda}} | \theta) \quad (2.6)$$

where θ is the set of trainable network parameters, $\mathbf{m}_{\Lambda}$ are acquired k-space samples in Λ pooled across all frames, and $\mathbf{m}_{cnn_{\Lambda}}$ are reconstructed k-space samples in Λ . The optimization objective is to minimize the discrepancy between reconstructed and acquired k-space data. Meanwhile, data consistency operations are only performed for k-space samples in Θ . Here the loss function is taken as normalized ℓ_1 - ℓ_2 loss:

$$\mathcal{L}(\mathbf{m}, \mathbf{m}_{cnn}) = \sum_{t=1}^{\tau} \frac{\|\mathbf{m}_t - \mathbf{m}_{t_{cnn}}\|_2}{\|\mathbf{m}_t\|_2} + \frac{\|\mathbf{m}_t - \mathbf{m}_{t_{cnn}}\|_1}{\|\mathbf{m}_t\|_1} \quad (2.7)$$

where during supervised training $\mathbf{m}$ and $\mathbf{m}_{cnn}$ correspond to the reference fully-sampled cine series in k-space and the network output in k-space respectively. In the case of self-supervised training, these refer to k-space samples with index Λ from the measurements and the network output respectively.

The success of the self-supervision strategy considered in this study relies on the resulting model’s ability to generalize across varying subsets of acquired and missing k-space samples.

2.2 Methods

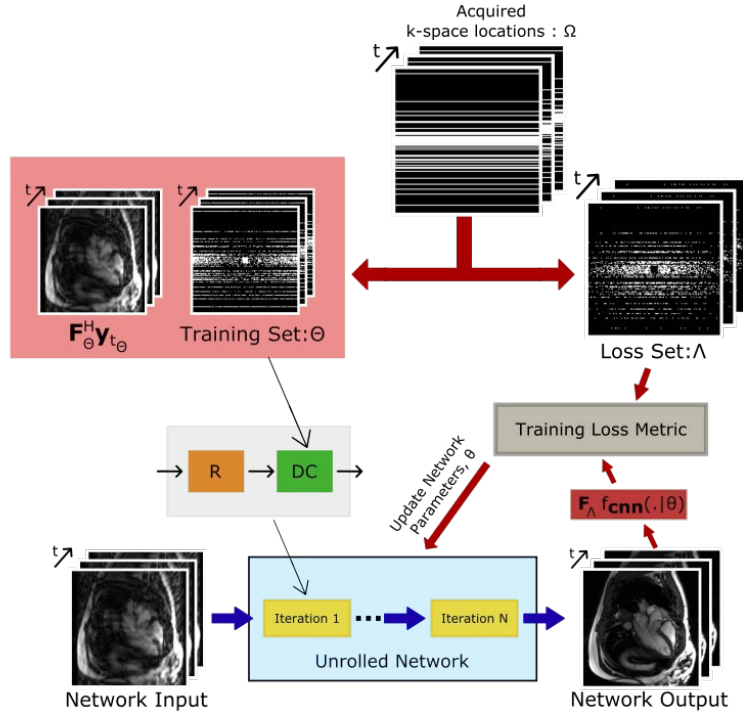


Figure 2.1: The self-supervised training strategy adapted to an unrolled network architecture for dynamic MRI reconstructions.

Self-supervision alleviates the harsh data requirements of supervised learning

for deep neural networks architectures with thousands-to-millions of parameters. However, training models on inherently lower-quality undersampled acquisitions might slow down learning procedures. Here, we hypothesized that self-supervised models would require greater amount of data for training, and so this would give a performance advantage to more compact models with fewer parameters over large models. To test this hypothesis, we evaluated several state-of-the-art deep architectures for dynamic MRI reconstruction in both supervised and self-supervised settings using the structure shown in Fig. 2.1. Critically, we experimented with architectures of varying orders of model complexity. All architectures were trained to solve Equation 2.3 by alternating between network-driven regularization and data consistency blocks.

Table 2.1: Performance comparisons (MSE, PSNR and SSIM) on dynamic cardiac data with x4 acceleration rate across experiment setups. MSE is scaled with 10^3 . Self-supervised schemes have 'SS' prefix while static reconstruction settings are indicated by '2D' suffix.

Supervised Training (x4)				
Method	PSNR	SSIM	MSE (10^{-3})	Complexity
2D-DRN	28.3257	0.89	0.0294	2,081,250
CascadeNet	31.6621	0.92	0.0158	678,924
CRNN	34.8252	0.95	0.0105	297,794
DRN	33.2935	0.93	0.0076	2,081,250
MODRN	35.4119	0.95	0.0044	21,198,822
Self-Supervised Training (x4)				
Method	PSNR	SSIM	MSE (10^{-3})	Complexity
SS-CascadeNet	27.2024	0.86	0.0388	678,924
SS-CRNN	33.6359	0.93	0.0069	297,794
SS-DRN	31.8702	0.93	0.0151	2,081,250
SS-MODRN	22.7439	0.78	0.1264	21,198,822

DRN: The dynamic reconstruction network is based on a U-Net backbone with recurrent modules to exploit redundancies in temporal and unrolling dimensions [62]. Here, we considered two variants of DRN models: a static 2D-DRN model that independently reconstructed each individual frame, and an aggregate DRN model that reconstructed all frames concurrently. A comparison among these variants allowed us to assess potential benefits of temporal correlations in dynamic MRI reconstruction.

Table 2.2: Performance comparisons (MSE, PSNR and SSIM) on dynamic cardiac data with x8 acceleration rate across experiment setups. MSE is scaled with 10^3 . Self-supervised schemes have 'SS' prefix while static reconstruction settings are indicated by '2D' suffix.

Supervised Training (x8)				
Method	PSNR	SSIM	MSE (10^{-3})	Complexity
2D-DRN	22.5123	0.81	0.0922	2,081,250
CascadeNet	26.0021	0.85	0.0592	678,924
CRNN	32.9124	0.92	0.0094	297,794
DRN	31.2637	0.91	0.0178	2,081,250
MODRN	33.4434	0.92	0.0071	21,198,822
Self-Supervised Training (x8)				
Method	PSNR	SSIM	MSE (10^{-3})	Complexity
SSS-CascadeNet	24.2161	0.81	0.0888	678,924
SS-CRNN	30.1501	0.91	0.0213	297,794
SS-DRN	28.1245	0.88	0.0419	2,081,250
SS-MODRN	16.6932	0.65	0.4244	21,198,822

MODRN: MODRN incorporates motion-estimation and motion-compensation networks into DRN to better align anatomy across separate frames [62], in an effort to better utilize correlated structural information. The motion estimation module takes as input fully-sampled reference frames to predict motion fields, as estimation performance is heavily reliant on having high-quality references. This renders MODRN unsuited to self-supervised learning strategies.

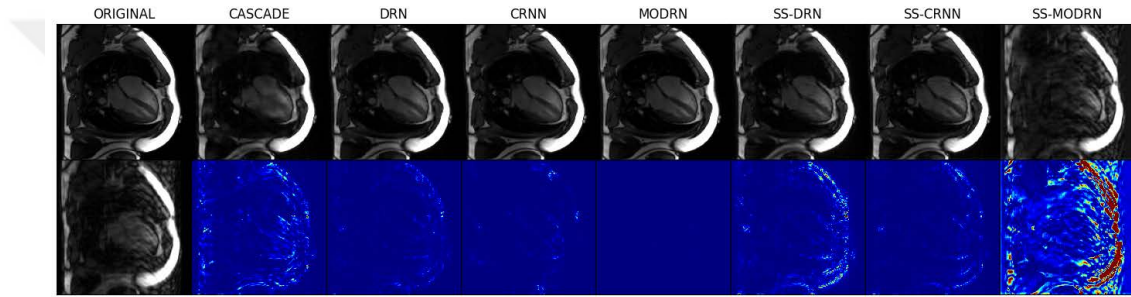
CascadeNet: The cascade networks is a baseline for static reconstruction of individual frames in dynamic MRI. It follows an unrolled architecture with interleaved data consistency and regularization blocks, and progressively suppressed aliasing artifacts in reconstructions [44].

CRNN: CRNN follows a similar unrolled architecture as CascadeNet, however, it utilizes recurrent connections to carry information along temporal and unrolling iteration dimensions as in DRN [45]. Information sharing across frames is achieved via bidirectional recurrent units to further refine reconstructions across iterations.

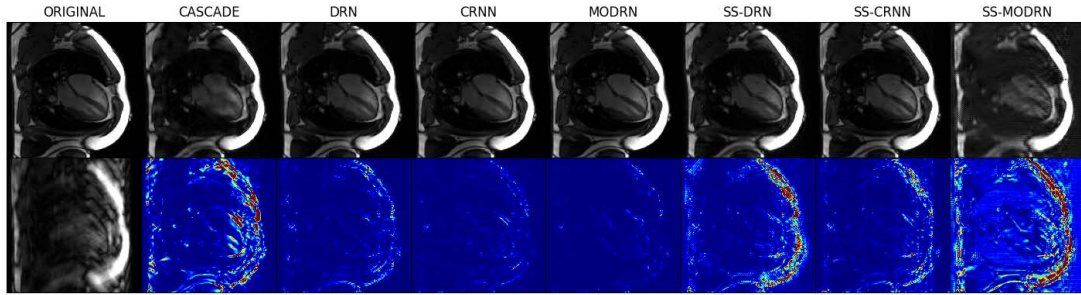
Experiments: All networks were trained using the Adam optimizer with parameters $\beta_1 = 0.99$ and $\beta_2 = 0.999$, a learning rate of 10^{-4} and batch size of 1. Models were implemented using pytorch library and executed on NVIDIA RTX 3090 GPUs. Experiments were conducted on fully-sampled MRI data from the public OCMR dataset containing CINE scans from 74 subjects [63]. Subjects had varying number of slices and frames, yielding a total of 183 slices. Data were split into independent training (155 slices) and test (28 slices) sets, with no subject overlap between the two sets. MRI data were retrospectively under-sampled to achieve acceleration rates of 4 and 8. A Gaussian sampling density with an autocalibration region containing 8 lines was used. For self-supervised learning, acquired k-space was split into two distinct sets at ratio 3:2 [59]. Reconstruction quality was assessed by measuring peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and mean-squared error (MSE) between the reconstructed and ground-truth images [64, 65].

2.3 Experimental Results

Figure 2.2 illustrates reconstructions in a representative test subject at 4x and 8x acceleration along with fully-sampled ground truth. PSNR, SSIM and MSE of all tested methods are presented in Tables 2.1 and 2.2, along with the embodied number of model parameters. While the supervised MODRN model with the greatest model complexity exhibits the highest performance, its self-supervised variant suffers heavily in the absence of fully-sampled reference data. Note that MODRN learns to create motion flow maps between the current frame being reconstructed and two reference frames marking a predetermined period T in the time axis. When supplied with undersampled frames, the estimated motion maps fail to capture anatomical motion due to undersampling artifacts greatly changing from frame to frame dominating the motion flow maps. Divorced of the motion estimation and compensation units, DRN offers somewhat lower performance for an order of magnitude fewer parameters. As it does not take any fully-sampled reference inputs, its deterioration in self-supervision is relatively lower.



(a) x4 Acceleration



(b) x8 Acceleration

Figure 2.2: Representative reconstructions from competing architectures at (a) 4-fold, (b) 8-fold acceleration. The first column displays Fourier reconstructions of fully-sampled and undersampled data. In remaining columns, the second row shows error maps for each architecture. The prefix 'SS' denotes a model trained via self supervision.

The static 2D-DRN and CascadeNet models naturally deteriorate around high-motion areas, and yield the poorest performances. Lastly, the compact CRNN model maintains on par performance to the MODRN model in the supervised setting, despite having two orders of magnitude fewer parameters. Furthermore, CRNN is able to maintain its performance reliably even in self-supervised setting. These results indicate that learning network models with self supervision increases data demands, since reference data are undersampled in this case. Therefore, compact network models have a remarkable edge for successful training via self supervision.

2.4 Conclusion

Here we introduced a self-supervision approach to deep reconstruction networks for dynamic cardiac MRI acquisitions. Experiments were conducted to systematically evaluate the amenability of models of varying complexity to self supervision. Our results indicate that a compact unrolled CNN architecture with bidirectional recurrent connections exhibits robustness against lowered quality of training data in self-supervised learning, and achieves on par performance with supervised duals.

Chapter 3

Segmentation-Aware MRI Reconstruction

The work described in this chapter has been published as *Acar, M., Çukur, T., Öksüz, İ. (2022). Segmentation-Aware MRI Reconstruction. In: Haq, N., Johnson, P., Maier, A., Qin, C., Würfl, T., Yoo, J. (eds) Machine Learning for Medical Image Reconstruction. MLMIR 2022. Lecture Notes in Computer Science, vol 13587. Springer, Cham. https://doi.org/10.1007/978-3-031-17247-2_6*

Self-attention mechanisms have been proposed to help focus the model’s attention on subregions where there is greater tendency for introducing reconstruction errors [66]. Note that vanilla attention mechanisms are not explicitly informed regarding the underlying anatomy. Other studies have considered more direct guidance from segmentation maps to focus on reconstruction performance in regions of interest. [67], [68], [69]. Pre-trained segmentation models have been transferred to mitigate problems associated with joint training of the reconstruction-segmentation network. Few recent studies have considered joint training of reconstruction-segmentation models, where an alternating optimization is performed between the two networks [70]. However, such alternative optimization is prone to premature stopping of learning where one of the networks performs unacceptably poorly.

In this chapter, we propose a segmentation-aware reconstruction method. The proposed method is based on a sequential architecture containing two networks for reconstruction and segmentation. To avoid premature stopping of learning, a stabilization approach is introduced for end-to-end training of the model. In particular, the undersampling rate is gradually decreased during the course of training. A composite reconstruction-segmentation loss is used, and errors backpropagated from the segmentation stage are used to focus the reconstruction on critical image regions. Experiments were conducted on a public cardiac MRI dataset [63]. Our results clearly indicate that the proposed segmentation-aware reconstruction improves focal image quality over solo reconstruction methods and unstabilized joint reconstruction-segmentation methods. Such localized performance benefits guided by tissue segmentation may not only be relevant for cardiac but also other imaging applications such as neuroimaging [71, 72, 73, 74, 75, 76, 77, 78, 79].

3.1 Methods

3.1.1 Proposed Framework

We propose an end-to-end training of reconstruction and segmentation networks to enable high reconstruction quality for a target region of interest (ROI). Figure 3.1 illustrates the proposed framework for segmentation-aware reconstruction for cardiac MRI reconstruction. Our framework is general in the sense that many different state-of-the-art architecture for reconstruction and segmentation modules can be utilized. The specific pairs of architectures that we examine in the current study are described in implementation details.

3.1.2 Stabilization

Our framework rests on the key notion of employing segmentation loss to focus the reconstruction process on tissues of high diagnostic interest. Accordingly,

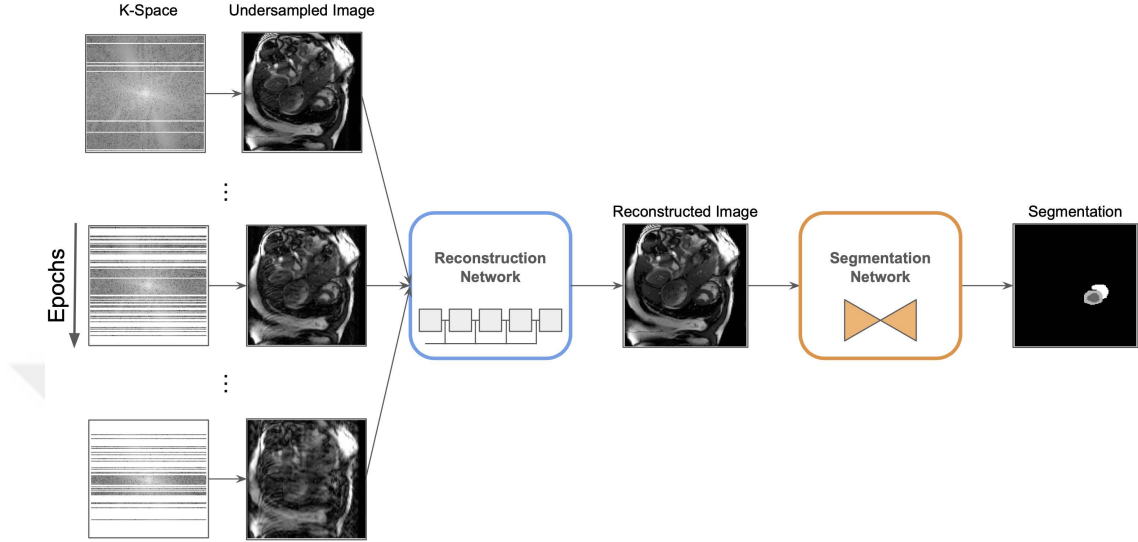


Figure 3.1: Segmentation-aware reconstruction framework. During training the network is trained end-to-end with stabilization. During inference an image is recollected from only the reconstruction network, along with an auxiliary segmentation map. Stabilization technique is illustrated by the sample inputs on the left across epochs

we perform end-to-end training of a sequential cascade of reconstruction and segmentation networks. Training of such compound networks is prone to undesirable convergence onto local minima where either network yields undesirably poor performance. During initial stages of training, the segmentation network receives an input computed by an insufficiently trained reconstruction network, so the input will contain a high degree of reconstruction artifacts. This will inevitably compromise the learning process for the segmentation stage, resulting in inaccurate segmentation maps. Note that the segmentation maps are then provided as guidance to the reconstruction network, so a vicious circle can be created where both network are compromised.

To address this critical issue, we propose to use a novel stabilization method where the undersampling rate of the acquisitions are gradually ramped up during the course of training. In the initial stages, the reconstruction network receives lightly undersampled data that is easy to reconstruct with few artifacts. Thus, the segmentation network receives as input high-quality reconstructions that will

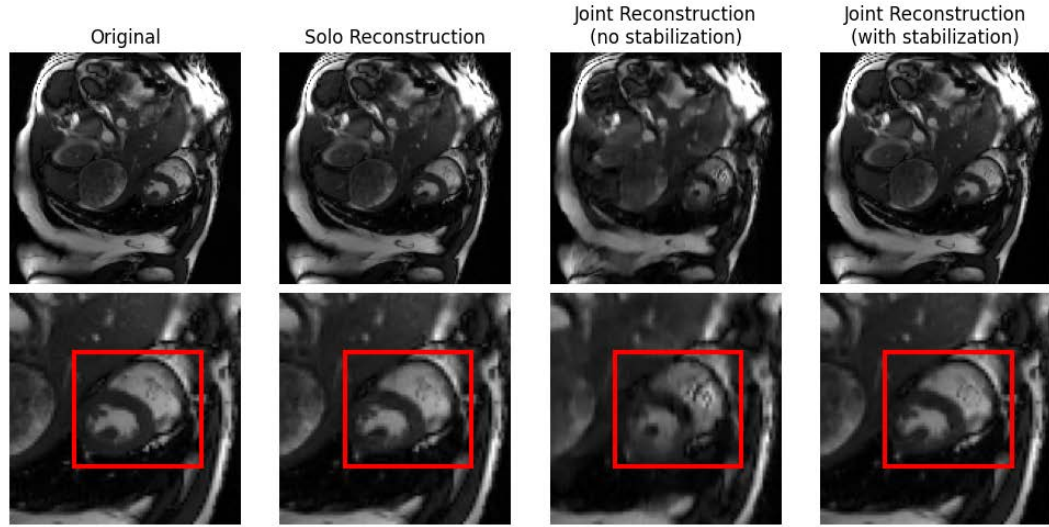
improve its learning capabilities. Once both network adapt to the instant acceleration rate, then the degree of undersampling can be elevated. Overall, the learning signals generated from the high-quality inputs in earlier epochs are propagated back to "warm-up" the model for the increasingly lower quality samples that are to come in later epochs. In particular, we impose an epoch-specific undersampling rate starting from $1 - \epsilon$ reaching to the desired undersampling rate r in an exponential manner as follows:

$$r_i = \begin{cases} (1 - \epsilon) \left(\frac{r}{1-\epsilon}\right)^{(i-1)/P} & 1 \leq i \leq P \\ r & i > P \end{cases} \quad (3.1)$$

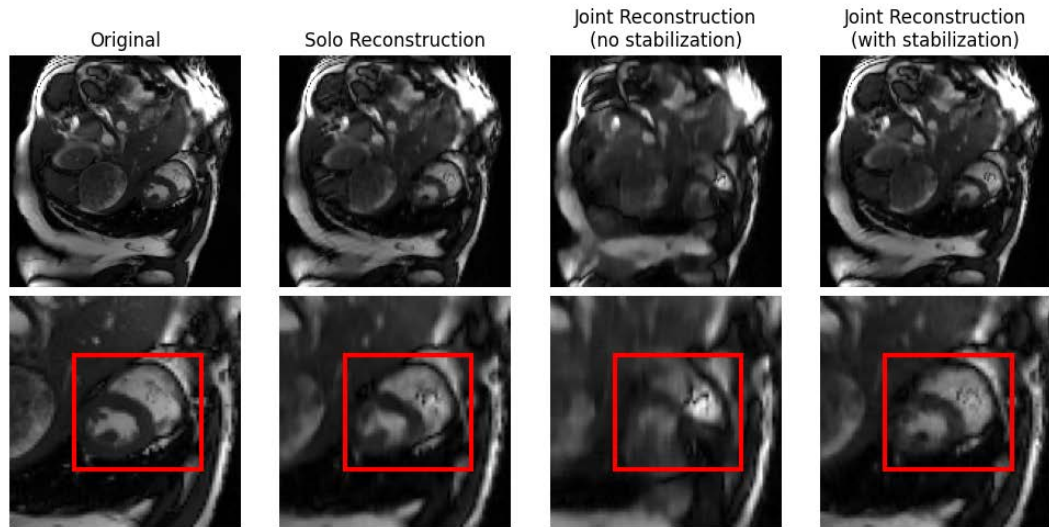
where i denotes the epoch number starting from 1, P governs the stabilization patience and ϵ is a small number close to 0. Therefore, the sub-optimal convergence problem can be mitigated with small perturbations on the task and both networks can be updated for P epochs with healthy gradients to prime the segmentation network to generate meaningful learning signal in focusing the reconstructions around the regions of interest which are dominated by the segmentation maps.

3.1.3 Model Architectures

We tested our proposed method with various configurations of reconstruction and segmentation networks. First, we used the U-Net architecture with depth of 4 and filter configuration of (32, 64, 128, 256) with kernel size 3×3 across all layers to implement the reconstruction and segmentation networks. This cascaded model with two sequential U-Nets was used to analyze the influence of end-to-end training and stabilization. For enhanced performance, we then adopted the Cascade Network for reconstruction and Multiscale Attention Network (MANet) for segmentation [44], [80]. Cascade Network follows an unrolled architecture with interleaved data consistency and regularization blocks, and progressively suppresses aliasing artifacts in reconstructions. The regularization blocks include residual connections to carry the input signal to the output and force the network to learn required the residual information [44]. In our setup, we used 6 cascades with 5



(a) x4 Acceleration



(b) x8 Acceleration

Figure 3.2: Representative reconstructions from competing techniques at (a) 4-fold, (b) 8-fold acceleration. The second row displays the localized reconstruction around the heart. The area on which the local performance calculations are taken is indicated with the red rectangle. First column displays Fourier reconstructions of fully-sampled data. In remaining columns, reconstruction with no help from the segmentation maps, joint reconstruction without stabilization and joint reconstruction with stabilization is shown respectively.

layers of 64 filters each. MANet improves upon UNet with multiscale information extraction achieved by point-wise and multiscale attention blocks [80]. Using dilated convolutions in the decoder distills multiscale information processed with the squeeze-and-excitation attention mechanism to capture dependencies among feature maps [81]. We used ResNet34 in the encoder with a depth of 4 and filter sizes (32, 64, 128, 256) for MANet [82, 83]. Finally, a multi-decoder architecture where the decoder head is split to separately perform reconstruction and segmentation is added to the experiments as an additional baseline [84]. For Multi-Decoder UNet, a common encoder of depth 4 with filter configuration (32, 64, 128, 256) is created for reconstruction and segmentation tasks, taking the undersampled images as input. The encoder is used to obtain disentangled feature representations which are then fed to the first decoder head to reconstruct the underlying image and to the second decoder to create the segmentation map.

Table 3.1: Performance comparisons of various baselines on cardiac MRI data across x4 and x8 acceleration rates. (S) suffix signifies the stabilization technique. "F" stands for focused measurements that are taken over the area of diagnostic interest. Stabilized training aids Multi-Decoder Unet and UNet $\rightarrow$ UNet in focused metrics indicating improved reconstruction quality for cardiac cavity.

x4 Acceleration						
Method	MSE	PSNR	SSIM	F-MSE	F-PSNR	F-SSIM
UNet	1.7693	26.9204	0.6247	25.5002	1.9009	0.5926
Multi-Decoder UNet	1.9693	26.3084	0.6651	25.3396	2.0209	0.6318
Multi-Decoder UNet(S)	1.7530	27.1221	0.7082	25.8338	1.7109	0.6735
UNet $\rightarrow$ UNet	1.9509	25.4176	0.5227	24.6887	2.5709	0.5488
UNet $\rightarrow$ UNet (S)	1.7474	27.0973	0.6729	26.4023	1.8174	0.6475
x8 Acceleration						
Method	MSE	PSNR	SSIM	F-MSE	F-PSNR	F-SSIM
UNet	2.2378	25.4062	0.6583	24.3128	2.7243	0.6033
Multi-Decoder UNet	2.4312	25.1512	0.6532	24.0712	2.9125	0.6219
Multi-Decoder UNet(S)	2.2441	25.2193	0.6646	24.7312	2.6217	0.6422
UNet $\rightarrow$ UNet	3.1743	23.2683	0.5491	23.0352	3.4719	0.5428
UNet $\rightarrow$ UNet (S)	2.2782	25.2931	0.6529	24.9023	2.4174	0.6375

3.1.4 Implementation Details

All networks were trained using the Adam optimizer with parameters $\beta_1 = 0.99$ and $\beta_2 = 0.999$, a learning rate of 10^{-4} and batch size of 16. Models were implemented using PyTorch library and executed on NVIDIA RTX 3090 GPUs. Experiments were conducted on fully-sampled MRI data from the public OCMR dataset containing CINE scans from 74 subjects [63]. Subjects had varying number of slices and frames, yielding a total of 183 slices, which were coil combined to simulate single-coil data acquisition. Data were split into independent training (155 slices) and test (28 slices) sets, with no subject overlap between the two sets. MRI data were retrospectively undersampled to achieve acceleration rates of 4 and 8. A Gaussian sampling density with an autocalibration region containing 8 lines was used. Magnitude images for the resulting reconstructions are used to generate segmentation map predictions. Ground-truth segmentation maps for MR images were created in-house via manual labeling with experts under the guidance of a senior radiologist. In all experiments, a composite loss function with an $\ell_1 - \ell_2$ term for the reconstruction task, and a Dice term for the segmentation task [85] is employed with equal weights. Global reconstruction quality was assessed by measuring peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and mean-squared error (MSE) between the reconstructed and ground-truth images. Local reconstruction quality was also measured via the same metrics, albeit the measurement regions containing the target tissues were selected based on the segmentation maps. Local measurements are denoted with the 'F' (for 'Focused') prefix in Table 3.2.

3.2 Experimental Results

Figure 3.2 illustrates reconstructions for a representative test subject at x4 and 8x acceleration along with the fully-sampled ground truth. PSNR, SSIM and MSE of all tested methods are presented in Tables 3.1 and 3.2 along with the localized measurements around the heart. Table 3.1 underlines the results of the

experiments done on the UNet architecture which is taken as a baseline for both the reconstruction and segmentation tasks. Additionally Multi-Decoder UNet is introduced to the experiments to compare against the segmentation-aware reconstruction framework. We see that in a stabilized setting, aided by segmentation, reconstruction of the diagnostically-relevant areas improve compared to single UNet. Multi-Decoder UNet seems to surpass the reconstruction-segmentation network performance at x4 acceleration task in MSE and SSIM measures around the regions of interest. However, when tasked with harsher undersampling rates, Table 3.1 highlights the effect of an end-to-end architecture as it yields better quality reconstructions in terms of PSNR and MSE metrics around the heart.

Table 3.2 shows, while Cascade Network performs better at solo reconstruction task in terms of global performance metrics, segmentation-aware reconstruction with stabilization outperforms competing methods in terms of localized metrics. Note that the jointly trained compound model for segmentation-aware reconstruction greatly suffers in the absence of stabilization. When supplied with high quality reconstructions, segmentation network is able to learn the mapping to create accurate segmentation maps. However, in the case of an insufficiently trained reconstruction network in the model, the segmentation network is exposed to heavy undersampling artifacts which in turn misguide the resulting segmentation output. Therefore, inaccurate segmentation information propagating into the reconstruction network damages overall quality of the reconstructions, leading to poor performance on both networks. As loss functions typically utilized for reconstruction is expressed over the entire image the learning signal coming directly from the reconstruction output largely governs the performance for the overall image in the global setting. However, such loss functions are mainly dominated by bright bone structures, diagnostically irrelevant body parts and overall low-frequency information. Back-propagating the errors from the segmentation network into the reconstruction network, indirectly emphasizes the areas of interest during training since the segmentation loss is concentrated around such diagnostically-relevant areas. Therefore, end-to-end training of a reconstruction-segmentation network creates a pseudo-attention effect to focus the efforts of the reconstruction network which in turn improves the localized performance around

regions of interest.

Table 3.2: Performance comparisons (MSE, PSNR and SSIM) on cardiac MRI data with x4 and x8 acceleration rates across experiment setups. MSE is scaled with 10^3 . (S) suffix indicates the stabilization technique. "F" stands for focused measurements that are taken over the area of diagnostic interest. Stabilized training improves the focused metrics indicating improved reconstruction quality for cardiac cavity.

x4 Acceleration						
Method	MSE	PSNR	SSIM	F-MSE	F-PSNR	F-SSIM
CascadeNet	0.6698	32.1988	0.9113	1.0342	29.1253	0.8331
CascadeNet $\rightarrow$ MANet	0.7640	31.5884	0.8563	1.1980	28.3712	0.8131
CascadeNet $\rightarrow$ MANet (S)	0.7114	31.4328	0.8991	0.9731	30.0000	0.8828
x8 Acceleration						
Method	MSE	PSNR	SSIM	F-MSE	F-PSNR	F-SSIM
CascadeNet	1.2825	28.4301	0.8259	1.7826	25.5156	0.7728
CascadeNet $\rightarrow$ MANet	1.7603	25.2837	0.7265	2.4456	23.3219	0.6673
CascadeNet $\rightarrow$ MANet (S)	1.3226	27.4895	0.8217	1.5453	26.3421	0.7931

3.3 Conclusion

Here we proposed a segmentation-aware reconstruction framework for cardiac MRI acquisitions. Experiments were conducted to systematically demonstrate the proposed method against solo reconstruction methods. As expected, solo reconstruction with global loss terms yields higher performance in global quality metrics. That said, our results clearly demonstrate that the segmentation-aware reconstruction outperform solo reconstruction in local quality metrics focused on the target ROI. Furthermore, we observe that stabilization of the acceleration rate during the course of joint network training is highly effective in mitigating convergence onto local minima with unacceptably poor reconstruction or segmentation performance.

Chapter 4

Region-of-Interest-Attentive MRI Reconstruction

Given their exceptional performance in solving inverse problems, deep neural networks have been rapidly embraced for MRI reconstruction [86, 87, 88, 41]. Reconstruction in this context involves mapping undersampled acquisitions to images that align as closely as possible with corresponding fully-sampled acquisitions. The literature has introduced several significant approaches to focus reconstruction efforts on specific regions of interest (ROIs).

One group of studies has incorporated attention mechanisms into the reconstruction model. These mechanisms selectively weigh input features based on underlying anatomical structures, refining resulting images around these locations [89, 90, 66, 42, 91, 92]. However, these studies lack explicit ROI information for network focusing. Instead, the attention mechanisms direct the network towards areas eliciting high error norms in a global loss function calculated over the entire image, which could be dominated by bright yet diagnostically irrelevant anatomical structures [93, 94, 95, 96]. It’s worth noting that diagnostically relevant ROIs often occupy smaller parts of the image and exhibit different pixel distributions [97].

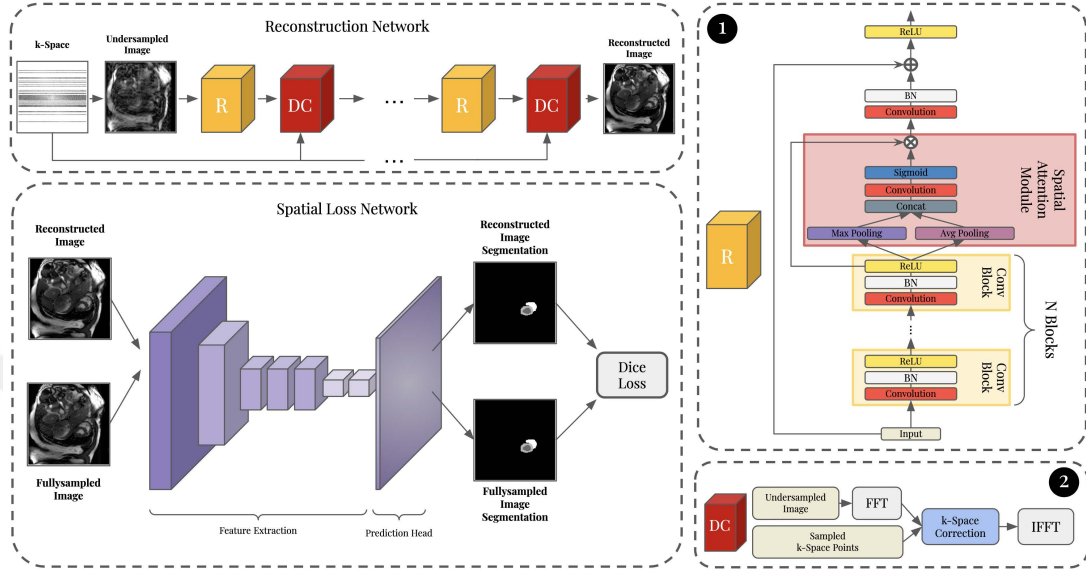


Figure 4.1: RATNet model with reconstruction network and loss function architecture. 1 depicts the regularization block architecture with the attention module addition. 2 details the data consistency scheme.

To provide explicit ROI guidance, another group of studies has developed secondary focusing networks that leverage spatial information through segmentation maps [92, 98, 97, 99]. These models divide the reconstruction process into a global reconstruction step, reconstructing the entire image with a global loss function, and a fine-tuning step. In the fine-tuning phase, the secondary network extracts spatial information from the reconstructed image and refines the reconstruction around the ROI using a second forward pass. While resulting images exhibit sharper ROI regions with refined reconstruction, this approach increases computational and time costs due to multiple reconstruction passes on the same data.

In this context, we present a novel approach to MRI reconstruction, specifically designed for regions of interest. Our approach utilizes spatial attention blocks accompanied by a dedicated pre-trained loss network, which operates alongside the global loss function to guide attention around the ROI in a single pass. This approach extends our prior work on segmentation-aware reconstruction [100]. We advance this concept with the ROI-Attentive Network (RATNet), integrating the forward mapping of a convolutional neural network (CNN) with the fine-tuning

step to learn a focusing operator for the reconstruction process. Before training the reconstruction network, a segmentation network is supplied with paired fully-sampled images and segmentation masks, allowing it to learn a segmentation prior. These parameters are then frozen for use as a spatially informed loss function. Our proposed architecture includes spatial attention blocks that bridge low-dimensional latent variables and high-dimensional image features, thereby integrating spatial information into the reconstruction step. During inference, the learned attention structure is combined with the reconstruction effort by optimizing network parameters through a weighted sum of the global loss function for image fidelity and a spatial loss function for the spatial attention blocks to learn diagnostically relevant focusing information.

The proposed method performs spatially-informed reconstruction on retrospectively undersampled MRI images derived from fully-sampled MRI acquisitions. Adapting the spatial attention blocks for the reconstruction task contributes to improved image quality specifically around diagnostically relevant ROIs.

4.1 Related Work

Deep learning-based reconstruction models currently represent the state-of-the-art in MRI, with a diverse array of successful architectures reported. These encompass vanilla CNNs [86], variational networks [87], physics-guided unrolled CNN networks [101], generative adversarial networks [102], and more recently, vision transformers [103]. However, conventional MRI reconstruction methods are typically trained to minimize loss norms across entire images. In clinical practice, certain sub-regions of the image hold greater diagnostic relevance, containing crucial details pertaining to underlying anatomy. To enhance the applicability of MRI acceleration in such cases, a common strategy has been to introduce focusing mechanisms into reconstruction models, aiming to selectively emphasize the reconstruction effort within specific ROIs.

While convolutional neural networks achieve high performance due to their

effective extraction of local image features, the convolution operator possesses a limited influence range or local receptive field. This limitation can be problematic when synthesizing signals from a wide spectrum of inputs. Expanding the receptive field can be achieved by deepening the convolutional neural network or utilizing a hierarchical network architecture. However, a deep stack of convolutional operations not only becomes computationally prohibitive for high-resolution images, but also introduces optimization challenges when long-range dependencies are propagated across numerous layers.

Alternatively, several models have integrated self-attention mechanisms [66, 104, 91]. These mechanisms allow signal extraction at a given position by considering all positions within the same image, rather than solely neighboring ones [105]. This approach permits the model to selectively enhance the reconstruction of sub-regions linked to correlated features. However, it doesn't guarantee that the focused region is clinically significant or relevant.

In contrast, spatially informed models are trained using explicit spatial information provided by proxy tasks such as segmentation [92, 97, 99]. These tasks offer guidance to reconstruction models by amplifying gradients on ROIs through a proxy module. This pseudo-attention effect is achieved by defining a spatially informed loss, such as a segmentation loss, that amplifies the global loss output for relevant tissues using an ROI extractor network. This is followed by an additional finetuning pass over the reconstructed image to refine the ROI features [97]. Although the finetuning approach enhances reconstruction performance around diagnostically relevant regions, the reconstruction model isn't directly conditioned on a semantically meaningful sub-region. Besides the added computational complexity of a second finetuning pass, the reconstruction remains unaware of the importance and unique features of the focused ROI due to the decoupled nature of the ROI extraction process. Consequently, optimization becomes challenging and susceptibility to artifacts at the edges arises due to sharp binary masks modulating gradients within the ROI.

In order to endow the reconstruction operator with attending to diagnostically relevant sub-regions of the image, a different approach is to give the model

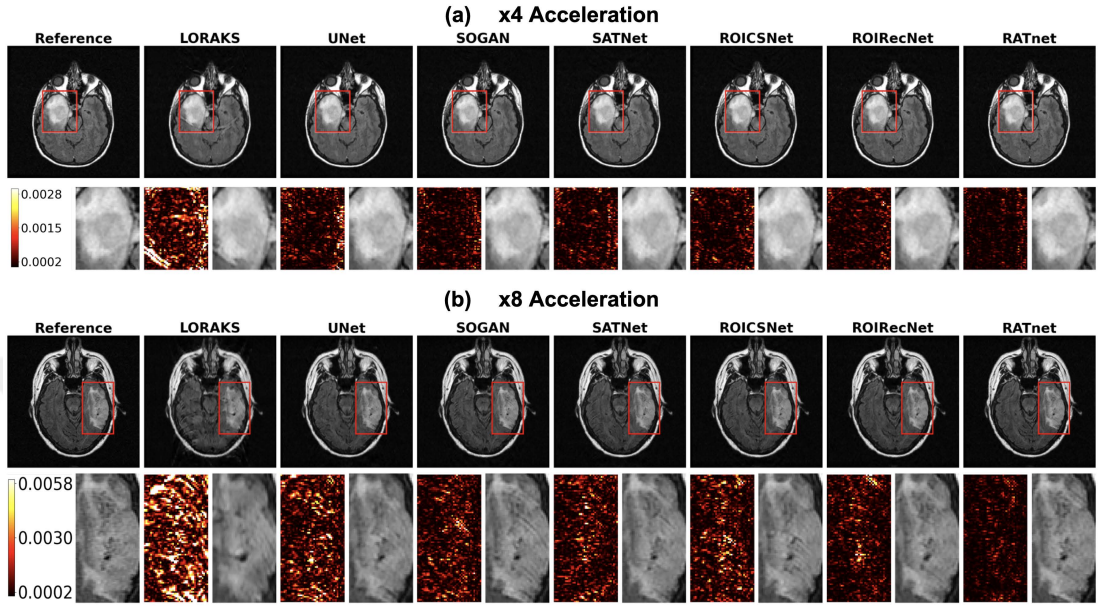


Figure 4.2: Reconstructions of representative brain magnitude acquisitions from TCGA-LGG dataset at 4-fold and 8-fold acceleration rates are shown for self-attention based methods (SATNet, SOGAN), focusing based methods (ROIRecNet, ROICSNet), proposed RATNet along with reference image. Zoom-in display windows are added to aid visualization of performance differences around diagnostically relevant sub-regions.

necessary attentive complexity to learn from the injected spatial information. Incorporating a spatial attention module extracts distinctive representations of image features and uses a soft attention map to modulate the features to emphasize spatial locations which elicit high error norm [106]. Therefore, it is crucial to supply the training procedure with spatially dominant loss function in order to utilize these modules effectively [107].

4.2 Theory

We introduce a reconstruction method for accelerated MRI which selectively focuses around the ROI in the image. Unlike previous methods, this model includes spatial attention modules to capture diagnostically relevant area features. Before training a fully convolutional network (FCN) tasked with segmentation which

learns to locate ROI to be used as a spatially-informed loss function as a frozen model for the training the reconstruction network. The reconstruction network learns to focus the limited reconstruction effort around the diagnostically relevant areas of the image. First, we provide an overview the inverse problem formulation in accelerated MRI. Then, we describe the fundamental building blocks of RATNet which can also be seen in Fig. 4.1.

4.2.1 Accelerated MRI Reconstruction

Accelerated MRI reconstruction seeks to accurately reconstruct MR images from highly undersampled k-space measurements to speed up scans:

$$y_s = F_u x + e \quad (4.1)$$

where $x \in \mathbb{C}^{M \times 1}$ is the underlying M-dimensional vectorized MR image and $y \in \mathbb{C}^{N \times 1}$ ($N \ll M$) is the undersampled N-dimensional k-space measurements with much lower dimension. Here $F_u \in \mathbb{C}^{N \times M}$ in the data fidelity term denotes the undersampled Fourier encoding matrix. Since undersampled acquisitions violate the Nyquist condition, Eq. 4.1 is underdetermined and therefore its solutions benefits from prior information. Prior knowledge can be incorporated as a regularization term:

$$\hat{x} = \operatorname{argmin}_x \|F_u x - y_s\|_2^2 + \mathcal{R}(x) \quad (4.2)$$

Here $\mathcal{R}$ is the regularization term which encodes prior information on the desired property of image x . The optimization of objective in Eq. 4.2 aims to minimize the data fidelity loss and regularization loss simultaneously. Conventionally, the sparse priors are used to constrain the ill-posed problem like ℓ_1 and ℓ_2 norms.

From the objective in Eq. 4.2, it is worth noting that all the pixels in the MR image x are typically weighted equally, regardless of the specific tissues. Hence,

conventional MRI reconstruction methods lack the ability to provide better reconstruction quality around a region of interest (ROI). However, in clinical practice scenarios, different tissues in the same MRI represent different biological information and for efficient clinical practice it is necessary to focus on certain tissues over others.

4.2.2 Reconstruction Network

RATNet adopts a deep neural network model for MRI reconstruction based on cascaded CNNs. The input of the reconstruction network is the zero filled MRI $F_u^H y$ obtained by padding zeros in the undersampled positions in the k-space and taking an inverse 2D FFT of the Fourier coefficients. The reconstructed MRIs are output to approximate the fully-sampled ground truth.

Reconstruction stacks N convolutional blocks with a global residual shortcut to help stabilize training. Overall, M blocks are cascaded together with a differentiable data consistency block proposed by Schlemper et. al. interleaved in between to enforce sampled data fidelity in between the unrolled reconstruction steps that progressively suppress aliasing artifacts [44]:

$$\text{DC}(\mathbf{z}) = \begin{cases} F\mathbf{z}(i) & \text{if } i \in \Omega \\ \frac{F\mathbf{z}(i) + \mu F\mathbf{z}_0(i)}{1 + \mu} & \text{if } i \notin \Omega \end{cases} \quad (4.3)$$

where F is the Fourier Transform operator, $\mathbf{z}$ is the intermediate reconstruction, Ω denotes the already-sampled k-space locations and $\mathbf{z}_0$ is the zero-filled complex-valued image. Here the parameter μ is inversely proportional to the noise power in the acquisitions. It can be seen that in the limit $\mu \rightarrow \infty$ the data-consistency operator converges to strict data fidelity.

As illustrated in Fig. 4.1, in each of the M blocks a spatial attention module is integrated at the output of the N^{th} block to extract a spatial attention map from the low-level latent image features. The input to the attention module is then modulated with the resulting attention map to emphasize sub-regions which

elicit higher error norms for the given loss function.

4.2.3 Spatial Attention Module

Discriminant feature representations are essential for scene understanding, which could be obtained by capturing long-range contextual information. In order to model rich contextual relationships over local features, we introduce a spatial attention module into the last reconstruction layer of each block of the network [106]. The spatial attention module encodes a wider range of contextual information into local features, thus enhancing their representation capability.

The module initially receives the refined image features to generate a spatial attention map using the inter-spatial relationship of features. In order to generate an attention map, average-pooling and max-pooling operators are applied along the channel axis in parallel and then the resulting activations are concatenated along the channel axis to generate an efficient feature descriptor. This descriptor is then convolved by a convolution layer with receptivity field of (7×7) to produce spatial attention features. Sigmoid activation is used to map these features into probabilities, creating the desired spatial attention map. The input features are then modulated using the resulting attention map to focus the activations around the ROI. Overall the operator can be described as:

$$SAM(F) = F \times \sigma(\text{conv}([F_{avg}; F_{max}])) \quad (4.4)$$

where $F \in \mathbb{R}^{C \times H \times W}$ is the input features with C channels, $F_{avg} \in \mathbb{R}^{1 \times H \times W}$ and $F_{max} \in \mathbb{R}^{1 \times H \times W}$ are the average and max pooled features respectively. σ denotes the sigmoid activation function.

4.2.4 Spatial Loss Network

The spatial loss network adopts an fully convolutional network (FCN) architecture [108]. The encoder components extract features at different scales using

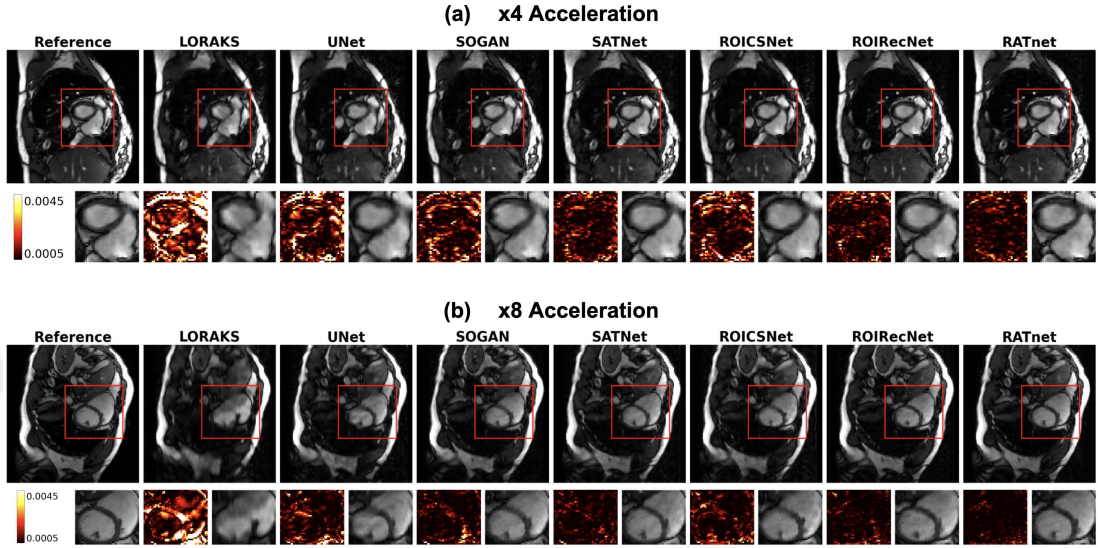


Figure 4.3: Reconstructions of representative coil-combined cardiac acquisitions at 4-fold and 8-fold acceleration rates are shown for the classical CS based LORAKS method, self-attention based methods (SATNet, SOGAN), focusing based methods (ROIRecNet, ROICSNet), proposed RATNet along with reference image. Zoom-in display windows are added to aid visualization of performance differences around diagnostically relevant sub-regions.

convolution and pooling, while the classification component is used to transform the image from feature space back to image space.

We task the spatial loss network with segmentation using fully-sampled MR images paired with corresponding segmentation labels. Dice loss function is employed to differentiate between classes [109]. After the training the weights are frozen to be used as a spatially informed loss function.

4.2.5 Learning Procedures

The reconstruction and the segmentation networks described above are used as building blocks for our proposed architecture which we show in Fig. 4.1. The undersampled k-space measurements in the training datasets are fed into the reconstruction network to obtain the reconstruction for the MRI. This reconstruction, along with the fully-sampled data pair, is fed into a global normalized

$\ell_1 - \ell_2$ Loss function to refine global features of the underlying image:

$$\mathcal{L}_{global}(x, \hat{x}) = \frac{1}{2} \left(\frac{\|x - \hat{x}\|_1}{\|x\|_1} + \frac{\|x - \hat{x}\|_2}{\|x\|_2} \right) \quad (4.5)$$

where $x \in \mathbb{C}^{M \times 1}$ is the vectorized MR image and $\hat{x} \in \mathbb{C}^{M \times 1}$ is the resulting reconstruction.

Additionally the reconstruction and the fully-sampled image are also fed into the pre-trained segmentation network which is utilized to produce segmentation features dominant in spatial information around the region of interest. The resulting features are fed into a Dice loss function for training the spatial attention module of the reconstruction network:

$$\begin{aligned} \mathcal{L}_{spatial}(x, \hat{x}) &= Dice(f_{segment}(x|\theta), f_{segment}(\hat{x}|\theta)) \\ &= 1 - \frac{2m\hat{m} + 1}{m + \hat{m} + 1} \end{aligned} \quad (4.6)$$

where m and $\hat{m}$ are the resulting segmentation masks and $f_{segment}(\cdot|\theta)$ is the pre-trained segmentation network parameterized with fixed vector θ . The high-level semantic information from the spatial loss network can guide the reconstruction to focus on the ROI. Note that the reconstruction network is trained to focus the reconstruction effort on the region of interest. The resulting loss function can be expressed as:

$$\mathcal{L}(x, \hat{x}) = \lambda_g \mathcal{L}_{global}(x, \hat{x}) + \lambda_s \mathcal{L}_{spatial}(x, \hat{x}) \quad (4.7)$$

using λ_g and λ_s as mixing hyper-parameters for the global and spatial loss functions respectively.

Compared with the entire image, ROIs usually contain less variation, which can significantly simplify the function mapping of the deep network. As detailed in experimental results section, fewer errors in ROI at the expense of more errors in the uninteresting regions can result from this approach.

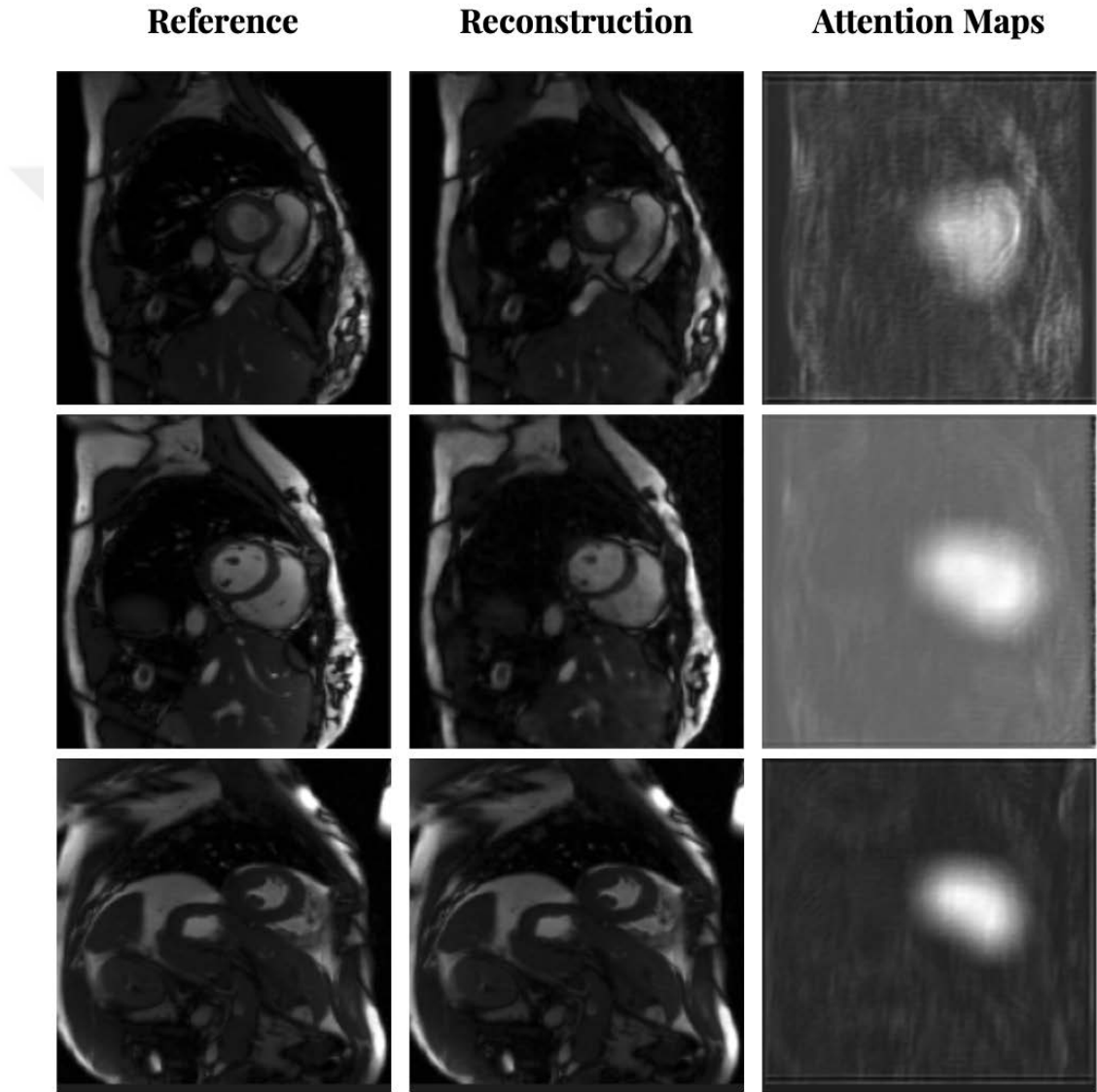


Figure 4.4: Attention maps in RATNet for three different test samples from OCMR dataset in coil-combined setup. Map activations are taken from the attention module of the last regularization block.

Experiment	PSNR	SSIM (%)	F-PSNR	F-SSIM (%)
OCMR Dataset (x4)				
LORAKS	29.3	79.1	27.3	73.7
U-Net	35.3	87.7	31.8	84.0
SATNet	39.2	96.3	34.8	91.0
SOGAN	38.5	95.2	34.7	91.2
ROIRecNet	38.3	93.9	35.3	92.0
ROICSNet	38.0	95.0	34.2	90.7
RATNet	38.2	94.6	36.9	92.9
TCGA-LGG Dataset (x4)				
LORAKS	28.5	77.4	26.1	74.6
U-Net	32.5	83.8	29.2	80.4
SATNet	36.1	92.0	31.8	86.9
SOGAN	35.4	91.0	31.9	87.3
ROIRecNet	35.2	89.8	32.5	87.9
ROICSNet	34.9	90.8	31.5	86.9
RATNet	34.9	90.5	33.9	89.0

Table 4.1: Reconstruction performance of competing methods on coil-combined OCMR and TCGA-LGG datasets at 4-fold acceleration. "F" prefix stands for the focused measurements taken around the diagnostically relevant sub-regions of the image.

4.3 Methods

4.3.1 Network Architecture

This section provides details of the utilized network architectures in our framework. First, we give details of the reconstruction network, then we explain the spatial attention modules and the spatial loss network.

4.3.1.1 Reconstruction Network

The reconstruction network is a cascaded architecture where the aliasing artifacts are progressively suppressed in an unrolled setting. Each cascade consists of a regularization block and a data consistency block. The regularization blocks include

residual connections to carry the input signal to the output and force the network to learn required the residual information. In our setup, we used 6 cascades with 5 layers of 64 filters. Each layer follows convolution-batch normalization-rectified linear unit setup with (3×3) kernel size. The output of the last layer is connected to a spatial attention unit to selectively emphasize the ROI. At the last layer the attended activations are fed into a convolutional block mapping the feature space into a complex residual image represented by the imaginary and real channels at the output. Lastly a rectified linear unit activation produces the final reconstruction of the block from the summation of the input and the obtained residual information.

Regularization block operates over the entire image to generate an intermediate reconstruction. In our setup, 6 of these blocks are concatenated to simulate the unrolled reconstruction. Data consistency blocks are interleaved with the regularization blocks to keep the signal in already sampled k-space locations consistent with the measured data. This is achieved by taking a Fast Fourier Transform (FFT) of the intermediate reconstruction to map the complex image to k-space measurements. At this stage the measurements at already sampled locations are replaced with the sampled values to be mapped back to a complex image which is more faithful to the sampled data. The output of the reconstruction network was set to (144×144) for both datasets.

4.3.1.2 Spatial Attention Modules

Spatial attention modules are plugged into the regularization blocks following the last convolutional reconstruction layer. We feed the refined intermediate reconstruction features $F \in \mathbb{R}^{C \times H \times W}$ to the spatial attention modules to modulate the features with the attention maps to emphasize diagnostically relevant sub-regions in the $[H \times W]$ features. In total, M attention blocks are employed in the overall network.

Experiment	PSNR	SSIM (%)	F-PSNR	F-SSIM (%)
OCMR Dataset (x8)				
LORAKS	27.5	76.9	26.0	72.0
U-Net	32.5	81.3	29.3	78.0
SATNet	37.2	93.5	33.1	87.4
SOGAN	36.4	92.6	32.8	88.1
ROIRecNet	36.0	91.3	33.8	89.4
ROICSNet	35.6	92.4	32.7	89.0
RATNet	36.1	92.0	35.1	90.5
TCGA-LGG Dataset (x8)				
LORAKS	26.6	74.2	25.2	70.1
U-Net	29.9	77.8	27.1	74.6
SATNet	34.3	89.3	30.1	84.2
SOGAN	33.5	88.5	30.5	85.0
ROIRecNet	3.1	87.3	31.1	85.3
ROICSNet	32.9	88.1	30.6	85.7
RATNet	33.3	88.0	32.4	86.7

Table 4.2: Reconstruction performance of competing methods on coil-combined OCMR and TCGA-LGG datasets at 8-fold acceleration. "F" prefix stands for the focused measurements taken around the diagnostically relevant sub-regions of the image.

4.3.1.3 Spatial Loss Network

While a general network computes a general nonlinear function, a net with only cascaded convolutional layers computes a nonlinear filter that combines layers of the feature hierarchy and refines the spatial precision of the output [108]. We task such a network with segmentation in order to learn features that are dominant in spatial information. The network is supplied with fully-sampled images and ground truth segmentation map pairs and optimized to reduce Dice Loss. The segmentation map is produced by progressively filtering and resampling the complex 2-channel input image. The network employs a ResNet-101 backbone for the convolutional feature extracting base, and diluted convolution for mapping the segmentation features to a segmentation map [110]. After training the segmentation model weights are frozen and the model is used as a spatially aware loss function during the reconstruction training task .

4.3.2 Methods of Comparison

RATNet was comparatively demonstrated against state-of-the-art techniques based on externally fine-tuned attention mechanisms and built-in attention layers as well as a traditional reconstruction method. For each technique, hyperparameter optimization was performed via cross-validation on a three-way split of subjects. Optimization was performed for number of epochs and weights for regularization terms based on performance on the validation set. All models were implemented using PyTorch library and executed on NVIDIA RTX 3090 GPUs. Experiments were conducted on fully-sampled MRI data from the public OCMR and TCGA-LGG datasets containing cine scans and brain scans with FLAIR sequence respectively [63, 111]. Data were split into independent training, test and validation sets, with no subject overlap between the three sets. In all experiments, a composite loss function with an $\ell_1 - \ell_2$ term for the reconstruction task, and a Dice Loss for the segmentation task is employed [109].

Global reconstruction quality was assessed by measuring peak signal-to-noise ratio (PSNR), structural similarity index (SSIM) and mean-squared error (MSE) between the reconstructed and ground-truth images. Local reconstruction quality was also measured via the same metrics, albeit the measurement regions containing the target tissues were selected based on the segmentation maps.

RATNet: RATNet was trained on retrospectively undersampled MR images. In the OCMR dataset, the model was trained to map onto complex images with real and imaginary channel outputs. Training was performed via the Adam optimizer with $\beta_1 = 0.6$, $\beta_2 = 0.999$, batch size of 16 and a learning rate of 0.0001 [112]. An early stopping criteria was set to terminate training if the validation performance does not improve in 50 consecutive epochs. A learning rate scheduling scheme was employed to reduce the learning rate of by a factor of 10 when the validation performance does not improve in 15 consecutive epochs. Network weights were randomly initialized using a standard normal distribution. For both OCMR and TCGA-LGG datasets $\lambda_g = 0.6$ and $\lambda_s = 0.2$ mixing parameters were selected. For the spatial loss network, a learning rate of 0.001, batch size of 16 and the

same early stopping and learning rate scheduling schemes were employed.

LORAKS: A traditional parallel-imaging reconstruction based on low-rank modeling of local k-space neighborhoods was performed via libraries in the LORAKS V2.1 toolbox [113]. Here, an auto-calibrated reconstruction was performed where the structured low-rank matrix was formed based on limited image support assumption [113]. Accordingly, the k-space neighborhood radius and the rank of the resultant matrix were selected via cross-validation as (2,8) for OCMR dataset.

U-Net: A vanilla reconstruction network with U-Net architecture was trained with an encode and decoder depth of 4 with channels [32, 64, 128, 256] respectively [114]. Skip connections are incorporated to the architecture for residual learning. For training, the optimization hyper-parameters were set identical to RATNet setup.

SATNet: A volumetric hierarchical deep residual convolutional neural network with self-attention modules integrated in convolutional blocks (SATNet) was trained using paired ground-truth and undersampled acquisitions. Network architecture and loss functions were adopted from [66]. Training was performed via the Adam optimizer, with $\beta_1 = 0.9$, $\beta_2 = 0.99$, learning rate of 0.0001 with the previously explained early stopping and learning rate scheduling strategy [112]. Network weights were randomly initialized using a standard normal distribution.

SOGAN: A spatial orthogonal attention generative adversarial network was trained on undersampled and fully-sampled data pairs. The network architecture and training strategy were adopted from [91]. Network weights were randomly initialized using a standard normal distribution. A batch size of 2 was used to train this architecture due to additional computational complexity that the self-attention modules introduce.

ROIRecNet: Focusing the reconstruction effort around a region of interest was also implemented using a finetuning approach by ROIRecNet [97]. The network architecture and the training strategy were adopted from [97]. Optimization of network weights was achieved using Adam optimizer with $\beta_1 = 0.9$, $\beta_2 =$

0.99, a learning rate of 0.00005 [112]. Network weights were randomly initialized using a standard normal distribution. A secondary segmentation network is also trained as a finetuning proxy network. A masked mean squared error loss on the externally predicted ROI was utilized to finetune the initial reconstruction.

ROICSNet: A compressed sensing inspired reconstruction network was trained on undersampled images with fully-sampled data pairs. The network architecture, hyper-parameters and training strategy was adopted from [99] for fair comparison against RATNet.

4.3.3 Datasets

Reconstruction experiments were conducted using multi-coil cardiac MRI dataset from OCMR and coil-combined brain MRI data from The Cancer Genome Atlas (TCGA) lower-grade glioma (LGG) dataset [63, 111]. For TCGA-LGG dataset FLAIR sequence were considered. While raw acquisitions provided in OCMR dataset allowed for a drawing classical CS baseline with LORAKS, TCGA-LGG dataset is presented in real magnitude image format, prohibiting a meaningful comparison. Therefore experiments involving LORAKS method was skipped for this dataset. OCMR dataset was coil-combined for single-coil reconstruction experiments using estimated coil-sensitivity maps produced by Walsh method with smoothing parameter of 5. In the OCMR dataset, 54 subjects were used for training, 5 for validation and 10 for testing. In total 152 dynamic scans of 1899 frames were included. The results are presented for both coil-combined and multi-coil setups. For the TCGA-LGG dataset, scans from 88 subjects used for training, 7 for validation and 15 for testing, totaling 3929 frames. For both datasets, the data from multiple sites are included with no common protocol and subject selection and splitting was done sequentially. MRI data were retrospectively undersampled to achieve acceleration rates of 4 and 8. A Gaussian sampling density with an auto-calibration region containing 8 lines was used. Ground-truth segmentation maps for OCMR cardiac MR images were created in-house via manual labeling with experts under the guidance of a senior radiologist.

OCMR Dataset				
Experiment	PSNR	SSIM (%)	F-PSNR	F-SSIM (%)
x4				
LORAKS	32.2	81.6	28.2	77.3
U-Net	32.2	83.3	28.9	79.9
SATNet	35.0	91.5	31.0	86.5
SOGAN	35.3	90.7	31.1	86.0
ROIRecNet	34.9	89.5	32.3	88.2
ROICSNet	34.7	90.4	31.5	86.9
RATNet	35.3	90.2	33.6	88.6
x8				
LORAKS	29.2	73.3	23.6	68.5
U-Net	29.7	77.4	26.7	74.3
SATNet	33.1	88.3	29.6	83.4
SOGAN	33.8	89.0	29.8	84.0
ROIRecNet	33.0	86.9	30.7	84.9
ROICSNet	32.6	88.0	30.2	85.5
RATNet	33.2	87.5	31.8	86.1

Table 4.3: Reconstruction performance of competing methods on multi-coil OCMR dataset at 4-fold and 8-fold accelerations. "F" prefix stands for the focused measurements taken around the diagnostically relevant sub-regions of the image.

4.3.4 Evaluation Metrics

To assess reconstruction quality, quantitative comparisons were performed against reference images Fourier-reconstructed from fully-sampled acquisitions. The fully-sampled points were normalized to one in magnitude prior to retrospective undersampling. Peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) were calculated between the reconstructed and reference images with the definitions. Aside from global reconstructions, local reconstruction quality was also measured via the same metrics around the target tissues selected based on the segmentation maps. Local measurements are denoted with the 'F' (for 'Focused') prefix in results. Precisely:

$$PNSR(y, \hat{y} | \Pi) = 10 \log_{10} \left(\frac{|\Pi|}{\sum_{i=1}^{|\Pi|} (y_{\Pi_i} - \hat{y}_{\Pi_i})^2} \right) \quad (4.8)$$

$$SSIM(y, \hat{y} | \Pi) = \frac{(2\mu_{y|\Pi}\mu_{\hat{y}|\Pi} + c_1)(2\sigma_{y\hat{y}|\Pi} + c_2)}{(\mu_{y|\Pi}^2\mu_{\hat{y}|\Pi}^2 + c_1)(\sigma_{y|\Pi}^2\sigma_{\hat{y}|\Pi}^2 + c_2)} \quad (4.9)$$

where $y, \hat{y} \in \mathbb{R}^{M \times 1}$ are the reference and estimated vectorized MR images respectively. μ_y, σ_y and $\mu_{\hat{y}}, \sigma_{\hat{y}}$ denotes the pixel sample means and covariances of y and $\hat{y}$. Similarly $\sigma_{y\hat{y}}$ is the covariance of y and $\hat{y}$. $c_1 = 0.0001$ and $c_2 = 0.0009$ are coefficients to stabilize the division with weak denominator. Here Π denotes the set of image locations on which the metrics are calculated. For a global measurement $|\Pi| = M$ can be set. In order to get localized measurements like F-PSNR or F-SSIM, Π can be populated with the locations of the ROI.

4.4 Experimental Results

In this section we provide results of our tests on OCMR (cardiac MRI) and TCGA-LGG (brain MRI) datasets. First, single-coil results are provided for both datasets. Then, we provide multi-coil reconstruction results on OCMR data. Finally we provide an ablation study of the proposed method on both datasets.

4.4.1 Single-Coil Reconstruction

In order to assess the influence of proposed architecture, we examined performance on coil-combined complex cine images and magnitude brain images across 4-folds and 8-folds retrospective undersampling rates (Tables 4.1 - 4.2). RATNet out-performs all competing methods on focused metrics measured around ROIs and, although not the best performing model, it is a close competitor on global metrics. On average across datasets and undersampling rates, RATNet yields 1.6 dB focused PSNR, 1.25% focused SSIM improvement over the second-best reconstruction method, demonstrating the efficacy of RATNet in capturing ROI specific features. Tables 4.1 - 4.2 show that RATNet achieves a PSNR drop of 3.18% and 3.69% for going from global to focused metrics in 4 and 8-fold

acceleration respectively. These results are 5.22% and 4.66% lower the closest competitor which is ROIRecNet with 8.42% and 8.35% drops, indicating a better global/focused reconstruction quality trade-off.

4.4.2 Multi-Coil Reconstruction

Differing from single-coil experiments, multi-coil reconstruction experiments were conducted on only the OCMR dataset (Table 4.3). RATNet outperforms all competing methods similarly. On average, across undersampling rates, RATNet outputs 1.7 dB focused PSNR, 1.33% focused SSIM improvement over the second-best reconstruction method. Table 4.3 shows that RATNet achieves 4.94% and 4.05% PSNR drop around the ROIs for 4 and 8-fold acceleration rates. The closest competitor in these experiments is also ROIRecNet with 8.82% and 9% decrease in PSNR.

4.4.3 Ablation Study

Lastly, we conducted an ablation study to assess the contributions of spatial attention layers and spatial loss function of the proposed method to the reconstruction performance. To do this, we compared RATNet against a conventional global reconstruction network with no spatial attention or loss function, a variant with only the attention layers but not the loss function and another variant that uses the spatial loss function without the attention layers. All ablation experiments were done in a single-coil setup. The same performance metrics were utilized in the reported evaluation in Table 4.4. This table demonstrates that RATNet outperforms all variants in focused metrics consistently across datasets. These results indicate the importance of the embedded attention module and spatially-informed loss function in RATNet to MRI reconstruction performance over ROI.

Experiment	PSNR	SSIM (%)	F-PSNR	F-SSIM (%)
OCMR Dataset (x4)				
L1L2	39.4	95.4	34.9	91.8
L1L2 + A	40.8	96.5	35.9	92.1
L1L2 + SL	36.2	92.1	33.1	91.4
L1L2 + A + SL	38.8	94.6	36.9	92.9
TCGA-LGG Dataset (x4)				
L1L2	35.9	91.7	32.7	88.2
L1L2 + A	37.0	92.7	33.4	88.4
L1L2 + SL	32.7	88.3	30.6	87.7
L1L2 + A + SL	34.9	90.5	33.9	89.0

Table 4.4: Ablation study on proposed method. "A" stands for spatial attention modules and "SL" stands for the use of spatial loss function. Illustrates the effect of each sub-component on the reconstruction quality. Performance metrics around the diagnostically relevant sub-regions of the image are indicated via the prefix "F" for "focused".

4.5 Discussion

In clinical scenarios, underlying anatomical structures in MR images carry diagnostically relevant information heterogeneously. Recent studies have proposed approaches for focusing the reconstruction effort of accelerated MRI scans around regions-of-interest using self-attention modules and proxy networks [97, 66, 91, 99]. Yet, self-attention modules are not explicitly informed on the relevant sub-regions, thus are not guaranteed to correlate ROI pixels with each other. Proxy network approach decouples the ROI finding operation from the reconstruction and results in increased time and computational complexity. In contrast, RATNet couples the reconstruction step with a spatial attention in order to guide the reconstruction efforts around the ROIs, resulting in well-informed coupled ROI-attentive MRI reconstructions with single forward pass during test time. Experiments on datasets with segmentation information around the ROIs is available demonstrate that RATNet yields superior performance against competing methods across different anatomies and undersampling rates. Our demonstrations clearly indicate the superiority of the proposed method over self-supervised approaches (SATNet, SOGAN), fine-tuning based ROI reconstructions (ROIRecNet, ROICSNet) and conventional applications (LORAKS, U-Net). RATNet

offers slightly worse performance on global reconstruction quality metrics while boasting a superior performance against all of the competing methods over ROI reconstruction focused quality metrics.

A practical concern for MRI reconstruction is the computational cost of training and inference. Here, we considered convolutional self-attention networks and fine-tuning proxy network approaches. Previous self-attention methods additionally compute the correlation of each feature location with all of the other feature locations globally [66, 115, 116]. This architecture aims to activate correlated neurons together to better distribute reconstruction effort around high error norm inducing areas. However, use of global loss functions in optimization of such networks does not permit self-attention layers to learn discriminative features for benefit of the clinical scenarios. Additionally, fine-tuning approaches introduce the RATNet trains a segmentation model to be used as a loss function which explicitly informs the spatial attention modules which are superior in lack of computationally complexity requirements to the self-attention layers. Thus, once trained, RATNet offers a cheaper inference procedure which inherently looks for diagnostically relevant areas of the image in a single pass.

The effect of spatial loss function is visualized in Fig. 4.4. Due the signal coming from still prominent global loss function, spatial attention module can be seen to emphasize global structures which elicit high error norms. However, brightly activated ROIs are reinforced by the spatial loss function. Due the error signal dominated by the ROI, the attention module selectively highlights the areas that contribute to this error largest. Without the spatial loss function attention is still useful, due to the heterogeneous distribution of the global learning signal. Furthermore, spatial loss function, without an attention mechanism to utilize it, hinders the model performance due to irregular loss flowing into the reconstruction pipeline.

Several lines of development can be pursued for the proposed technique. First, the dependency on high quality human-labeled segmentation datasets for the training of spatial loss function hinders the training of larger and better models while limiting the applicability of the approach. Pixel clustering and assignment

based self-supervised segmentation approaches can alleviate the need for human annotation and increase the efficiency of this approach [117]. Second, RATNet learns a coil-combined MRI prior and subsequently incorporates coil-sensitivity information during reconstruction. This is achieved by passing coil-sensitivity maps across regularization steps for the output to be back-projected onto individual coils. Producing multi-coil output by consistent coil priors can generate higher fidelity images. Third, RATNet uses a full segmentation network as a loss function which introduces inherent biasing information of segmentation training on the reconstruction process. Employing direct approaches for optimization of the attention maps using gradient methods can consolidate the imaging prior and prevent unpredictable artifacts [118, 119].

4.6 Conclusion

Here we introduce a novel ROI-attentive MRI reconstruction based on a cascaded convolutional network modified with spatial attention blocks. RATNet, leverages spatially dominant learning signal coming from deployed segmentation based spatial loss function to guide the attention modules in unrolled regularization blocks to highlight relevant ROIs in clinical and diagnostic practice. Benefits of RATNet over state-of-the-art attention and fine-tuning approaches were demonstrated in cardiac and brain MRI in both single-coil and multi-coil scenarios. Obtained inherent ROI guidance and computational efficiency render RATNet promising candidate for high-performance accelerated MRI.

Chapter 5

Conclusion

Three main contributions were described in this thesis. In the second chapter detailing the first contribution, a self-supervision approach was proposed for deep networks in dynamic cardiac MRI reconstruction. A systematic assessment was conducted between model complexity and performance in self-supervised reconstruction to identify that simpler models offer learning benefits in this setting, unlike in supervised settings where more complex models can elicit performance benefits.

In the third chapter detailing the second contribution, a segmentation-aware framework was proposed for cardiac MRI reconstruction. Segmentation guided reconstructions were compared against solo reconstructions with global loss terms. Systematic evaluations indicates that guided reconstructions improved performance metrics locally near target regions at the expense of moderately lower global performance metrics. Stabilization of the learning procedure was also observed to help maintain performance in joint network training. Future work is warranted to assess strategies that can help build a network that produces a variable output that can focus on optimizing image quality in user-selected regions of interest.

In the fourth chapter detailing the third contribution, an ROI-attentive MRI

reconstruction based on a cascaded convolutional network was proposed. The introduced model segmentation-driven spatial loss functions to guide attention modules embedded within an unrolled network architecture that reconstructed images. Comprehensive demonstrations indicates that the proposed method offers higher performance than many baselines in terms of both local image quality and computational efficiency.



Bibliography

- [1] M. Chung, I. J. Dahabreh, N. Hadar, S. J. Ratichek, J. M. Gaylor, T. A. Trikalinos, and J. Lau, “Emerging mri technologies for imaging musculoskeletal disorders under loading stress,” *Comparative Effectiveness Technical Briefs*, 2011.
- [2] T. Çukur, J. M. Santos, J. M. Pauly, and D. G. Nishimura, “Variable-density parallel imaging with partially localized coil sensitivities,” *IEEE transactions on medical imaging*, vol. 29, no. 5, pp. 1173–1181, 2010.
- [3] B. Moraal, S. Roosendaal, P. Pouwels, H. Vrenken, R. Schijndel, D. Meier, C. Guttman, J. Geurts, and F. Barkhof, “Multi-contrast, isotropic, single-slab 3d MR imaging in multiple sclerosis,” *European Radiology*, vol. 18, pp. 2311–2320, 2008.
- [4] T. Çukur, N. K. Bangerter, and D. G. Nishimura, “Enhanced spectral shaping in steady-state free precession imaging,” *Magnetic Resonance in Medicine*, vol. 58, no. 6, pp. 1216–1223, 2007.
- [5] T. Çukur and D. G. Nishimura, “Fat-water separation with alternating repetition time balanced ssfp,” *Magnetic Resonance in Medicine*, vol. 60, no. 2, pp. 479–484, 2008.
- [6] T. Çukur and D. G. Nishimura, “Multiple repetition time balanced steady-state free precession imaging,” *Magnetic Resonance in Medicine*, vol. 62, no. 1, pp. 193–204, 2009.

- [7] T. Çukur, J. H. Lee, N. K. Bangerter, B. A. Hargreaves, and D. G. Nishimura, “Non-contrast-enhanced flow-independent peripheral mr angiography with balanced ssfp,” *Magnetic Resonance in Medicine*, vol. 61, no. 6, pp. 1533–1539, 2009.
- [8] T. Çukur, M. Yamada, W. R. Overall, P. Yang, and D. G. Nishimura, “Positive contrast with alternating repetition time ssfp (parts): A fast imaging technique for spio-labeled cells,” *Magnetic Resonance in Medicine*, vol. 63, no. 2, pp. 427–437, 2010.
- [9] K. Keskin, U. Yılmaz, and T. Çukur, “Constrained ellipse fitting for efficient parameter mapping with phase-cycled bssfp mri,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 1, pp. 14–26, 2022.
- [10] E. Bıyık, K. Keskin, S. UH Dar, A. Koç, and T. Çukur, “Factorized sensitivity estimation for artifact suppression in phase-cycled bssfp mri,” *NMR in Biomedicine*, vol. 33, no. 4, p. e4228, 2020.
- [11] T. Çukur, M. Lustig, and D. G. Nishimura, “Multiple-profile homogeneous image combination: application to phase-cycled ssfp and multicoil imaging,” *Magnetic resonance in medicine*, vol. 60, no. 3, pp. 732–738, 2008.
- [12] T. Çukur, M. Lustig, E. U. Saritas, and D. G. Nishimura, “Signal compensation and compressed sensing for magnetization-prepared mr angiography,” *IEEE transactions on medical imaging*, vol. 30, no. 5, pp. 1017–1027, 2011.
- [13] T. Çukur, A. Shimakawa, H. Yu, B. A. Hargreaves, B. S. Hu, D. G. Nishimura, and J. H. Brittain, “Magnetization-prepared ideal bssfp: a flow-independent technique for noncontrast-enhanced peripheral angiography,” *Journal of Magnetic Resonance Imaging*, vol. 33, no. 4, pp. 931–939, 2011.
- [14] N. K. Bangerter, T. Cukur, B. A. Hargreaves, B. S. Hu, J. H. Brittain, D. Park, G. E. Gold, and D. G. Nishimura, “Three-dimensional fluid-suppressed t2-prep flow-independent peripheral angiography using balanced ssfp,” *Magnetic resonance imaging*, vol. 29, no. 8, pp. 1119–1124, 2011.

- [15] M. Aksoy, C. Forman, M. Straka, T. Çukur, J. Hornegger, and R. Bammer, “Hybrid prospective and retrospective head motion correction to mitigate cross-calibration errors,” *Magnetic resonance in medicine*, vol. 67, no. 5, pp. 1237–1251, 2012.
- [16] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Çukur, “Image synthesis in multi-contrast MRI with conditional generative adversarial networks,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.
- [17] M. Yurt, O. Dalmaz, S. U. H. Dar, M. Özbey, B. Tınaz, K. K. Oğuz, and T. Çukur, “Semi-supervised learning of MRI synthesis without fully-sampled ground truths,” *IEEE Trans Med Imaging*, vol. 41, no. 12, pp. 3895–3906, 2022.
- [18] W. A. Edelstein, M. Mahesh, and J. A. Carrino, “Mri: time is dose—and money and versatility,” *Journal of the American College of Radiology*, vol. 7, no. 8, pp. 650–652, 2010.
- [19] M. Lustig, D. Donoho, J. Santos, and J. Pauly, “Compressed sensing mri,” *Signal Processing Magazine, IEEE*, vol. 25, pp. 72 – 82, 04 2008.
- [20] M. Shahdloo, E. Ilıcak, M. Tofghi, E. U. Saritas, A. E. Çetin, and T. Çukur, “Projection onto epigraph sets for rapid self-tuning compressed sensing mri,” *IEEE transactions on medical imaging*, vol. 38, no. 7, pp. 1677–1689, 2018.
- [21] B. Gözcü, R. K. Mahabadi, Y.-H. Li, E. Ilıcak, T. Cukur, J. Scarlett, and V. Cevher, “Learning-based compressive mri,” *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1394–1406, 2018.
- [22] S. U. H. Dar, M. Özbey, A. B. Çath, and T. Çukur, “A transfer-learning approach for accelerated mri using deep neural networks,” *Magnetic resonance in medicine*, vol. 84, no. 2, pp. 663–685, 2020.
- [23] T. Cukur, M. Lustig, and D. G. Nishimura, “Improving non-contrast-enhanced steady-state free precession angiography with compressed sensing,” *Magnetic Resonance in Medicine*, vol. 61, no. 5, pp. 1122–1131, 2009.

- [24] T. Çukur, “Accelerated phase-cycled ssfp imaging with compressed sensing,” *IEEE transactions on medical imaging*, vol. 34, no. 1, pp. 107–115, 2014.
- [25] S. Ilbey, C. B. Top, A. Güngör, T. Çukur, E. U. Saritas, and H. E. Güven, “Comparison of system-matrix-based and projection-based reconstructions for field free line magnetic particle imaging,” *International Journal on Magnetic Particle Imaging*, vol. 3, no. 1, pp. 1–8, 2017.
- [26] E. Kopanoglu, A. Güngör, T. Kilic, E. U. Saritas, K. K. Oguz, T. Çukur, and H. E. Güven, “Simultaneous use of individual and joint regularization terms in compressive sensing: Joint reconstruction of multi-channel multi-contrast mri acquisitions,” *NMR in Biomedicine*, vol. 33, no. 4, p. e4247, 2020.
- [27] T. Sanchez, B. Gözcü, R. B. van Heeswijk, A. Eftekhari, E. Ilıcak, T. Çukur, and V. Cevher, “Scalable learning-based sampling optimization for compressive dynamic mri,” in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 8584–8588, IEEE, 2020.
- [28] T. Çukur, J. M. Santos, D. G. Nishimura, and J. M. Pauly, “Varying kernel-extent gridding reconstruction for undersampled variable-density spirals,” *Magnetic Resonance in Medicine*, vol. 59, no. 1, pp. 196–201, 2008.
- [29] S. U. H. Dar, M. Yurt, and T. Çukur, “A few-shot learning approach for accelerated mri via fusion of data-driven and subject-driven priors,” *arXiv preprint arXiv:2103.07790*, 2021.
- [30] A. Koç, B. Bartan, E. Gundogdu, T. Çukur, and H. M. Ozaktas, “Sparse representation of two-and three-dimensional images with fractional fourier, hartley, linear canonical, and haar wavelet transforms,” *Expert Systems with Applications*, vol. 77, pp. 247–255, 2017.
- [31] A. Güngör, E. Kopanoğlu, T. Çukur, and H. E. Güven, “Fast recovery of compressed multi-contrast magnetic resonance images,” in *Medical Imaging 2017: Image Processing*, vol. 10133, pp. 503–508, SPIE, 2017.

- [32] M. Blaimer, F. Breuer, M. Mueller, R. Heidemann, M. Griswold, and P. Jakob, “Smash, sense, pils, grappa: how to choose the optimal method,” *Topics in magnetic resonance imaging : TMRI*, vol. 15, pp. 223–36, 09 2004.
- [33] K. P. Pruessmann, M. Weiger, M. B. Scheidegger, and P. Boesiger, “Sense: sensitivity encoding for fast mri,” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 42, no. 5, pp. 952–962, 1999.
- [34] M. A. Griswold, P. M. Jakob, R. M. Heidemann, M. Nittka, V. Jellus, J. Wang, B. Kiefer, and A. Haase, “Generalized autocalibrating partially parallel acquisitions (grappa),” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 47, no. 6, pp. 1202–1210, 2002.
- [35] W. E. Kyriakos, L. P. Panych, D. F. Kacher, C.-F. Westin, S. M. Bao, R. V. Mulkern, and F. A. Jolesz, “Sensitivity profiles from an array of coils for encoding and reconstruction in parallel (space rip),” *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 44, no. 2, pp. 301–308, 2000.
- [36] J. C. Ye, “Compressed sensing mri: a review from signal processing perspective,” *BMC Biomedical Engineering*, vol. 1, no. 1, pp. 1–17, 2019.
- [37] K. G. Hollingsworth, “Reducing acquisition time in clinical mri by data undersampling and compressed sensing reconstruction,” *Phys Med Biol*, vol. 60, no. 21, p. R297, 2015.
- [38] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, “Learning a variational network for reconstruction of accelerated mri data,” *Magnetic resonance in medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.
- [39] A. Güngör, B. Askin, D. A. Soydan, C. B. Top, E. U. Saritas, and T. Çukur, “Deq-mpi: A deep equilibrium reconstruction with learned consistency for magnetic particle imaging,” *IEEE Transactions on Medical Imaging*, 2023.

- [40] Y. Korkmaz, T. Cukur, and V. Patel, “Self-supervised mri reconstruction with unrolled diffusion models,” *arXiv preprint arXiv:2306.16654*, 2023.
- [41] Y. Korkmaz, S. U. Dar, M. Yurt, M. Özbey, and T. Cukur, “Unsupervised mri reconstruction via zero-shot learned adversarial transformers,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1747–1763, 2022.
- [42] Y. Korkmaz, M. Yurt, S. U. H. Dar, M. Özbey, and T. Cukur, “Deep mri reconstruction with generative vision transformers,” in *Machine Learning for Medical Image Reconstruction: 4th International Workshop, MLMIR 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, October 1, 2021, Proceedings 4*, pp. 54–64, Springer, 2021.
- [43] G. Elmas, S. U. Dar, Y. Korkmaz, E. Ceyani, B. Susam, M. Ozbey, S. Avestimehr, and T. Çukur, “Federated learning of generative image priors for mri reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 42, no. 7, pp. 1996–2009, 2023.
- [44] J. Schlemper, J. Caballero, J. V. Hajnal, A. N. Price, and D. Rueckert, “A deep cascade of convolutional neural networks for dynamic mr image reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 37, no. 2, pp. 491–503, 2018.
- [45] C. Qin, J. Schlemper, J. Caballero, A. N. Price, J. V. Hajnal, and D. Rueckert, “Convolutional recurrent neural networks for dynamic MR image reconstruction,” *IEEE Transactions on Medical Imaging*, vol. 38, no. 1, pp. 280–290, 2019.
- [46] S. U. Dar, M. Yurt, M. Shahdloo, M. E. Ildız, B. Tınaz, and T. Çukur, “Prior-guided image reconstruction for accelerated multi-contrast mri via generative adversarial networks,” *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, pp. 1072–1087, 2020.
- [47] M. Akçakaya, S. Moeller, S. Weingärtner, and K. Uğurbil, “Scan-specific robust artificial-neural-networks for k-space interpolation (raki) reconstruction: Database-free deep learning for fast imaging,” *Magnetic resonance in medicine*, vol. 81, no. 1, pp. 439–453, 2019.

- [48] A. Gungor, B. Aşkın, D. A. Soydan, E. U. Saritas, C. B. Top, and T. Çukur, “A denoiser scaling technique for plug-and-play mpi reconstruction,” *International Journal on Magnetic Particle Imaging IJMPI*, vol. 9, no. 1 Suppl 1, 2023.
- [49] M. Özbey, O. Dalmaz, S. U. Dar, H. A. Bedel, , Ş. Öztürk, A. Güngör, and T. Çukur, “Unsupervised medical image translation with adversarial diffusion models,” *IEEE Transactions on Medical Imaging*, 2023.
- [50] A. Güngör, S. U. Dar, Şaban Öztürk, Y. Korkmaz, H. A. Bedel, G. Elmas, M. Ozbey, and T. Çukur, “Adaptive diffusion priors for accelerated mri reconstruction,” *Medical Image Analysis*, vol. 88, p. 102872, 2023.
- [51] S. U. Dar, Ş. Öztürk, M. Özbey, and T. Çukur, “Learning deep mri reconstruction models from scratch in low-data regimes,” *arXiv preprint arXiv:2301.02613*, 2023.
- [52] O. Dalmaz, U. Mirza, G. Elmas, M. Özbey, S. U. Dar, E. Ceyani, S. Avestimehr, and T. Çukur, “One model to unite them all: Personalized federated learning of multi-contrast mri synthesis,” *arXiv preprint arXiv:2207.06509*, 2022.
- [53] O. Dalmaz, U. Mirza, G. Elmas, M. Özbey, S. U. Dar, and T. Çukur, “A specificity-preserving generative model for federated mri translation,” in *Distributed, Collaborative, and Federated Learning, and Affordable AI and Healthcare for Resource Diverse Global Health: Third MICCAI Workshop, DeCaF 2022, and Second MICCAI Workshop, FAIR 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18 and 22, 2022, Proceedings*, pp. 79–88, Springer, 2022.
- [54] M. Yurt, S. U. Dar, A. Erdem, E. Erdem, K. K. Oguz, and T. Çukur, “mustGAN: multi-stream generative adversarial networks for MR image synthesis,” *Medical Image Analysis*, vol. 70, p. 101944, 2021.
- [55] O. Dalmaz, U. Mirza, G. Elmas, M. Özbey, S. U. Dar, and T. Çukur, “A specificity-preserving generative model for federated mri translation,” in *Distributed, Collaborative, and Federated Learning, and Affordable AI*

and Healthcare for Resource Diverse Global Health: Third MICCAI Workshop, DeCaF 2022, and Second MICCAI Workshop, FAIR 2022, Held in Conjunction with MICCAI 2022, Singapore, September 18 and 22, 2022, Proceedings, pp. 79–88, Springer, 2022.

- [56] I. Aytekin, O. Dalmaz, H. Ankishan, E. U. Saritas, U. Bagci, T. Cukur, and H. Celik, “Detecting COVID-19 from respiratory sound recordings with transformers,” in *Medical Imaging 2022: Computer-Aided Diagnosis* (K. Drukker and K. M. Iftikharuddin, eds.), vol. 12033, p. 1203306, International Society for Optics and Photonics, SPIE, 2022.
- [57] I. Aytekin, O. Dalmaz, K. Gonc, H. Ankishan, E. U. Saritas, U. Bagci, H. Celik, and T. Cukur, “Covid-19 detection from respiratory sounds with hierarchical spectrogram transformers,” *arXiv preprint arXiv:2207.09529*, 2022.
- [58] N. Fuin, A. Bustin, T. Küstner, I. Oksuz, J. Clough, A. P. King, J. A. Schnabel, R. M. Botnar, and C. Prieto, “A multi-scale variational neural network for accelerating motion-compensated whole-heart 3d coronary mr angiography,” *Magnetic Resonance Imaging*, vol. 70, pp. 155–167, 2020.
- [59] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, “Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data,” *Magnetic Resonance in Medicine*, vol. 84, p. 3172–3191, Jul 2020.
- [60] B. Yaman, S. A. H. Hosseini, and M. Akçakaya, “Zero-shot self-supervised learning for mri reconstruction,” *arXiv:2102.07737*, 2021.
- [61] B. Yaman, C. Shenoy, Z. Deng, S. Moeller, H. El-Rewaidy, R. Nezafat, and M. Akçakaya, “Self-supervised physics-guided deep learning reconstruction for high-resolution 3d lge cmr,” in *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pp. 100–104, 2021.
- [62] Q. Huang, Y. Xian, D. Yang, H. Qu, J. Yi, P. Wu, and D. N. Metaxas, “Dynamic mri reconstruction with end-to-end motion-guided network,” *Medical Image Analysis*, vol. 68, p. 101901, 2021.

- [63] C. Chen, Y. Liu, P. Schniter, M. Tong, K. Zareba, O. Simonetti, L. Potter, and R. Ahmad, “Ocmr (v1.0)–open-access multi-coil k-space dataset for cardiovascular magnetic resonance imaging,” 2020.
- [64] E. Ilicak, E. U. Saritas, and T. Çukur, “Automated parameter selection for accelerated mri reconstruction via low-rank modeling of local k-space neighborhoods,” *Zeitschrift für Medizinische Physik*, vol. 33, no. 2, pp. 203–219, 2023.
- [65] L. K. Senel, T. Kilic, A. Gungor, E. Kopanoglu, H. E. Guven, E. U. Saritas, A. Koc, and T. Cukur, “Statistically segregated k-space sampling for accelerating multiple-acquisition mri,” *IEEE transactions on medical imaging*, vol. 38, no. 7, pp. 1701–1714, 2019.
- [66] Y. Wu, Y. Ma, J. Liu, J. Du, and L. Xing, “Self-attention convolutional neural network for improved mr image reconstruction,” *Information sciences*, vol. 490, pp. 317–328, 2019.
- [67] Q. Huang, D. Yang, P. Wu, H. Qu, J. Yi, and D. Metaxas, “Mri reconstruction via cascaded channel-wise attention network,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1622–1626, 2019.
- [68] Z. Yuan, M. Jiang, Y. Wang, B. Wei, Y. Li, P. Wang, W. Menpes-Smith, Z. Niu, and G. Yang, “Sara-gan: Self-attention and relative average discriminator based generative adversarial networks for fast compressed sensing mri reconstruction,” *Frontiers in Neuroinformatics*, vol. 14, 2020.
- [69] A. Pramanik and M. Jacob, “Reconstruction and segmentation of parallel mr data using image domain deep-slr,” *arXiv:2102.01172*, 2021.
- [70] Q. Huang, D. Yang, J. Yi, L. Axel, and D. Metaxas, “Fr-net: Joint reconstruction and segmentation in compressed sensing cardiac mri,” in *International conference on functional imaging and modeling of the heart*, pp. 352–360, Springer, 2019.

- [71] T. Çukur, S. Nishimoto, A. G. Huth, and J. L. Gallant, “Attention during natural vision warps semantic representation across the human brain,” *Nature neuroscience*, vol. 16, no. 6, pp. 763–770, 2013.
- [72] T. Çukur, A. G. Huth, S. Nishimoto, and J. L. Gallant, “Functional subdomains within human ffa,” *Journal of Neuroscience*, vol. 33, no. 42, pp. 16748–16766, 2013.
- [73] T. Çukur, A. G. Huth, S. Nishimoto, and J. L. Gallant, “Functional subdomains within scene-selective cortex: parahippocampal place area, retrosplenial complex, and occipital place area,” *Journal of Neuroscience*, vol. 36, no. 40, pp. 10257–10273, 2016.
- [74] E. Çelik, U. Keles, İ. Kiremitçi, J. L. Gallant, and T. Çukur, “Cortical networks of dynamic scene category representation in the human brain,” *Cortex*, vol. 143, pp. 127–147, 2021.
- [75] E. Çelik, S. U. H. Dar, Ö. Yılmaz, Ü. Keleş, and T. Çukur, “Spatially informed voxelwise modeling for naturalistic fmri experiments,” *NeuroImage*, vol. 186, pp. 741–757, 2019.
- [76] M. Shahdloo, E. Çelik, and T. Çukur, “Biased competition in semantic representation during natural visual search,” *Neuroimage*, vol. 216, p. 116383, 2020.
- [77] M. Shahdloo, E. Çelik, B. A. Urgen, J. L. Gallant, and T. Çukur, “Task-dependent warping of semantic representations during search for visual action categories,” *Journal of Neuroscience*, vol. 42, no. 35, pp. 6782–6799, 2022.
- [78] I. Kiremitçi, Ö. Yılmaz, E. Çelik, M. Shahdloo, A. G. Huth, and T. Çukur, “Attentional modulation of hierarchical speech representations in a multitalker environment,” *bioRxiv*, pp. 2020–12, 2021.
- [79] Ö. Yılmaz, E. Çelik, and T. Çukur, “Informed feature regularization in voxelwise modeling for naturalistic fmri experiments,” *European Journal of Neuroscience*, vol. 52, no. 5, pp. 3394–3410, 2020.

- [80] T. Fan, G. Wang, Y. Li, and H. Wang, “Ma-net: A multi-scale attention network for liver and tumor segmentation,” *IEEE Access*, vol. 8, pp. 179656–179665, 2020.
- [81] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, “Squeeze-and-excitation networks,” *arXiv:1709.01507*, 2017.
- [82] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” *arXiv:1512.03385*, 2015.
- [83] S. Ozturk, E. Celik, and T. Cukur, “Content-based medical image retrieval with opponent class adaptive margin loss,” *Information Sciences*, vol. 637, p. 118938, 2023.
- [84] A. Amyar, R. Modzelewski, H. Li, and S. Ruan, “Multi-task deep learning based ct imaging analysis for covid-19 pneumonia: Classification and segmentation,” *Computers in Biology and Medicine*, vol. 126, p. 104037, 2020.
- [85] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, “Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations,” in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 240–248, Springer International Publishing, 2017.
- [86] K. Kwon, D. Kim, and H. Park, “A parallel mr imaging method using multilayer perceptron,” *Medical physics*, vol. 44, no. 12, pp. 6209–6224, 2017.
- [87] K. Hammernik, T. Klatzer, E. Kobler, M. P. Recht, D. K. Sodickson, T. Pock, and F. Knoll, “Learning a variational network for reconstruction of accelerated mri data,” *Magnetic resonance in medicine*, vol. 79, no. 6, pp. 3055–3071, 2018.
- [88] M. J. Muckley, B. Riemenschneider, A. Radmanesh, S. Kim, G. Jeong, J. Ko, Y. Jun, H. Shin, D. Hwang, M. Mostapha, *et al.*, “State-of-the-art machine learning mri reconstruction in 2020: Results of the second fastmri challenge,” *arXiv preprint arXiv:2012.06318*, vol. 2, no. 6, p. 7, 2020.

- [89] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, “Attention gated networks: Learning to leverage salient regions in medical images,” *Medical image analysis*, vol. 53, pp. 197–207, 2019.
- [90] Q. Huang, D. Yang, P. Wu, H. Qu, J. Yi, and D. Metaxas, “Mri reconstruction via cascaded channel-wise attention network,” in *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, pp. 1622–1626, IEEE, 2019.
- [91] W. Zhou, H. Du, W. Mei, and L. Fang, “Spatial orthogonal attention generative adversarial network for mri reconstruction,” *Medical Physics*, vol. 48, no. 2, pp. 627–639, 2021.
- [92] E. Ilicak, S. Cetin, E. Bulut, K. K. Oguz, E. U. Saritas, G. Unal, and T. Çukur, “Targeted vessel reconstruction in non-contrast-enhanced steady-state free precession angiography,” *NMR in Biomedicine*, vol. 29, no. 5, pp. 532–544, 2016.
- [93] S. Turk, A. Demirkaya, M. Y. Turali, C. Hepdurgun, S. U. Dar, A. K. Karabulut, A. Azizova, M. Orman, I. Tamsel, U. Aydingoz, M. Argin, and T. Cukur, “Jointnet: A deep model for predicting active sacroiliitis from sacroiliac joint radiography,” *arXiv preprint arXiv:2301.10769*, 2023.
- [94] F. T. Y. Vural, S. D. Newman, T. Çukur, and I. Ö. Ertugrul, “Machine learning methods for human brain imaging,” *Frontiers in Neuroinformatics*, vol. 17, 2023.
- [95] I. Sivgin, H. A. Bedel, Ş. Öztürk, and T. Çukur, “A plug-in graph neural network to boost temporal sensitivity in fmri analysis,” *arXiv preprint arXiv:2301.00439*, 2023.
- [96] Ş. Öztürk and T. Çukur, “Deep clustering via center-oriented margin free-triplet loss for skin lesion detection in highly imbalanced datasets,” *IEEE Journal of Biomedical and Health Informatics*, vol. 26, no. 9, pp. 4679–4690, 2022.

- [97] L. Sun, Z. Fan, X. Ding, Y. Huang, and J. Paisley, “Region-of-interest undersampled mri reconstruction: a deep convolutional neural network approach,” *Magnetic resonance imaging*, vol. 63, pp. 185–192, 2019.
- [98] Z. Fan, L. Sun, X. Ding, Y. Huang, C. Cai, and J. Paisley, “A segmentation-aware deep fusion network for compressed sensing mri,” in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 55–70, 2018.
- [99] L. Sun, H. Huang, X. Ding, Y. Huang, X. Liu, and Y. Yu, “Fast magnetic resonance imaging on regions of interest: From sensing to reconstruction,” in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*, pp. 97–106, Springer, 2021.
- [100] M. Acar, T. Çukur, and İ. Öksüz, “Segmentation-aware mri reconstruction,” in *Machine Learning for Medical Image Reconstruction: 5th International Workshop, MLMIR 2022, Held in Conjunction with MICCAI 2022, Singapore, September 22, 2022, Proceedings*, pp. 53–61, Springer, 2022.
- [101] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, “Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data,” *Magnetic resonance in medicine*, vol. 84, no. 6, pp. 3172–3191, 2020.
- [102] G. Yang, S. Yu, H. Dong, G. Slabaugh, P. L. Dragotti, X. Ye, F. Liu, S. Arridge, J. Keegan, Y. Guo, *et al.*, “Dagan: deep de-aliasing generative adversarial networks for fast compressed sensing mri reconstruction,” *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1310–1321, 2017.
- [103] O. Dalmaz, M. Yurt, and T. Çukur, “Resvit: residual vision transformers for multimodal medical image synthesis,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2598–2614, 2022.
- [104] Z. Yuan, M. Jiang, Y. Wang, B. Wei, Y. Li, P. Wang, W. Menpes-Smith, Z. Niu, and G. Yang, “Sara-gan: Self-attention and relative average discriminator based generative adversarial networks for fast compressed sensing mri reconstruction,” *Frontiers in Neuroinformatics*, vol. 14, p. 611666, 2020.

- [105] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” *Advances in neural information processing systems*, vol. 30, 2017.
- [106] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, “Cbam: Convolutional block attention module,” in *Proceedings of the European conference on computer vision (ECCV)*, pp. 3–19, 2018.
- [107] I. Mazumdar and J. Mukherjee, “Fully automatic mri brain tumor segmentation using efficient spatial attention convolutional networks with composite loss,” *Neurocomputing*, vol. 500, pp. 243–254, 2022.
- [108] L. Bi, D. Feng, and J. Kim, “Dual-path adversarial learning for fully convolutional network (fcn)-based medical image segmentation,” *The Visual Computer*, vol. 34, pp. 1043–1052, 2018.
- [109] R. Zhao, B. Qian, X. Zhang, Y. Li, R. Wei, Y. Liu, and Y. Pan, “Rethinking dice loss for medical image segmentation,” in *2020 IEEE International Conference on Data Mining (ICDM)*, pp. 851–860, IEEE, 2020.
- [110] S. Targ, D. Almeida, and K. Lyman, “Resnet in resnet: Generalizing residual architectures,” *arXiv preprint arXiv:1603.08029*, 2016.
- [111] M. A. Mazurowski, K. Clark, N. M. Czarnek, P. Shamsesfandabadi, K. B. Peters, and A. Saha, “Radiogenomics of lower-grade glioma: algorithmically-assessed tumor shape is associated with tumor genomic subtypes and patient outcomes in a multi-institutional study with the cancer genome atlas data,” *Journal of neuro-oncology*, vol. 133, pp. 27–35, 2017.
- [112] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [113] T. Kim and J. Haldar, “Loraks software version 2.0: Faster implementation and enhanced capabilities,” *University of Southern California, Los Angeles, CA, Tech. Rep. USC-SIPI-443*, 2018.
- [114] O. Ronneberger, P. Fischer, and T. Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *Medical Image Computing and*

Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241, Springer, 2015.

- [115] A. Gungor, B. Askin, D. A. Soydan, E. U. Saritas, C. B. Top, and T. Çukur, “TranSMS: Transformers for super-resolution calibration in magnetic particle imaging,” *IEEE Transactions on Medical Imaging*, vol. 41, no. 12, pp. 3562–3574, 2022.
- [116] H. A. Bedel, I. Sivgin, O. Dalmaz, S. U. Dar, and T. Çukur, “Bolt: Fused window transformers for fmri time series analysis,” *Medical Image Analysis*, vol. 88, p. 102841, 2023.
- [117] Y. Wang, W. Zhuo, Y. Li, Z. Wang, Q. Ju, and W. Zhu, “Fully self-supervised learning for semantic segmentation,” *arXiv preprint arXiv:2202.11981*, 2022.
- [118] M. U. Mirza, O. Dalmaz, H. A. Bedel, G. Elmas, Y. Korkmaz, A. Gungor, S. U. Dar, and T. Çukur, “Learning fourier-constrained diffusion bridges for mri reconstruction,” *arXiv preprint arXiv:2308.01096*, 2023.
- [119] H. A. Bedel and T. Çukur, “Dreamr: Diffusion-driven counterfactual explanation for functional mri,” *arXiv preprint arXiv:2307.09547*, 2023.