Journal of Machine Learning for Biomedical Imaging. 2022:018. pp 1-28 Submitted 10/2021; Published 06/2022

Joint Frequency and Image Space Learning for MRI Reconstruction and Analysis

Nalini M. Singh nmsingh@mit.edu
Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA
Dept. of Health Sciences & Technology, MIT, Cambridge, MA, USA

Juan Eugenio Iglesias e.iglesias@ucl.ac.uk
A. A. Martinos Center, Massachusetts General Hospital, Boston, MA, USA
Harvard Medical School, Cambridge, MA, USA
Centre for Medical Image Computing, UCL, London, UK
Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA

Elfar Adalsteinsson elfar@mit.edu
Research Laboratory of Electronics, MIT, Cambridge, MA, USA
Dept. of Electrical Engineering & Computer Science, MIT, Cambridge, MA, USA

Adrian V. Dalca adalca@mit.edu
A. A. Martinos Center, Massachusetts General Hospital, Boston, MA, USA
Harvard Medical School, Cambridge, MA, USA
Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA

Polina Golland polina@csail.mit.edu
Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, MA, USA
Dept. of Electrical Engineering & Computer Science, MIT, Cambridge, MA, USA

Abstract

We propose neural network layers that explicitly combine frequency and image feature representations and show that they can be used as a versatile building block for reconstruction from frequency space data. Our work is motivated by the challenges arising in MRI acquisition where the signal is a corrupted Fourier transform of the desired image. The proposed joint learning schemes enable both correction of artifacts native to the frequency space and manipulation of image space representations to reconstruct coherent image structures at every layer of the network. This is in contrast to most current deep learning approaches for image reconstruction that treat frequency and image space features separately and often operate exclusively in one of the two spaces. We demonstrate the advantages of joint convolutional learning for a variety of tasks, including motion correction, denoising, reconstruction from undersampled acquisitions, and combined undersampling and motion correction on simulated and real world multicoil MRI data. The joint models produce consistently high quality output images across all tasks and datasets. When integrated into a state of the art unrolled optimization network with physics-inspired data consistency constraints for undersampled reconstruction, the proposed architectures significantly improve the optimization landscape, which yields an order of magnitude reduction of training time. This result suggests that joint representations are particularly well suited for MRI signals in deep learning networks. Our code and pretrained models are publicly available at `https://github.com/nalinimsingh/interlacer`.

1. Introduction

Magnetic resonance imaging (MRI) (Lauterbur, 1973) acquires frequency space data and converts these measurements to images for visualization and downstream analysis. Practical

`https://www.melba-journal.org/papers/2022:018.html`.

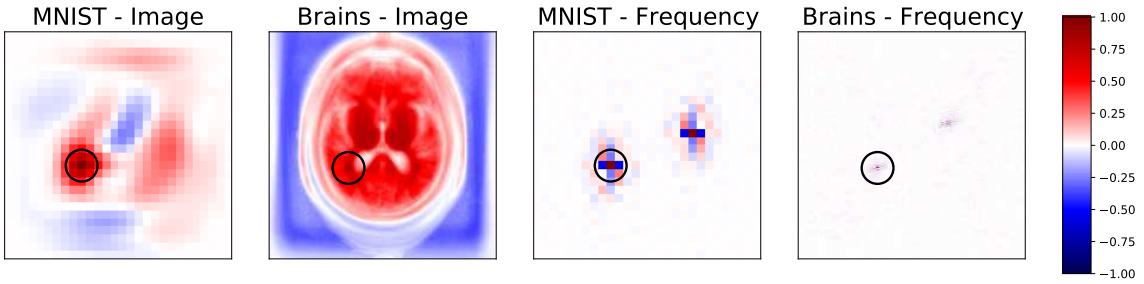

Figure 1: Maps of correlation coefficients between a single pixel (center of circle) and all other pixels in image (left two panels) and frequency space (right two panels) representations of MNIST and a brain MRI dataset. All maps show strong local correlations useful for inferring missing or corrupted data in both spaces. Frequency space correlations also display conjugate symmetry characteristic of Fourier transforms of real images.

imaging considerations often affect the data acquisition process. For example, motion occurs during acquisition (Andre et al., 2015), noise affects sensor readings (Macovski, 1996), and sub-Nyquist undersampling is routinely used to speed up data acquisition (Lustig et al., 2008). Traditionally, the acquired frequency space signals are converted to image space reconstructions via an inverse Fourier transform, with each individual frequency space measurement contributing to all output pixels in the image space. As a result, local changes in the acquired frequency space data induce global effects on the entire output image. To produce accurate image reconstructions, modeling tools for Fourier imaging must correct these global artifacts in addition to performing fine-scale image space processing.

Recently, neural networks have emerged as an alternative approach for MRI reconstruction (Aggarwal et al., 2018; Hammernik et al., 2018; Hyun et al., 2018; Lee et al., 2017; Putzky and Welling, 2019; Quan et al., 2018; Schlemper et al., 2017; Sun et al., 2016; Yang et al., 2017; Aggarwal et al., 2018; Hammernik et al., 2018; Cheng et al., 2018; Han et al., 2019; Zhu et al., 2018; Duffy et al., 2021; Haskell et al., 2019; Johnson and Drangova, 2019; Küstner et al., 2019; Pawar et al., 2018; Shaw et al., 2020; Oksuz et al., 2019; Usman et al., 2020; Benou et al., 2017; Jiang et al., 2018; Manjón and Coupe, 2018). Most existing architectures are based on purely frequency space representations or purely image space representations. Here, we propose and demonstrate joint frequency-image space representations that enable networks to learn a wide set of tasks including and beyond the extensively studied undersampled reconstruction. To motivate our approach, we examine the correlation structure for frequency and image space representations in Fig. 1. Local neighborhoods around a pixel exhibit strong correlations, suggesting that local convolution operations, which are widely successful on image space computer vision tasks, might also be useful when applied to frequency space data to capture this local structure. Convolutional operations in frequency space promise to enable direct correction of local frequency space artifacts corresponding to global image space effects, while convolutional image space processing facilitates complementary correction of artifacts that are best captured in the image domain.

1.1 Prior Work

We study joint representations in the context of three corruption processes that arise during the imaging process.

Motion. Previous retrospective motion correction strategies (Batchelor et al., 2005; Haskell et al., 2018) are cast as large, non-convex optimization problems with iterative solutions that are slow to compute. Deep learning methods (Duffy et al., 2021; Haskell et al., 2019; Johnson and Drangova, 2019; Küstner et al., 2019; Pawar et al., 2018; Shaw et al., 2020; Usman et al., 2020) solve the motion correction problem with a neural network operating purely in the image space, even though motion artifacts are induced directly in the frequency space during data acquisition. An alternative approach has been demonstrated recently that detects motion directly on frequency space data, followed by motion correction via an image space network (Oksuz et al., 2019).

Noise. Previous work on MRI denoising applies classical signal processing techniques including filtering (Manjón et al., 2008) and wavelet-based methods (Anand and Sahambi, 2010; Nowak, 1999). Deep learning methods employ convolutional networks solely on image space data (Benou et al., 2017; Jiang et al., 2018; Manjón and Coupe, 2018).

Undersampling. Classical undersampled reconstruction techniques either construct the output image as a least-squares estimate from the acquired frequency space data (Pruessmann et al., 1999) or combine convolutional filters in the frequency space with an inverse Fourier transform (Griswold et al., 2002; Lustig and Pauly, 2010). Many deep learning methods apply convolutions to image space reconstructions of the acquired undersampled frequency data (Aggarwal et al., 2018; Hammernik et al., 2018; Hyun et al., 2018; Lee et al., 2017; Putzky and Welling, 2019; Quan et al., 2018; Schlemper et al., 2017; Sun et al., 2016; Yang et al., 2017). To improve the quality and fidelity of the reconstruction, the convolutional layers can be combined into an architecture that emulates unrolled optimization, with a convolutional regularizer coupled with a physics-inspired data consistency constraint that is enforced after each iteration (Aggarwal et al., 2018; Hammernik et al., 2018). Alternatively, the convolutional architectures can act directly on the frequency space data (Akçakaya et al., 2019; Cheng et al., 2018; Han et al., 2019). The notably different AUTOMAP architecture uses fully-connected layers to convert frequency space data to the image space and then applies further image space convolutions (Zhu et al., 2018), incurring prohibitive memory complexity of $\mathcal{O}(N^4)$ for a $N \times N$ image.

More recently, solutions that combine frequency and image space convolutions have been demonstrated in the context of undersampled reconstruction. One approach is to combine separately trained pure frequency and pure image space networks into a common architecture (Eo et al., 2018; Souza and Frayne, 2019; Wang et al., 2019). The most closely related work to ours integrates frequency and image space blocks within the same network (Zhou and Zhou, 2020), effectively implementing one of the two variants we consider in this paper. Here we propose an additional layer architecture that also tightly couples frequency and image space representations and evaluate both variants on a wide variety of tasks, well beyond the undersampled reconstruction scenario for which the previously combined architectures have been proposed.

In our experiments, a basic network that simply concatenates joint layers outperforms its pure frequency and image counterparts across a large set of artifacts and reconstruction quality metrics. To investigate how the joint layer architecture interacts with the data consistency constraints often used in undersampled reconstruction, we train the basic network with such a constraint and observe that it compares favorably with the state of the art task-specific undersampled reconstruction networks (Eo et al., 2018; Schlemper et al., 2017) that also incorporate a data consistency constraint. Moreover, we probe the relationship between the proposed joint layers and the widely used unrolled optimization architectures by replacing image convolutional layers with our joint layers in a state of the art unrolled optimization network, MoDL (Aggarwal et al., 2018). Using the proposed joint layers improves the training landscape and reduces training time by about an order of magnitude.

To summarize, our contributions are as follows:

1. We define two task-independent convolutional layer architectures that tightly couple frequency and image representations of an input image that can be used in conjunction with unrolled optimization, data consistency constraints, and other sophisticated strategies for building and training reconstruction neural networks.

2. We demonstrate in simulation experiments that joint networks outperform pure image or pure frequency space networks for reconstructing high quality images in the presence of (i) extreme motion, (ii) heavy noise, and (iii) combination of artifacts, such as motion and undersampling.

3. We demonstrate that the proposed joint learning strategy is compatible with a data consistency constraint and performs favorably relative to state-of-the-art networks specifically designed for the undersampled reconstruction task.

4. We demonstrate on complex-valued, multicoil, real world data that incorporating joint layers into unrolled optimization networks results in more effective training and an order of magnitude decrease of training time, suggesting that the proposed architectures are particularly well suited for image representation in MRI reconstruction networks.

This paper is organized as follows. In the next section, we define the proposed layer and network architectures. Section 3 provides the implementation details and describes our ablation studies. Section 4 reports experimental results, followed by the discussion of the proposed layers, their limitations, and conclusions in Section 5.

2. Joint Networks

MRI acquires Fourier transform measurements, referred to as k-space data. We assume a 2D multislice MRI acquisition. For each slice in this setup, the goal of image reconstruction is to generate an image I from the acquired Fourier transform measurements $F = \mathscr{F}\{I\}$. Classically, this reconstruction is computed via a 2D inverse Fourier transform, producing an estimated image $\hat{I} = \mathscr{F}^{-1}\{F\}$. In practice, corrupted and possibly undersampled measurements $\tilde{F}$ are acquired instead of F, and the goal is to estimate the desired image I from the corrupted signal $\tilde{F}$. Many strategies exist for selecting which measurements to acquire in

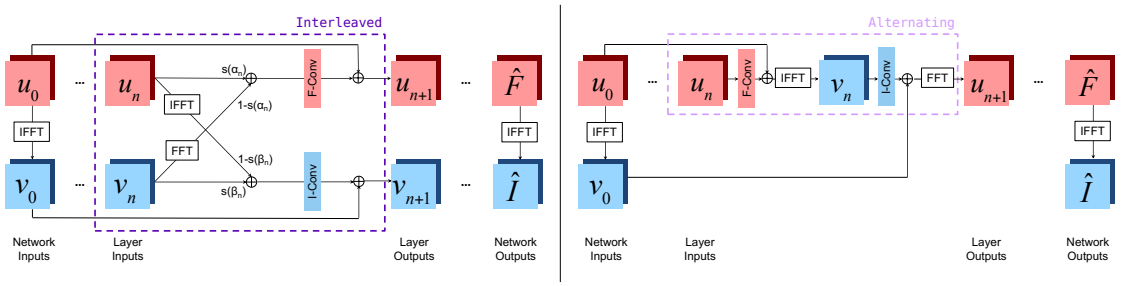

Figure 2: The `Interleaved` (left) and `Alternating` (right) layers, embedded within full network architectures. Each 'F-Conv' or 'I-Conv' block applies Batch Normalization (BN), a convolution, and an activation function in the frequency or image space, respectively.

frequency space. Here we consider Cartesian sampling, where measurement coordinates k_x and k_y are evenly sampled across the 2D Fourier plane, but our method can be generalized to other acquisition schemes. In this section, we define two neural network layer variants that combine image and frequency space convolutional features, referred to as `Interleaved` and `Alternating`, specify the network architectures, and describe the learning procedure.

2.1 Joint Layer Structures

Fig. 2 illustrates the layer structures of the two joint networks. We use u_n to denote the frequency space input and v_n to denote the image space input of layer n. Thus, $u_0 = \tilde{F}$ and $v_0 = \mathscr{F}^{-1}\{u_0\}$ represent the frequency space and image space inputs to the network.

In the `Interleaved` setup, layer inputs are combined via *learned*, layer-specific mixing parameters α_n and β_n that parameterize the sigmoid function $s(x) = (1+e^{-x})^{-1}$ to constrain the mixing coefficients to (0,1):

$$\begin{aligned} \hat{u}_n &= s(\alpha_n)\, u_n + (1 - s(\alpha_n))\, \mathscr{F}\{v_n\}, \\ \hat{v}_n &= s(\beta_n)\, v_n + (1 - s(\beta_n))\, \mathscr{F}^{-1}\{u_n\}. \end{aligned} \tag{1}$$

Real and imaginary parts of inputs are represented as separate channels at each layer and are joined appropriately to form complex numbers when computing the Fourier transform $\mathscr{F}\{\cdot\}$ or its inverse. Next, the layer applies batch normalization (BN), a convolution, and an activation function with a skip connection to produce the outputs:

$$\begin{aligned} u_{n+1} &= \sigma(w_n \circledast \mathrm{BN}(\hat{u}_n) + b_n) + u_0, \\ v_{n+1} &= \sigma'(w'_n \circledast \mathrm{BN}(\hat{v}_n) + b'_n) + v_0, \end{aligned} \tag{2}$$

where (w_n, b_n) are learned frequency space convolution weights and biases, (w'_n, b'_n) are learned image space convolution weights and biases, and $\sigma(\cdot)$ and $\sigma'(\cdot)$ are activation functions specific to the frequency space and image space network components, described later in this section.

This layer architecture is a generalization of networks that operate purely in frequency space, obtained by choosing $s(\alpha_n) = 1$ and $s(\beta_n) = 0$, and of networks that operate purely in image space, that arise when $s(\alpha_n) = 0$ and $s(\beta_n) = 1$. When $0 < s(\alpha_n) < 1$ and $0 <$

$s(\beta_n) < 1$, this layer represents a function that cannot be expressed solely via pure image or frequency space convolutional layers that do not invoke the Fourier transform or its inverse. Note that the frequency output u_n of layer n is not required to be the Fourier transform of the layer's image output v_n, only that the mixing is applied to either two frequency space outputs or two image space outputs. This additional flexibility ensures that u_n and v_n are not entirely redundant and the network learns the right features to capture MRI structure based on the input data and the task at hand.

In the `Alternating` setup, each layer sequentially incorporates frequency and image space convolutions with the appropriate batch normalization and activation function:

$$\begin{aligned} v_n &= \mathscr{F}^{-1}\left\{\sigma(w_n \circledast \mathrm{BN}(u_n) + b_n) + u_0\right\}, \\ u_{n+1} &= \mathscr{F}\left\{\sigma'(w'_n \circledast \mathrm{BN}(v_n) + b'_n) + v_0\right\}, \end{aligned} \tag{3}$$

i.e., the reconstruction alternates between convolutions in the frequency and image space. A version of this architecture was previously introduced as part of a task-specific network for undersampled reconstruction (Zhou and Zhou, 2020).

For both joint architectures, the frequency space convolutions represent element-wise multiplications in the image space. Since the convolution kernels have limited width, the learned convolutions cannot represent all such element-wise multiplications, but instead parameterize the subset whose 2D Fourier transform is zero outside of a central region. Coupled with nonlinearities in the frequency space, these operations enable the network to use global, spatially varying operations not captured by image space convolutions.

Although both of these layers explicitly include the Fourier transform and its inverse, no parameters are associated with those transforms. Thus, we learn only convolutional weights, biases, and possibly mixing coefficients. Since our networks incorporate Fourier transforms, they have an overall $\mathcal{O}(N^2 \log N)$ space complexity for $N \times N$ images.

2.2 Activation Functions

Adopting the standard practice of using the ReLU nonlinearity for image data, we define $\sigma'(x) = \mathrm{ReLU}(x)$ for all convolutions in the image space. This operation is applied separately to real and imaginary channels of each image space convolution output (Trabelsi et al, 2018). However, the zero-gradient of this nonlinearity for negative values is ill-suited for networks that operate on frequency space data, as individual inputs can take on a large range of positive and negative values. We introduce an alternative nonlinear activation function that we apply to both the real and imaginary channels of each frequency space convolution output:

$$\sigma(x) = x + ReLU\left(\frac{x-1}{2}\right) + ReLU\left(-\frac{x+1}{2}\right). \tag{4}$$

This nonlinearity's magnitude increases with that of the input everywhere, while preserving the distinction between positive and negative inputs. We found that networks using this nonlinearity consistently outperformed networks that employed ReLU activation functions on frequency space convolution outputs.

2.3 Learning

The networks evaluated in this paper can be trained with any differentiable loss function $\mathcal{L}$. In our experiments, we investigate a wide variety of loss functions. We train the joint network $f(\cdot;\theta_{\mathrm{f}},\theta_{\mathrm{i}})$ for image reconstruction by optimizing a set of frequency space parameters θ_{f} and a set of image space parameters θ_{i} over the training dataset $\mathcal{D} = \{(\tilde{F}_m, I_m)\}$ using stochastic gradient descent-based strategies to obtain

$$(\theta_{\mathrm{f}}^*, \theta_{\mathrm{i}}^*) = \arg\min_{(\theta_{\mathrm{f}},\theta_{\mathrm{i}})} \sum_{m=1}^{|\mathcal{D}|} \mathcal{L}\left(I_m, \mathscr{F}^{-1}\left(f(\tilde{F}_m;\theta_{\mathrm{f}},\theta_{\mathrm{i}})\right)\right), \tag{5}$$

where θ_{f} and θ_{i} depend on the setup of the joint layer.

3. Implementation Details and Ablation Architectures

We construct each joint network to contain 10 joint frequency and image space layers. We performed a hyperparameter sweep and observed that the accuracy of reconstruction on the validation set stopped improving for networks that included more than 10 joint layers. A single 2D convolutional layer acts on the frequency space output u_{10} of the final joint layer to produce the final 2-channel complex output $\hat{F}$. The estimated image $\hat{I}$ is the inverse Fourier transform of the network's output, i.e., $\hat{I} = \mathscr{F}^{-1}\{\hat{F}\}$. All convolution blocks within both types of joint layers have kernel size 3x3 and 64 output features, resulting in a total of 670,622 parameters for the `Interleaved` network and 706,438 parameters for the `Alternating` network.

To evaluate the utility of combined frequency and image space layers as a network building block for manipulating Fourier imaging data, we compare performance of the `Interleaved` and `Alternating` architectures to two similarly structured baseline architectures with only frequency or only image space operations.

First, we create an architecture `Frequency` that performs convolutions only on frequency space data and train the network $g(\cdot;\theta_{\mathrm{f}})$ to identify frequency space parameters

$$\theta_{\mathrm{f}}^* = \arg\min_{\theta_{\mathrm{f}}} \sum_{m=1}^{|\mathcal{D}|} \mathcal{L}\left(I_m, \mathscr{F}^{-1}\left(g(\tilde{F}_m;\theta_{\mathrm{f}})\right)\right). \tag{6}$$

The network contains 20 convolution layer to match the joint networks' 10 pairs of 2 convolution layers. As in the `Interleaved` and `Alternating` networks, each convolution layer has kernel size 3x3 and 64 output features, followed by the final, two-feature 2D convolutional layer, resulting in 706,438 parameters. This network captures the convolution strategy used in (Akçakaya et al., 2019; Han et al., 2019; Kim et al., 2019), which incorporate frequency space convolutions in the context of other task-specific architectures and loss choices.

We also implement an image space network `Image`. The network $g(\cdot;\theta_{\mathrm{i}})$ is trained by optimizing

$$\theta_{\mathrm{i}}^* = \arg\min_{\theta_{\mathrm{i}}} \sum_{m=1}^{|\mathcal{D}|} \mathcal{L}\left(I_m, g\left(\mathscr{F}^{-1}\left(\tilde{F}_m\right);\theta_{\mathrm{i}}\right)\right). \tag{7}$$

This network's architecture is identical to that of `Frequency` and also contains 706,438 parameters, but it operates on image space data. This network captures the convolution

strategy used in prior work that incorporates image space convolutions with task-specific architectures and loss function choices, e.g., unrolled optimization and data consistency constraints (Aggarwal et al., 2018; Hammernik et al., 2018; Haskell et al., 2019; Hyun et al., 2018; Küstner et al., 2019; Lee et al., 2017; Manjón and Coupe, 2018; Pawar et al., 2018; Putzky and Welling, 2019; Quan et al., 2018; Schlemper et al., 2017; Sun et al., 2016; Yang et al., 2017).

We initialize all convolution weights using the He normal initializer (He et al., 2015) and use the Adam optimizer (Kingma and Ba, 2014) (learning rate 0.001) until convergence. We initialize $s(\alpha)$ and $s(\beta)$ to 0.5. Training each model requires one day on an NVIDIA RTX 2080 Ti GPU. Our code and pre-trained models for each of these networks is available at `https://github.com/nalinimsingh/interlacer`.

4. Experiments

In this section, we evaluate the proposed joint layers in a set of experiments that progress from simulated data and basic networks to real world complex-valued multicoil MRI measurements and unrolled optimization frameworks with physics-inspired data consistency constraints. The experiments in this section are performed on brain MRIs from multiple datasets. Additional experiments on FastMRI single coil knee MRI, including comparisons with the top methods on FastMRI leaderboard, are provided in Appendix A.

4.1 No Data Consistency

In this section, we present experiments where no data consistency contraint is employed in training our networks. These experiments directly compare the performance of the different layer types described in Sections 2 and 3. These experiments are particularly useful for understanding the relative performance of these methods in settings where direct data consistency may not be desirable because the acquired data is corrupted by an artifact.

Data. In this experiment, we simulate artifacts of interest in a set of 6,276 T_1-weighted brain MRI images from patients aged 55-90 collected as part of the Alzheimer's Disease Neuroimaging Initiative (ADNI) (Mueller et al., 2005). We select the central 2D axial image of each volume for training and evaluation. To simulate acquired data, we apply the 2D Fourier transform to each image. After simulating the artifacts as described below, we normalize each input and output training pair by dividing by the maximum value in the corrupted image. The k-space data were zero-padded in this dataset during the original image reconstruction process, prior to our simulations. As a result, the quantitative results from these experiments do not represent model performance when deployed on raw, acquired k-space data (Shimron et al., 2022). Instead, these experiments probe the relative performance of competing methods on tasks for which large datasets of raw k-space are not readily available, such as motion correction and denoising. Subsequent experiments with raw, acquired frequency space data that have not been padded demonstrate that the proposed joint layers can also handle non-padded data. We split the dataset into 4,115 training images, 2,061 validation images, and 100 test images such that no subjects are shared across the training, validation, and test sets. Preliminary experiments and hyper-

parameters are evaluated on the validation dataset; the test set is only used for computing the performance statistics.

Training Loss and Evaluation Metric. We train `Frequency`, `Image`, `Interleaved`, and `Alternating` networks described in Section 3 using L1 loss on the real and imaginary components of the output and employ the SSIM scores (Wang et al., 2004) between the ground truth and reconstructed magnitude images to evaluate the quality of reconstruction on the test set.

4.1.1 Experimental Setup

Motion. Imaging subjects may move as measurements are being acquired at different points in the Fourier space. In practice, all points within a single line $F(\cdot, k_y)$ in frequency space are acquired rapidly together. Thus, it is commonly assumed that no motion occurs during acquisition of a single frequency space line. In this work, we use a rigid-body motion model for motion that occurs between acquisitions of successive lines.

If the imaged subject is affected by a rotation ϕ_{k_y} about the origin, a horizontal translation Δx_{k_y}, and a vertical translation Δy_{k_y} during acquisition of line k_y, the acquired signal corresponds to the rigidly transformed image $\tilde{I}_{k_y}$

$$\begin{aligned} \tilde{F}\left(\cdot, k_y\right) &= \mathscr{F}\left\{\tilde{I}_{k_y}\right\}\left(\cdot, k_y\right), \quad \text{where} \\ \tilde{I}_{k_y}\left(x, y\right) &= I\big(\left(x - \Delta x_{k_y}\right)\cos\phi_{k_y} - \left(y - \Delta y_{k_y}\right)\sin\phi_{k_y}, \\ &\qquad \left(x - \Delta x_{k_y}\right)\sin\phi_{k_y} + \left(y - \Delta y_{k_y}\right)\cos\phi_{k_y}\big). \end{aligned} \tag{8}$$

Eq. (8) forms a translated and rotated version of the desired image I. A pure translation without rotation in the image space corresponds to a phase shift in the frequency space:

$$\tilde{F}_t\left(k_x, k_y\right) = F\left(k_x, k_y\right)\exp\left\{-j2\pi\left(k_x\frac{\Delta x_{k_y}}{N} + k_y\frac{\Delta y_{k_y}}{N}\right)\right\} \tag{9}$$

for a $N \times N$ image. A pure rotation about the center of the image space without translation corresponds to a rotation by the same angle in the frequency space:

$$\begin{aligned} \tilde{F}_r\left(k_x, k_y\right) = F\big(&k_x\cos\phi_{k_y} - k_y\sin\phi_{k_y}, \\ &k_x\sin\phi_{k_y} + k_y\cos\phi_{k_y}\big). \end{aligned} \tag{10}$$

To simulate motion artifacts during image acquisition as described in Eq. (8), we sample three motion parameters at various lines in frequency space: a horizontal translation Δx, vertical translation Δy, and rotation ϕ. We report results for the case when the fraction γ_m of the total number of lines at which motion occurs is 0.03, though the trends in our results hold for several different values of this parameter. We apply the sampled motion parameters to contiguous lines in frequency space between consecutive motion line samples. Translation parameter values are drawn uniformly from the range $[-8\text{px}, 8\text{px}]$, corresponding to physical translations on the range $[-8\text{mm}, 8\text{mm}]$. Rotation parameter values are drawn uniformly from the range $[-11°, 11°]$. These parameter ranges are chosen to include extreme motion at the upper limit of what might be expected in a typical MRI scan. For a Cartesian, fully-sampled acquisition, the resulting combined frequency space data represents the signal acquired when the imaging subject shifts according to the sampled motion parameters at each of the randomly sampled lines in frequency space.

Noise. Noisy MRI data can be modeled via an additive i.i.d. complex Gaussian distribution:

$$\begin{aligned} \tilde{F}(k_x, k_y) &= F(k_x, k_y) + \epsilon_1 + j\epsilon_2, \\ \epsilon_1, \epsilon_2 &\sim \mathcal{N}(0, \sigma^2 \mathbb{I}_{N\times N}),\ \epsilon_1 \perp\!\!\!\perp \epsilon_2, \end{aligned} \tag{11}$$

where $\mathcal{N}(\mu, \Sigma)$ represents the Gaussian distribution with mean μ and covariance Σ. This noise distribution gives rise to the standard Rician distribution on MRI image space pixel magnitudes (Cárdenas-Blanco et al., 2008).

To simulate noisy acquisitions as described in Eq. (11), we sample pixelwise independent noise from a zero-mean Gaussian distribution. We report results in the case where this noise has standard deviation γ_n of 10,000, though our observed trends are consistent for both smaller and larger values of this parameter. This value was chosen because it visually results in an aggressive noise corruption on the magnitude image; the average resulting magnitude image has SNR$\approx$1.5.

Undersampling. To speed up image acquisition, a common approach is to only acquire data at a subset S_y of discrete "lines," i.e., values of $k_y \in S_y$:

$$\tilde{F}(k_x, k_y) = \begin{cases} F(k_x, k_y) & k_y \in S_y \\ 0 & k_y \notin S_y. \end{cases} \tag{12}$$

We simulate undersampling as described in Eq. (12) with sampling frequency $\gamma_s = 25\%$ (equivalent to an acceleration factor of 4), where the selected line indices S_y are sampled at random. These lines are selected without a bias toward the low-frequency lines at the center of the Fourier plane of each image, independently of the sampling pattern in all other images. This challenging undersampling pattern measures how well different layer architectures perform under non-traditional acquisition schemes, for example, when using scan-specific acquisition patterns (Bahadir et al., 2020). Our subsequent experiments evaluate the proposed layers with more conventional undersampling schemes. As an aside, the ground truth data in this experiment has conjugate symmetry in the frequency space, so in the hypothetical case of γ_s=50% with our random sampling scheme it is possible that all of the data required to perfectly reconstruct the image is present in the input. This is impossible for the acceleration factor of γ_s=25% in this study.

Undersampling with Motion. Undersampling reduces scan time and thus is commonly used to limit the time during which motion can occur. We analyze the setting where both motion corruption and undersampling occur simultaneously (Fig. 3), forcing the reconstruction algorithms to correct both types of artifacts. As in the pure motion experiments, for each slice, we set the fraction of lines $\gamma_m = 0.03$. For each line affected by motion, we sample three parameters of motion: Δx_i, Δy_i, and ϕ_i, corresponding respectively to a horizontal translation, vertical translation, and counterclockwise rotation about the slice origin. We simulate the corresponding motion-corrupted frequency space as described in Eq. (8). We then sample the full center 8% of k_y-lines and sample the remainder of the line indices from a uniform distribution to achieve an overall 4x acceleration factor.

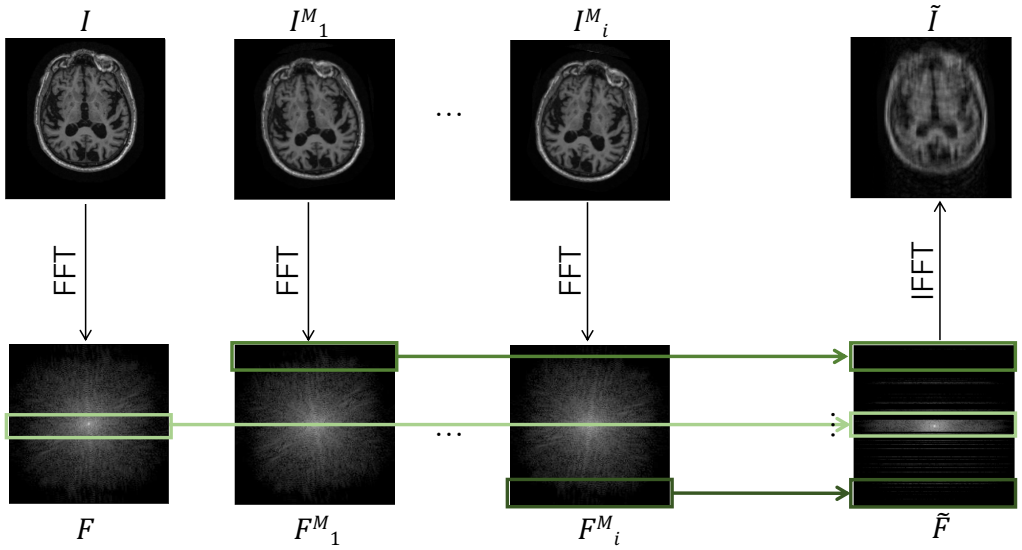

Figure 3: Data generation procedure for undersampling in the presence of motion. At line L_i in frequency space, the original image I is rotated and translated to form I_i^M. Lines from the corresponding Fourier transforms F and F_i^M are mixed and undersampled to generate motion-corrupted frequency space data $\tilde{F}$ that would have been acquired under the illustrated motion pattern. A similar method is used to simulate pure motion corruption without undersampling, where all frequency space lines are maintained to generate $\tilde{F}$.

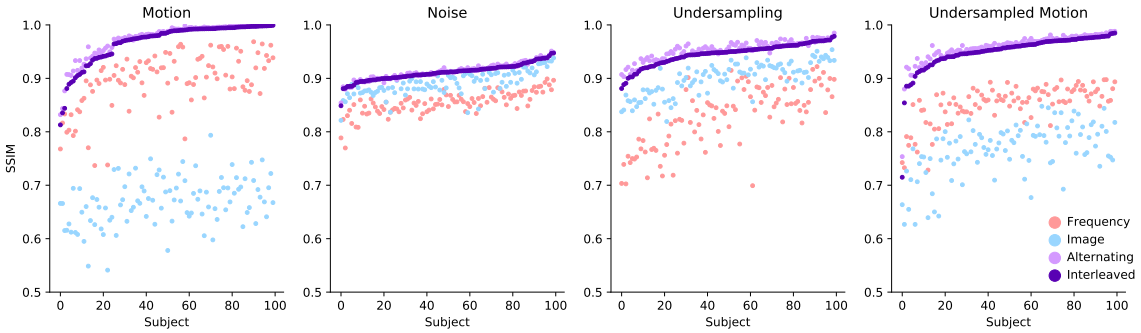

Figure 4: Subjectwise SSIM comparison for all brain MRI tasks without data consistency constraints. Subjects are sorted by performance of the `Interleaved` network. For all tasks, networks combining frequency and image space convolutions outperform single-domain networks.

4.1.2 Results

Fig. 4 reports reconstruction quality statistics for all four types of simulations described in Section 4.1.1: motion, noise, undersampling, and motion combined with undersampling. The `Interleaved` and `Alternating` architectures outperform the baseline architectures for nearly every task and subject. Across all tasks and nearly all subjects, the `Interleaved` and `Alternating` architectures are quite similar in numerical performance. Sample image reconstructions for the motion, motion with undersampling and denoising tasks are shown

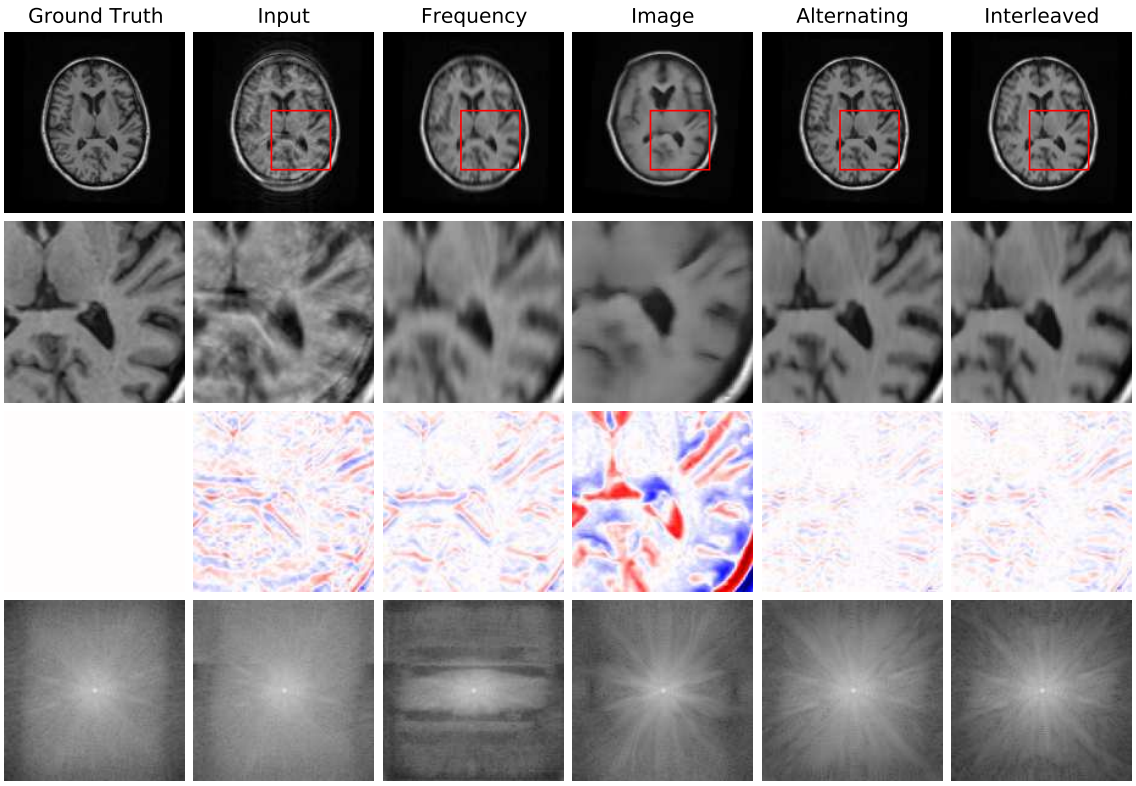

Figure 5: Example reconstructions with motion at 3% of scanning lines, zoomed-in image patches, difference patches between reconstructions and ground truth images, and frequency space reconstructions. The log values are taken of the frequency space data to better visualize its dynamic range. In the patch difference, red pixels have a higher value in the reconstruction than in the ground truth, while blue pixels have a lower value in the reconstruction than in the ground truth. The `Interleaved` and `Alternating` architectures more accurately eliminate the 'shadow' of the moved brain and the induced blurring compared to the single-domain networks.

in Figs. 5-7. Qualitatively, for each task, the `Frequency` network provides a blurry version of the ground truth image. The `Image` network provides a reconstruction which effectively removes 'background' effects but has limited success in correcting these artifacts within the image. In contrast, the `Interleaved` and `Alternating` networks provide sharper, high-quality reconstructions across all tasks. Further, the frequency space reconstructions provided by those networks appear the most faithful to the ground truth frequency data.

4.2 Hard Data Consistency Constraint

Deep learning for undersampled reconstruction is an active area of research and several state of the art methods have emerged for this task. In this experiment, we compare `Interleaved` and `Alternating` networks to such methods on ADNI data introduced in Section 4.1.

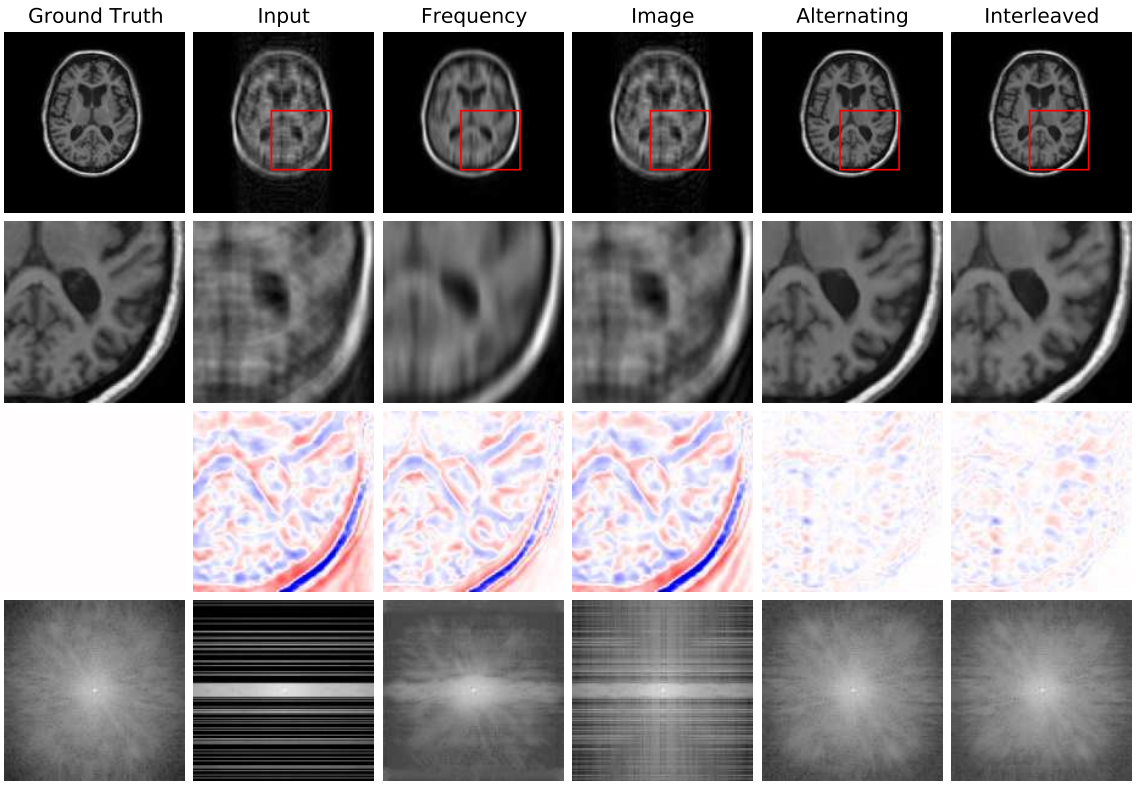

Figure 6: Example reconstructions from 4x undersampled, motion-corrupted data data, zoomed-in image patches, difference patches between reconstructions and ground truth images, and frequency space reconstructions. As in the motion corruption and undersampling examples, the `Interleaved` and `Alternating` architectures provide more accurate reconstructions of the ground truth images and reconstructing a more coherent k-space.

Undersampling is fundamentally different from motion and noise corruption, because the acquired data for lines $k_y \in S_y$ are the correct, desired outputs of the reconstruction algorithm at those frequency space locations. Data consistency can be enforced at test time and at intermediate layers of the network by substituting the appropriate k-space lines into the k-space representations of the image (final or intermediate) produced by the network. We enforce data consistency in `Interleaved` and `Alternating` networks by copying the acquired frequency space data into the network output.

We compare `Interleaved` and `Alternating` networks to a U-Net (Falk et al., 2019), the CascadeNet (Schlemper et al., 2017), which combines image space convolutions with forced data consistency at each layer of the network, and, most similar to our method, the KIKI network (Eo et al., 2018), which includes two separate image and frequency space networks. The KIKI-net architecture incorporates four networks operating in the frequency, image, frequency, and image spaces, respectively. This is in contrast to our networks, where every layer contains convolutions in both spaces and uses a custom nonlinearity for the frequency

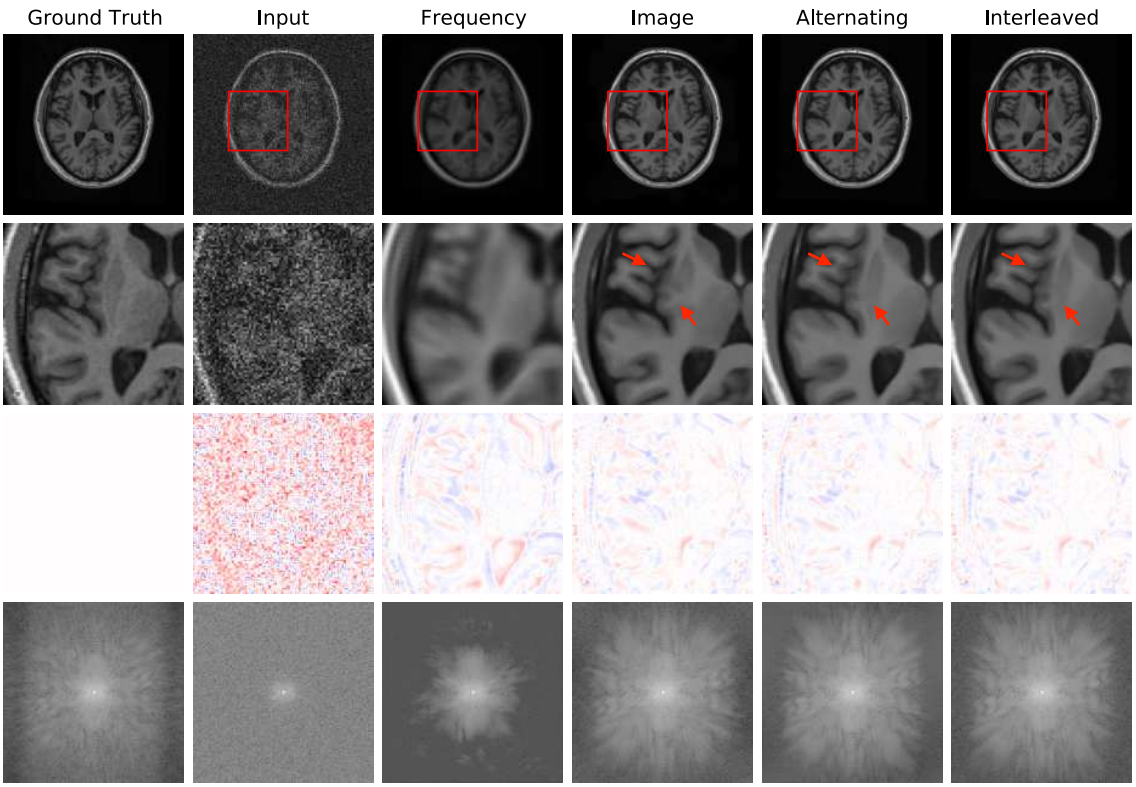

Figure 7: Example reconstructions with noise of standard deviation 10,000. The `Interleaved` and `Alternating` reconstructions remove the pixelated noise effect without over-smoothing, in contrast to the single-domain networks.

space layers. Moreover, the KIKI-net architecture imposes a data consistency constraint after each k-space subnetwork. For tasks other than undersampled image reconstruction, the data consistency constraints in CascadeNet and KIKI-net would incorrectly force the acquired k-space lines to be maintained in the final reconstruction; thus, we restrict comparisons with CascadeNet and KIKI-net to the undersampled reconstruction case.

We use implementations of the baseline methods available at `https://github.com/zaccharieramzi/fastmri-reproducible-benchmark` (Ramzi et al., 2020). We scale each network to have roughly 800,000 parameters for fair comparison with our joint architectures. We use an L1 loss function to train the networks and SSIM scores to evaluate their performance on the test set.

Undersampling patterns. In addition to the random sampling scheme in Section 4.1, we simulate two traditional undersampling patterns: (i) the central 8% of lines are fully sampled while every fourth line of the outer regions of k-space is sampled and (ii) the central 4% of lines are fully sampled while every eighth line of the outer regions of k-space is sampled.

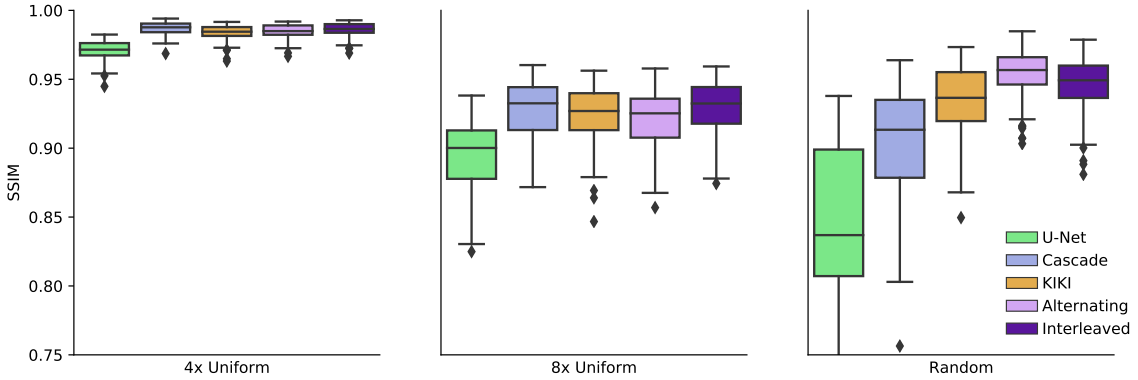

Figure 8: SSIM comparison of the joint networks with the state of the art undersampled reconstruction approaches on ADNI data. Results are reported for three undersampling patterns: 4x uniform undersampling with a fully-sampled central region (left), 8x uniform undersampling with a fully-sampled central region (middle), and 4x undersampling at random (right). In all cases, simple networks composed of repeated copies of our joint layers perform at least as well as other state of the art networks, and in the difficult case of a random sampling pattern, outperform the baseline networks.

4.2.1 Results

Fig. 8 reports statistics for U-Net, CascadeNet, KIKI-net, Joint and `Alternating` networks. Fig. 9 provides sample image reconstructions. `Interleaved` and `Alternating` networks perform comparably to other state of the art methods on the simpler uniform undersampling tasks and outperform the state of the art methods on the more complex random undersampling task.

4.3 Unrolled Optimization

Finally, we evaluate the performance of the proposed joint layers in the setting of an unrolled optimization architecture on real world multicoil MRI data. In this experiment, we replace the image space convolutional layers with our `Interleaved` layers in the MoDL framework (Aggarwal et al., 2018) for unrolled optimization. We use the authors' publicly available implementation of MoDL at `https://github.com/hkaggarwal/modl`. Each iteration of the MoDL network first passes the input through convolutional layers that serve as a data-driven regularizer and then applies an analytical update based on the data consistency term. To keep the total number of convolutions comparable, we train the baseline MoDL network with 10 image convolutional layers in each iteration and the joint MoDL network with 5 `Interleaved` layers in each iteration. We set $K = 5$ iterations for both networks. The authors use the strategy of first training a one-iteration MoDL network and using its weights to initialize the training of a multi-iteration MoDL network. This process speeds up training of the larger unrolled optimization network and avoids instabilities. We found that pre-training of a one-iteration model was unnecessary when using the joint layers, and train both the one-iteration and the five-iteration joint MoDL networks using

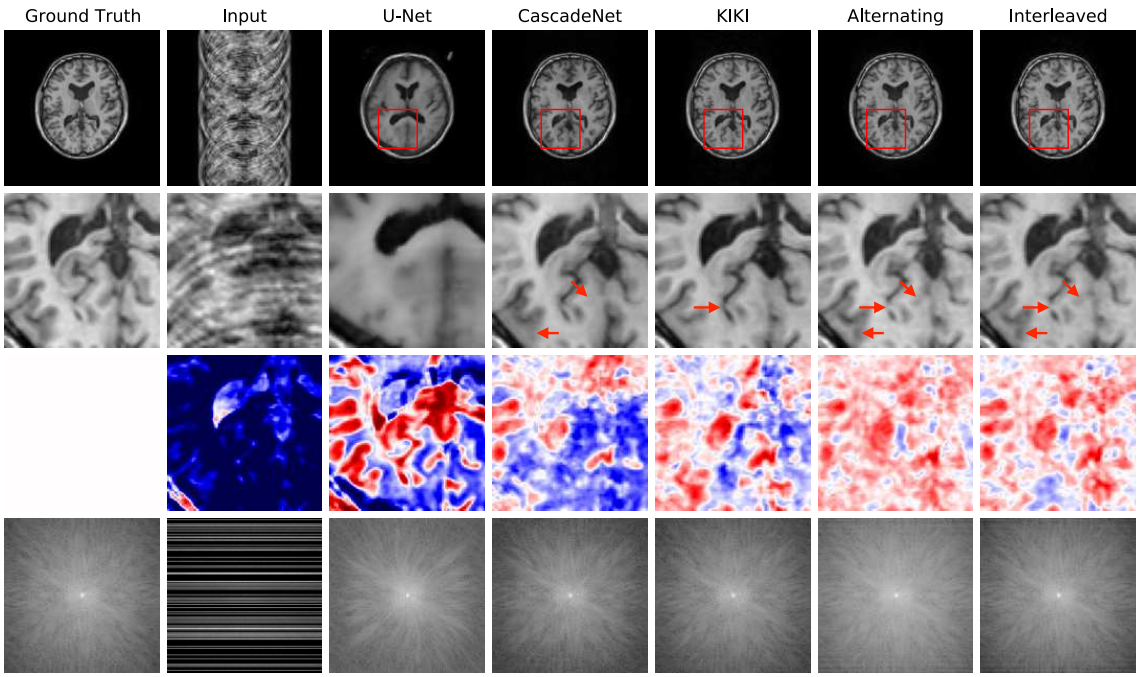

Figure 9: Example reconstructions from 4x undersampled data, with lines selected at random. The `Interleaved` and `Alternating` architectures provide more accurate reconstructions of the ground truth images, better eliminating 'ringing' and blurring artifacts.

random initializations. For consistency with the original MoDL training approach, we train all networks using L2 loss.

Data. We use the data from the original MoDL study (Aggarwal et al., 2018). This dataset contains raw k-space data from 3D T2 CUBE acquisitions with Cartesian readouts using a 12-channel head coil. The dataset contains 360 training slices from 4 training subjects and a single, separate test subject. We exclude some edge slices in this test volume and use the central 90 slices for our evaluations to match the training distribution. We train all networks using a variable density 6x undersampling mask as specified in the original paper.

4.3.1 Results

Figure 10 presents the training curves, validation SSIM, and sample reconstructions for all versions of the MoDL architecture. All networks attain similar validation SSIM values, but MoDL networks with joint layers achieve high reconstruction quality in roughly a third as many epochs as image space networks. Further, using our joint layers removes the need to pretrain a one-iteration network. The five-iteration network with joint layers trains successfully from random initializations. The resulting differences in wall clock training times are summarized in Table 1.

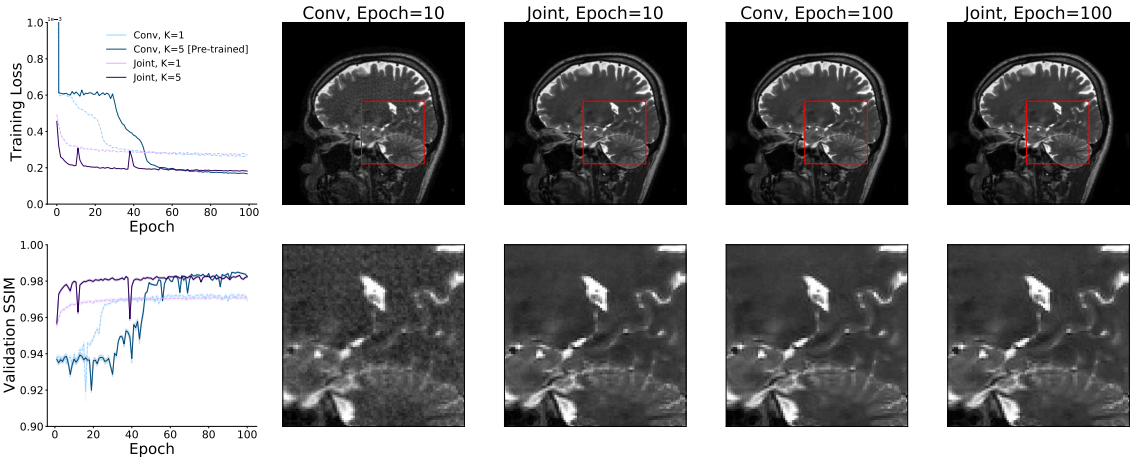

Figure 10: Example training loss and validation SSIM curves (left) and sample reconstructions and patches for MoDL networks with $K = 1, 5$ iterations trained with image convolutional layers and with the proposed joint (`Interleaved`) layers. MoDL networks with image convolutional layers do not converge if trained directly with $K = 5$. Instead, a $K = 1$ MoDL network must be trained and used to initialize the weights of a $K = 5$ MoDL network. MoDL networks trained with joint layers do not require pre-training and achieve the same loss and validation SSIM values as networks trained with image convolutions in significantly less time.

MoDL Layer	Pre-Training (Hrs)	Training (Hrs)	Total (Hrs)
Image Convolution	19	12	31
Joint Layer	0	4	4

Table 1: Training times for the full ($K = 5$) versions of the MoDL architecture to achieve validation SSIM≥ 0.98. For stable training, MoDL with image space convolutions must be initialized using the weights learned for a $K = 1$ MoDL network. MoDL architectures trained with our joint layers require no pre-training. In total, using joint layers results in roughly an 8x speed-up over the pure image space approach.

5. Discussion and Conclusions

We demonstrate the advantages of joint image and frequency space learning strategies for correcting corrupted MRI data. For tasks where data consistency constraints cannot be readily applied, our joint networks produce sharper reconstructions than the more blurry, artifacted versions generated by single space networks. For the well-studied task of undersampled reconstruction, where data consistency constraints can be imposed easily, we show that networks comprising joint layers can be trained with such constraints and compare favorably to other strategies that incorporate data consistency constraints to improve the quality of single space network reconstructions. For unrolled architectures that iteratively perform the steps of an optimization procedure to produce high quality reconstructions, the joint layers can straightforwardly replace image convolutional layers to improve train-

ing landscape and convergence. These results point to joint layers as a useful building block when designing neural network architectures for correcting frequency space artifacts.

While we demonstrate our method in a diverse set of acquisition scenarios, our analysis does not exhaustively cover all possible imaging artifacts. For example, we do not analyze the effects of interslice motion, which may occur in addition to the intraslice motion studied in this work and introduces new image content from an adjacent slice into the slice being imaged. Further, while we analyze extremely aggressive versions of motion, noise, and undersampling to demonstrate the effectiveness of our method in the most challenging scenarios, future versions of this method could tune these parameters to more closely match the statistics of the patient population being scanned. For example, empirically measured motion trajectories could be used to characterize the rate and severity of the induced motion artifacts.

In the future, we aim to develop additional strategies for applications where direct consistency with acquired data is not necessarily desirable, such as motion correction. We also plan to investigate local operations beyond convolutions that more directly capitalize on properties and symmetries of frequency space data for use in joint architectures. Local convolutions in the frequency space represent a subset of all possible element-wise multiplications in the image space. Thus, future work could perform these operations in the image space, saving the computational overhead of performing an FFT within each layer, or could take advantage of additional element-wise image space multiplications whose Fourier transforms are not bandlimited to the size of our filter kernels. The combination of these advances promises to significantly improve reconstruction and analysis of MRI data in the face of widely varying acquisition challenges and downstream applications.

Acknowledgments

The authors thank members of the Medical Vision Group at MIT CSAIL for useful discussions. This research was supported by NIBIB, NICHD, NIA, and NINDS of the National Institutes of Health under award numbers 5T32EB1680, P41EB015902, R01EB017337, R01HD100009, R01EB032708, 1R01AG064027-01A1, 5R01NS105820-02, 1R01AG070988-01 and 1RF1MH123195-01, by the European Research Council Starting Grant 677697, project "BUNGEE-TOOLS," by Alzheimer's Research UK (ARUK-IRG2019A-003), by an NSF Graduate Research Fellowship, and by a Google PhD Fellowship.

Ethical Standards

The work follows appropriate ethical standards in conducting research and writing the manuscript, following all applicable laws and regulations regarding treatment of animals or human subjects.

Conflicts of Interest

We declare we don't have conflicts of interest.

References

Hemant K Aggarwal, Merry P Mani, and Mathews Jacob. MoDL: Model-based deep learning architecture for inverse problems. *IEEE Transactions on Medical Imaging*, 38(2): 394–405, 2018.

Mehmet Akçakaya, Steen Moeller, Sebastian Weingärtner, and Kâmil Uğurbil. Scan-specific robust artificial-neural-networks for k-space interpolation (RAKI) reconstruction: Database-free deep learning for fast imaging. *Magnetic Resonance in Medicine*, 81(1): 439–453, 2019.

C Shyam Anand and Jyotinder S Sahambi. Wavelet domain non-linear filtering for MRI denoising. *Magnetic Resonance Imaging*, 28(6):842–861, 2010.

Jalal B Andre, Brian W Bresnahan, Mahmud Mossa-Basha, Michael N Hoff, C Patrick Smith, Yoshimi Anzai, and Wendy A Cohen. Toward quantifying the prevalence, severity, and cost associated with patient motion during clinical MR examinations. *Journal of the American College of Radiology*, 12(7):689–695, 2015.

Cagla D Bahadir, Alan Q Wang, Adrian V Dalca, and Mert R Sabuncu. Deep-learning-based optimization of the under-sampling pattern in MRI. *IEEE Transactions on Computational Imaging*, 6:1139–1152, 2020.

PG Batchelor, D Atkinson, P Irarrazaval, DLG Hill, J Hajnal, and D Larkman. Matrix description of general motion correction applied to multishot images. *Magnetic Resonance in Medicine*, 54(5):1273–1280, 2005.

Ariel Benou, Ronel Veksler, Alon Friedman, and Tammy Riklin Raviv. Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced MRI sequences. *Medical Image Analysis*, 42:145–159, 2017.

Arturo Cárdenas-Blanco, Cristian Tejos, Pablo Irarrazaval, and Ian Cameron. Noise in magnitude magnetic resonance images. *Concepts in Magnetic Resonance Part A: An Educational Journal*, 32(6):409–416, 2008.

Joseph Y Cheng, Morteza Mardani, Marcus T Alley, John M Pauly, and SS Vasanawala. DeepSPIRiT: Generalized parallel imaging using deep convolutional neural networks. In *Proc. 26th Annual Meeting of the ISMRM, Paris, France*, 2018.

Ben A Duffy, Lu Zhao, Farshid Sepehrband, Joyce Min, Danny JJ Wang, Yonggang Shi, Arthur W Toga, Hosung Kim, Alzheimer's Disease Neuroimaging Initiative, et al. Retrospective motion artifact correction of structural MRI images using deep learning improves the quality of cortical surface reconstructions. *Neuroimage*, 230:117756, 2021.

Taejoon Eo, Yohan Jun, Taeseong Kim, Jinseong Jang, Ho-Joon Lee, and Dosik Hwang. KIKI-net: Cross-domain convolutional neural networks for reconstructing undersampled magnetic resonance images. *Magnetic resonance in medicine*, 80(5):2188–2201, 2018.

Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, et al. U-net: Deep learning for cell counting, detection, and morphometry. *Nature methods*, 16(1): 67–70, 2019.

Mark A Griswold, Peter M Jakob, Robin M Heidemann, Mathias Nittka, Vladimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Haase. Generalized autocalibrating partially parallel acquisitions (GRAPPA). *Magnetic Resonance in Medicine*, 47(6):1202–1210, 2002.

Kerstin Hammernik, Teresa Klatzer, Erich Kobler, Michael P Recht, Daniel K Sodickson, Thomas Pock, and Florian Knoll. Learning a variational network for reconstruction of accelerated MRI data. *Magnetic Resonance in Medicine*, 79(6):3055–3071, 2018.

Yoseob Han, Leonard Sunwoo, and Jong Chul Ye. k-space deep learning for accelerated MRI. *IEEE Transactions on Medical Imaging*, 2019.

Melissa W Haskell, Stephen F Cauley, and Lawrence L Wald. TArgeted Motion Estimation and Reduction (TAMER): Data consistency based motion mitigation for MRI using a reduced model joint optimization. *IEEE Transactions on Medical Imaging*, 37(5):1253–1265, 2018.

Melissa W Haskell, Stephen F Cauley, Berkin Bilgic, Julian Hossbach, Daniel N Splitthoff, Josef Pfeuffer, Kawin Setsompop, and Lawrence L Wald. Network Accelerated Motion Estimation and Reduction (NAMER): Convolutional neural network guided retrospective motion correction using a separable motion model. *Magnetic Resonance in Medicine*, 82 (4):1452–1461, 2019.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1026–1034, 2015.

Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of PSNR in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008.

Chang Min Hyun, Hwa Pyung Kim, Sung Min Lee, Sungchul Lee, and Jin Keun Seo. Deep learning for undersampled MRI reconstruction. *Physics in Medicine & Biology*, 63(13): 135007, 2018.

Dongsheng Jiang, Weiqiang Dou, Luc Vosters, Xiayu Xu, Yue Sun, and Tao Tan. Denoising of 3D magnetic resonance images with multi-channel residual learning of convolutional neural network. *Japanese Journal of Radiology*, 36(9):566–574, 2018.

Patricia M Johnson and Maria Drangova. Conditional generative adversarial network for 3D rigid-body motion correction in MRI. *Magnetic Resonance in Medicine*, 82(3):901–910, 2019.

Tae Hyung Kim, Pratyush Garg, and Justin P Haldar. LORAKI: Autocalibrated recurrent neural networks for autoregressive MRI reconstruction in k-space. *arXiv preprint arXiv:1904.09390*, 2019.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Machine Learning*, 2014.

Thomas Küstner, Karim Armanious, Jiahuan Yang, Bin Yang, Fritz Schick, and Sergios Gatidis. Retrospective correction of motion-affected MR images using deep learning frameworks. *Magnetic Resonance in Medicine*, 82(4):1527–1540, 2019.

Paul C Lauterbur. Image formation by induced local interactions: Examples employing nuclear magnetic resonance. *Nature*, 242(5394):190–191, 1973.

Dongwook Lee, Jaejun Yoo, and Jong Chul Ye. Deep residual learning for compressed sensing MRI. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 15–18. IEEE, 2017.

Michael Lustig and John M Pauly. SPIRiT: Iterative self-consistent parallel imaging reconstruction from arbitrary k-space. *Magnetic resonance in medicine*, 64(2):457–471, 2010.

Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. Compressed sensing MRI. *IEEE Signal Processing Magazine*, 25(2):72–82, 2008.

Albert Macovski. Noise in MRI. *Magnetic Resonance in Medicine*, 36(3):494–497, 1996.

José V Manjón and Pierrick Coupe. MRI denoising using deep learning. In *International Workshop on Patch-based Techniques in Medical Imaging*, pages 12–19. Springer, 2018.

José V Manjón, José Carbonell-Caballero, Juan J Lull, Gracián García-Martí, Luís Martí-Bonmatí, and Montserrat Robles. MRI denoising using non-local means. *Medical Image Analysis*, 12(4):514–523, 2008.

Susanne G Mueller, Michael W Weiner, Leon J Thal, Ronald C Petersen, Clifford Jack, William Jagust, John Q Trojanowski, Arthur W Toga, and Laurel Beckett. The Alzheimer's Disease Neuroimaging Initiative. *Neuroimaging Clinics*, 15(4):869–877, 2005.

Robert D Nowak. Wavelet-based Rician noise removal for magnetic resonance imaging. *IEEE Transactions on Image Processing*, 8(10):1408–1419, 1999.

Ilkay Oksuz, James Clough, Bram Ruijsink, Esther Puyol-Antón, Aurelien Bustin, Gastao Cruz, Claudia Prieto, Daniel Rueckert, Andrew P King, and Julia A Schnabel. Detection and correction of cardiac MRI motion artefacts during reconstruction from k-space. In *International conference on medical image computing and computer-assisted intervention*, pages 695–703. Springer, 2019.

Kamlesh Pawar, Zhaolin Chen, N Jon Shah, and Gary F Egan. MoCoNet: Motion correction in 3D MPRAGE images using a convolutional neural network approach. *arXiv preprint arXiv:1807.10831*, 2018.

Klaas P Pruessmann, Markus Weiger, Markus B Scheidegger, and Peter Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magnetic Resonance in Medicine*, 42(5):952–962, 1999.

Patrick Putzky and Max Welling. Invert to learn to invert. In *Advances in Neural Information Processing Systems*, pages 446–456, 2019.

Tran Minh Quan, Thanh Nguyen-Duc, and Won-Ki Jeong. Compressed sensing MRI reconstruction using a generative adversarial network with a cyclic loss. *IEEE Transactions on Medical Imaging*, 37(6):1488–1497, 2018.

Zaccharie Ramzi, Philippe Ciuciu, and Jean-Luc Starck. Benchmarking MRI reconstruction neural networks on large public datasets. *Applied Sciences*, 10(5):1816, 2020.

Jo Schlemper, Jose Caballero, Joseph V Hajnal, Anthony N Price, and Daniel Rueckert. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *IEEE Transactions on Medical Imaging*, 37(2):491–503, 2017.

Richard Shaw, Carole H Sudre, Thomas Varsavsky, Sébastien Ourselin, and M Jorge Cardoso. A k-space model of movement artefacts: Application to segmentation augmentation and artefact removal. *IEEE transactions on medical imaging*, 39(9):2881–2892, 2020.

Efrat Shimron, Jonathan I Tamir, Ke Wang, and Michael Lustig. Implicit data crimes: Machine learning bias arising from misuse of public data. *Proceedings of the National Academy of Sciences*, 119(13):e2117203119, 2022.

Roberto Souza and Richard Frayne. A hybrid frequency-domain/image-domain deep network for magnetic resonance image reconstruction. In *2019 32nd SIBGRAPI Conference on Graphics, Patterns and Images (SIBGRAPI)*, pages 257–264. IEEE, 2019.

Jian Sun, Huibin Li, Zongben Xu, et al. Deep ADMM-Net for compressive sensing MRI. In *Advances in Neural Information Processing Systems*, pages 10–18, 2016.

Muhammad Usman, Siddique Latif, Muhammad Asim, Byoung-Dai Lee, and Junaid Qadir. Retrospective motion correction in multishot MRI using generative adversarial network. *Scientific Reports*, 10(1):1–11, 2020.

Guanhua Wang, Enhao Gong, Suchandrima Banerjee, John Pauly, and Greg Zaharchuk. Accelerated MRI reconstruction with dual-domain generative adversarial network. In *International Workshop on Machine Learning for Medical Image Reconstruction*, pages 47–57. Springer, 2019.

Zhou Wang, Eero P Simoncelli, and Alan C Bovik. Multiscale structural similarity for image quality assessment. In *The Thirty-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, volume 2, pages 1398–1402. IEEE, 2003.

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

Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, et al. DAGAN: Deep de-aliasing generative adversarial networks for fast compressed sensing MRI reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1310–1321, 2017.

Jure Zbontar, Florian Knoll, Anuroop Sriram, Matthew J Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J Geras, Joe Katsnelson, Hersh Chandarana, et al. fastMRI: An open dataset and benchmarks for accelerated MRI. *arXiv preprint arXiv:1811.08839*, 2018.

Bo Zhou and S Kevin Zhou. DuDoRNet: Learning a dual-domain recurrent network for fast MRI reconstruction with deep T1 prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4273–4282, 2020.

Bo Zhu, Jeremiah Z Liu, Stephen F Cauley, Bruce R Rosen, and Matthew S Rosen. Image reconstruction by domain-transform manifold learning. *Nature*, 555(7697):487–492, 2018.

Appendix A. FastMRI Experiments

We compare the `Interleaved` and `Alternating` networks with `Frequency` and `Image` baseline methods, as well as the top three methods submitted to the single coil track of the FastMRI challenge at `https://fastmri.org`.

A.1 Data

We train and evaluate all networks on the proton density knee MRI frequency space data from the single coil FastMRI Dataset (Zbontar et al., 2018). We train separate networks for signals acquired with and without fat suppression. We apply the FastMRI 4x undersampling scheme at both training and test time. The 4x undersampling scheme acquires all of the central 8% of lines and samples lines outside of the central region from a uniform distribution such that 25% of all lines are sampled in total. After undersampling the signals, we normalize each input and output training pair by dividing by the maximum value in the corrupted image. We use the standard FastMRI split of 34,742 training slices from 973 volumes and 7,135 validation slices from 199 volumes. No subjects are shared across these sets. We treat the FastMRI validation set as our test set and use it only for evaluation by comparing the network's output to the high quality fully sampled images provided as part of the FastMRI dataset.

A.2 Training Loss and Evaluation Metrics

We evaluate and compare the networks trained with a variety of loss functions and assess reconstruction quality via different quality metrics. We train `Frequency`, `Image`, `Interleaved` and `Alternating` networks with seven loss functions: image space L1 error, frequency space L1 error, a joint L1 metric summing image and frequency L1 errors, SSIM (Wang et al., 2004), multiscale SSIM (Wang et al., 2003), and PSNR (Huynh-Thu and Ghanbari, 2008). The joint L1 metric weighs the frequency space L1 error by 0.1 relative to the image space L1 error to account for differences in the error magnitudes. The SSIM and multiscale SSIM scores are computed with window size 7×7 and constants $k_1 = 0.01$, $k_2 = 0.03$.

We also compare the joint networks with top single coil methods on the FastMRI benchmark. For these experiments, we use a larger version of the `Interleaved` network comprised of 6 joint layers with two frequency space and two image space convolutions per layer, yielding roughly 3 million parameters total.

A.3 Results

Our results on the knee undersampled reconstruction task replicate the trends observed in the brain undersampled reconstruction task. Joint networks outperform single-domain networks, as reported in Table 2. This suggests that our joint layers can successfully process acquired, complex-valued MRI data. Further, Table 2 confirms that the success of joint learning is not specific to a certain loss landscape. Qualitative examples of reconstructions from networks trained with various loss functions are shown in Fig. 11.

The reconstructed images produced by the larger `Interleaved` network are qualitatively similar to those produced by the top three methods on the FastMRI leaderboard (Fig. 12).

Table 3 reports reconstruction quality measures for `Interleaved` network and the top single-slice methods on the FastMRI benchmark. `Interleaved` network achieves results that are close to the state of the art architectures specifically tuned for this task. We emphasize that our goal is not to attain state of the art performance on the FastMRI benchmark, but rather to show that simple layers comprised of both frequency and image space convolutions achieve reasonable performance on this benchmark while offering flexibility for correcting a wide range of other artifacts, and for correcting multiple artifacts present simultaneously.

Loss	Architecture	Freq L1 (↓)	Image L1 (↓)	Joint L1 (↓)	SSIM (↑)	MS SSIM (↑)	PSNR (↑)
Freq L1	Frequency	5.2 ± 1.4	0.089 ± 0.063	0.61 ± 0.19	0.60 ± 0.12	0.76 ± 0.06	19.8 ± 2.4
	Image	8.9 ± 3.4	0.079 ± 0.060	0.97 ± 0.40	0.70 ± 0.14	0.88 ± 0.06	22.4 ± 3.3
	Interleaved	**3.9 ± 1.5**	**0.040 ± 0.018**	**0.43 ± 0.16**	**0.73 ± 0.11**	**0.91 ± 0.05**	**26.6 ± 2.4**
	Alternating	4.1 ± 1.5	0.095 ± 0.023	0.51 ± 0.16	0.58 ± 0.10	0.77 ± 0.09	20.3 ± 1.9
Image L1	Frequency	7.9 ± 2.5	0.040 ± 0.015	0.83 ± 0.27	0.69 ± 0.12	0.88 ± 0.06	26.8 ± 2.2
	Image	21.9 ± 8.2	0.054 ± 0.034	2.24 ± 0.85	0.59 ± 0.14	0.85 ± 0.09	24.9 ± 2.7
	Interleaved	**6.9 ± 2.7**	**0.031 ± 0.018**	**0.72 ± 0.28**	**0.78 ± 0.12**	**0.92 ± 0.06**	**28.9 ± 2.5**
	Alternating	7.5 ± 2.9	0.032 ± 0.013	0.78 ± 0.31	0.76 ± 0.12	0.91 ± 0.06	28.5 ± 2.4
Joint L1	Frequency	5.2 ± 1.7	0.062 ± 0.068	0.58 ± 0.23	0.66 ± 0.15	0.86 ± 0.06	23.1 ± 3.1
	Image	8.9 ± 3.4	0.055 ± 0.060	0.95 ± 0.39	0.70 ± 0.14	0.88 ± 0.06	25.4 ± 3.4
	Interleaved	**3.9 ± 1.5**	**0.032 ± 0.019**	**0.43 ± 0.17**	**0.77 ± 0.12**	**0.92 ± 0.05**	**27.8 ± 2.4**
	Alternating	4.1 ± 1.5	0.035 ± 0.020	0.44 ± 0.17	0.75 ± 0.12	0.91 ± 0.05	26.8 ± 2.5
-SSIM	Frequency	7.3 ± 1.8	0.039 ± 0.018	0.76 ± 0.20	0.73 ± 0.12	0.90 ± 0.05	26.5 ± 2.4
	Image	12.0 ± 4.3	0.058 ± 0.059	1.25 ± 0.49	0.69 ± 0.14	0.87 ± 0.07	24.7 ± 3.3
	Interleaved	**6.4 ± 2.4**	**0.029 ± 0.015**	**0.67 ± 0.25**	**0.80 ± 0.12**	**0.94 ± 0.05**	**29.0 ± 2.3**
	Alternating	7.4 ± 2.8	0.031 ± 0.012	0.77 ± 0.29	0.79 ± 0.13	0.93 ± 0.06	27.7 ± 2.1
-MS SSIM	Frequency	9.3 ± 1.5	0.043 ± 0.023	0.98 ± 0.16	0.69 ± 0.14	0.91 ± 0.05	25.3 ± 2.0
	Image	15.6 ± 7.7	0.061 ± 0.045	1.63 ± 0.81	0.61 ± 0.13	0.86 ± 0.07	24.0 ± 3.5
	Interleaved	**8.6 ± 1.8**	**0.030 ± 0.017**	**0.89 ± 0.19**	**0.79 ± 0.12**	**0.94 ± 0.05**	**27.5 ± 1.9**
	Alternating	15.0 ± 4.3	0.031 ± 0.014	1.54 ± 0.44	**0.79 ± 0.12**	**0.94 ± 0.05**	23.8 ± 1.8
-PSNR	Frequency	8.0 ± 2.5	0.038 ± 0.012	0.84 ± 0.26	0.70 ± 0.12	0.89 ± 0.06	27.4 ± 2.2
	Image	15.3 ± 7.1	0.058 ± 0.043	1.59 ± 0.75	0.64 ± 0.13	0.85 ± 0.07	24.0 ± 2.6
	Interleaved	**7.3 ± 2.8**	**0.031 ± 0.012**	**0.76 ± 0.29**	**0.77 ± 0.12**	**0.92 ± 0.06**	**29.1 ± 2.2**
	Alternating	9.1 ± 4.7	0.037 ± 0.016	0.95 ± 0.49	0.70 ± 0.13	0.89 ± 0.08	27.6 ± 2.4

Table 2: Image reconstruction evaluation metrics (columns) for networks trained with varying loss functions (rows) on images acquired without fat suppression. Similar trends hold for images with fat suppression. MS SSIM stands for multiscale SSIM. For metrics labeled ↓, smaller values are better; for metrics labeled ↑, larger values are better. Across nearly every training loss function and metric, the `Interleaved` network performs best. In almost every case, the `Alternating` network architecture performs similarly or only slightly worse than the `Interleaved` network. This is particularly true in the case of SSIM-based loss functions, which provide the best overall quantitative results across all evaluation metrics.

Figure 11: Typical image reconstruction results for all architectures (rows) and loss functions (columns) on FastMRI images without fat suppression. The `Interleaved` and `Alternating` networks provide the sharpest reconstructions for all loss functions. Amongst these, both SSIM-based loss functions most sharply reconstruct high frequency structures within the zoomed-in patch. Similar results are observed in images with fat suppression.

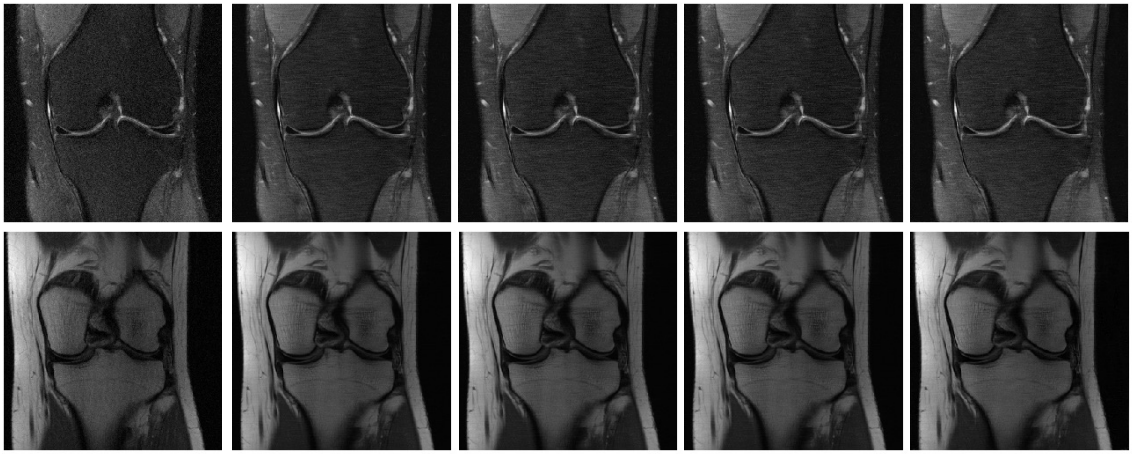

Figure 12: Comparison of the `Interleaved` reconstruction results with the top methods on the FastMRI single coil knee reconstruction challenge. All images were taken from the FastMRI online submission website. Our method produces a reconstruction qualitatively similar to those of the top three methods on the leaderboard.

Method	**MAE**	**SSIM**	**PSNR**
Interleaved (Ours)	0.0296	0.768	32.9
AIRS-Net	0.0266	0.784	33.8
SubtleMR	0.0270	0.781	33.7
i-RIM	0.0271	0.781	33.7

Table 3: Reconstruction quality statistics on the FastMRI leaderboard test dataset, at 4x undersampling. The FastMRI dataset contains images both with and without fat suppression. Simple `Interleaved` network comprised of joint layers is comparable to the three top models on the FastMRI leaderboard, yielding reconstructions with SSIM within 3% of the leading methods.