Joint Frequency and Image Space Learning for Fourier Imaging

Nalini M. Singh
CSAIL, MIT
nmsingh@mit.edu

Juan Eugenio Iglesias
MGH, HMS, CMIC, UCL, & CSAIL, MIT
e.iglesias@ucl.ac.uk

Elfar Adalsteinsson
RLE, IMES, EECS, MIT
elfar@mit.edu

Adrian V. Dalca
MGH, HMS & CSAIL, MIT
adalca@mit.edu

Polina Golland
CSAIL, MIT
polina@csail.mit.edu

Abstract

We demonstrate that neural network layers that explicitly combine frequency and image feature representations are a versatile building block for analysis of imaging data acquired in the frequency space. Our work is motivated by the challenges arising in MRI acquisition where the signal is a corrupted Fourier transform of the desired image. The joint learning schemes proposed and analyzed in this paper enable both correction of artifacts native to the frequency space and manipulation of image space representations to reconstruct coherent image structures. This is in contrast to most current deep learning approaches for image reconstruction that apply learned data manipulations solely in the frequency space or solely in the image space. We demonstrate the advantages of joint convolutional learning on three diverse tasks: image reconstruction from undersampled acquisitions, motion correction, and image denoising in brain and knee MRI. We further demonstrate advantages of the joint learning approaches across training schemes using a wide variety of loss functions. Unlike purely image based and purely frequency based architectures, the joint models produce consistently high quality output images across all tasks and datasets. Joint image and frequency space feature representations promise to significantly improve modeling and reconstruction of images acquired in the frequency space. Our code is available at https://github.com/nalinimsingh/interlacer.

1. Introduction

A wide range of imaging modalities acquire frequency space data and convert these measurements to images for visualization and downstream analysis, including magnetic resonance imaging (MRI) [25], Fourier space optical coherence tomography [27], Fourier ptychography [41], and synthetic aperture radar [45]. Practical imaging considerations often affect these data acquisition processes. For example, sub-Nyquist undersampling is routinely used to speed up data acquisition [28], motion occurs during acquisition [4], and noise affects sensor readings [30]. The acquired frequency space signals are typically 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 to produce coherent structure.

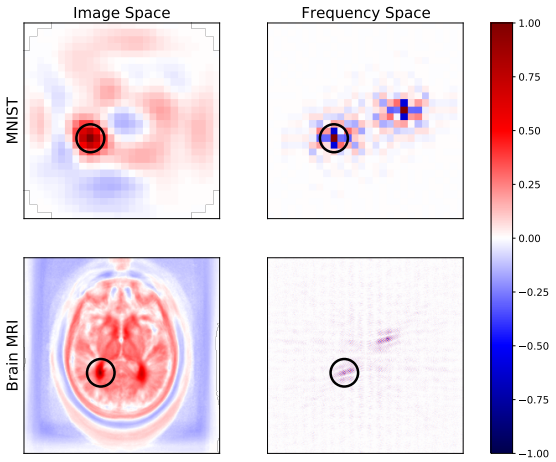

Figure 1. Maps of correlation coefficients between a single pixel (circled) and all other pixels in image and frequency space representations of MNIST (top) and a brain MRI dataset (bottom). All maps show strong local correlations useful for inferring missing or corrupted data. Frequency space correlations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

To that end, we consider the correlation structure for both

frequency and image space representations for a particular pixel of MNIST and brain MRIs, as illustrated in Fig. 1. Local neighborhoods around the pixel exhibit strong correlations, suggesting that local convolution operations, which have proven successful on image space computer vision tasks, might also be useful when applied to frequency space data. Convolutional operations in frequency space enable direct correction of local frequency space artifacts corresponding to global image space effects. However, resulting imprecisions in a frequency space representation yield jarring visual artifacts in the corresponding image space representation. Convolutional image space processing facilitates complementary correction of these artifacts.

We investigate joint networks that learn in both the frequency and image space. Our contributions are as follows:

- We describe two task-independent layer structures that jointly learn image and frequency space convolutions.
- We demonstrate that joint networks outperform pure image or pure frequency space networks for correcting a diverse set of MRI data corruptions: (1) aggressive undersampling, (2) extreme motion, and (3) heavy noise. This finding holds when reconstructing from real-world undersampled MRI signals.
- We demonstrate that the superiority of joint networks holds across a wide range of loss functions and evaluation metrics, suggesting that the success of joint networks is not tied to a particular loss landscape.

Taken together, our results provide a broadly applicable insight for practitioners: joint networks combining both image and frequency space convolutions should be the starting point when designing neural network architectures for correcting and analyzing a wide variety of imaging artifacts induced in the frequency space.

2. Background and Related Work

MRI acquires Fourier transform measurements. For each 2D slice in an MRI volume, the goal of image reconstruction is to generate an image I from the acquired discrete Fourier transform measurements $F = \mathscr{F}\{I\}$. Classically, this reconstruction is computed via a 2D inverse Fourier transform, producing estimated image $\hat{I} = \mathscr{F}^{-1}\{F\}$. Many strategies exist for selecting which measurements to acquire in frequency space. Here we consider Cartesian sampling, where measurement coordinates k_x and k_y are evenly sampled across the 2D Fourier plane. In this section, we describe three processes by which corrupted measurements $\tilde{F}$ are acquired instead of F and review prior methods for estimating the desired image I from $\tilde{F}$.

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

$$\tilde{F}[k_x, k_y] = \begin{cases} F[k_x, k_y] & k_y \in S_y \\ 0 & k_y \notin S_y. \end{cases} \quad (1)$$

Classical image reconstruction techniques for undersampled data vary in their choice of operating domain for performing reconstruction. SENSE reconstruction reduces the problem to least-squares estimation of the image space from undersampled frequency space data [36]. GRAPPA [12] and SPIRiT [29] apply convolutions in Fourier space to estimate missing lines, and then apply the inverse Fourier transform to reconstruct the image. Most deep learning methods apply convolutions to image space reconstructions of the input frequency space data [1, 13, 19, 26, 37, 38, 39, 40, 46]. A few recent methods apply convolutional architectures directly to frequency space data [2, 9, 14]. One method separately trains two pure frequency space networks and two pure image space networks and subsequently applies them sequentially after training [11]. An alternative strategy, AUTOMAP, uses fully-connected layers to effectively convert frequency space data to the image space and then applies further image space convolutions [50]. The size of such a network is quadratic in the image size, incurring immense memory costs for reconstructing larger images ($\mathcal{O}(N^4)$ for an $N \times N$ image). In contrast, our joint learning schemes use only convolutions and Fourier transforms and incur a memory cost that is $\mathcal{O}(N^2 \log N)$ for an $N \times N$ image.

Most relevant to our work, a recent method also incorporates both frequency and image space processing blocks within the same network [49]. In contrast to our networks, the technique requires a fully-sampled auxiliary image of a different MRI contrast. While such imaging is sometimes available in the research setting, we demonstrate the utility of networks combining frequency and image space convolutions on a range of varied tasks *without* access to any auxiliary images, as is common in clinical settings. Further, while the network in [49] is designed carefully for a particular task, we focus on identifying simple network layer structures that provide a generally useful starting point for a much broader set of tasks. Task-specific insights (e.g., a data consistency constraint) could be further applied to the basic joint architectures presented here.

Motion. 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 and that rigid-body motion occurs between acquisition of successive lines.

If the imaging 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

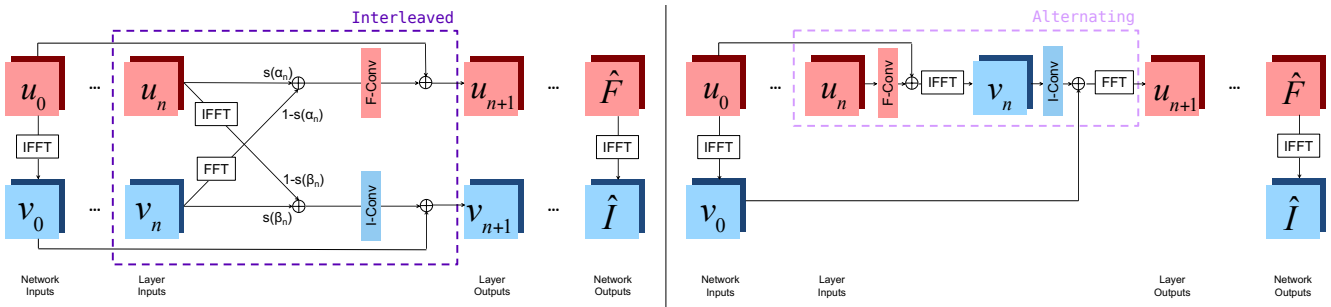

Figure 2. The `Interleaved` (left) and `Alternating` (right) layers, embedded within full network architectures. Each 'F-Conv' or 'I-Conv' block applies Batch Normalization (BN), a convolution, and an activation function in the frequency or image space, respectively.

signal corresponds to the rigidly transformed image $\tilde{I}_{k_y}$:

$$\tilde{F}[\cdot, k_y] = \mathscr{F}\left\{\tilde{I}_{k_y}\right\}[\cdot, k_y], \quad \text{where} \tag{2}$$

$$\tilde{I}_{k_y}[x, y] = I\big[\left(x - \Delta x_{k_y}\right)\cos\phi_{k_y} - \left(y - \Delta y_{k_y}\right)\sin\phi_{k_y}, \\ \left(x - \Delta x_{k_y}\right)\sin\phi_{k_y} + \left(y - \Delta y_{k_y}\right)\cos\phi_{k_y}\big].$$

Eq. (2) 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:

$$\tilde{F}_t[k_x, k_y] = F[k_x, k_y]\exp\left\{-j2\pi\left(k_x\frac{\Delta x_{k_y}}{N} + k_y\frac{\Delta y_{k_y}}{N}\right)\right\} \tag{3}$$

for an $N \times N$ 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:

$$\tilde{F}_r[k_x, k_y] = F\big[k_x\cos\phi_{k_y} - k_y\sin\phi_{k_y}, \\ k_x\sin\phi_{k_y} + k_y\cos\phi_{k_y}\big]. \tag{4}$$

Previous retrospective motion correction strategies [6, 16] are cast as large, non-convex optimization problems with iterative solutions that are slow to compute. More recent deep learning approaches [10, 15, 21, 24, 35, 42] solve the motion correction problem with a neural network operating purely in image space, even though motion artifacts are induced directly in frequency space during data acquisition.

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

$$\tilde{F}[k_x, k_y] = F[k_x, k_y] + \epsilon_1 + j\epsilon_2, \\ \epsilon_1, \epsilon_2 \sim \mathcal{N}(0, \sigma^2\mathbb{I}_{N\times N}),\ \epsilon_1 \perp\!\!\!\perp \epsilon_2, \tag{5}$$

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 [8]. Previous work on MRI denoising applies classical signal processing techniques including filtering [31] and wavelet-based methods [3, 34], while deep learning methods employ convolutional networks solely on image space data [7, 20, 32].

3. Joint Networks

We use a neural network to estimate a complex ground truth image I from a complex, corrupted acquired frequency space signal $\tilde{F}$. In this section, we describe two layer variants combining image and frequency space features, referred to as `Interleaved` and `Alternating`, and specify the network architectures and learning procedure.

3.1. 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 $s(\cdot)$ to constrain the mixing coefficients to range from 0 to 1:

$$\hat{u}_n = s(\alpha_n)\,u_n + (1 - s(\alpha_n))\,\mathscr{F}\{v_n\}, \\ \hat{v}_n = s(\beta_n)\,v_n + (1 - s(\beta_n))\,\mathscr{F}^{-1}\{u_n\}. \tag{6}$$

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:

$$u_{n+1} = \sigma(w_n \circledast \text{BN}(\hat{u}_n) + b_n) + u_0, \\ v_{n+1} = \sigma'(w'_n \circledast \text{BN}(\hat{v}_n) + b'_n) + v_0, \tag{7}$$

where (w_n, b_n) are learned frequency space convolution weights and biases, (w'_n, b'_n) are their image space counterparts, 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 structure 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.

In the `Alternating` setup, each layer sequentially incorporates frequency and image space convolutions with the appropriate batch normalization and activation function (Fig. 2, right):

$$\begin{aligned} v_n &= \mathscr{F}^{-1}\left\{\sigma(w_n \circledast \text{BN}(u_n) + b_n) + u_0\right\}, \\ u_{n+1} &= \mathscr{F}\left\{\sigma'(w'_n \circledast \text{BN}(v_n) + b'_n) + v_0\right\}, \end{aligned} \tag{8}$$

i.e., the reconstruction alternates between convolutions in the frequency and image space.

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, yielding $\mathcal{O}(N^2 \log N)$ space complexity for $N \times N$ images.

3.2. Activation Functions

Building on standard image space neural networks that often use the ReLU nonlinearity, we set $\sigma'(x) = \text{ReLU}(x)$ for all convolutions in the image space. 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:

$$\sigma(x) = x + ReLU\left(\frac{x-1}{2}\right) + ReLU\left(-\frac{x+1}{2}\right). \tag{9}$$

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 (see Supplementary Material).

3.3. 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 networks of the form $f(\cdot; \theta_f, \theta_i)$ 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:

$$\left(\theta_f^*, \theta_i^*\right) = \arg\min_{(\theta_f, \theta_i)} \sum_{m=1}^{|\mathcal{D}|} \mathcal{L}\left(I_m, \mathscr{F}^{-1}\left(f(\tilde{F}_m; \theta_f, \theta_i)\right)\right), \tag{10}$$

where θ_f and θ_i depend on the setup of the joint layer.

4. Implementation Details and Baseline Models

We construct each joint network to contain 8 joint frequency and image space layers. A single 2D convolutional layer acts on the frequency space output u_8 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 9x9 and 32 output features, resulting in a total of 1,178,202 parameters for the `Interleaved` network and 1,256,006 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 focus on comparing performance of the `Interleaved` and `Alternating` architectures to two similarly structured baseline architectures with only frequency or only image space operations, instead of comparing our network to task-specific architectures. Task-specific network design strategies, such as cascading [39], can be easily integrated with either of the joint network architectures for downstream applications.

First, we create an architecture `Frequency` that performs convolutions only on frequency space data and train it to identify frequency space parameters:

$$\theta_f^* = \arg\min_{\theta_f} \sum_{m=1}^{|\mathcal{D}|} \mathcal{L}\left(I_m, \mathscr{F}^{-1}\left(g(\tilde{F}_m; \theta_f)\right)\right). \tag{11}$$

The network contains 16 convolution blocks to match the joint networks' 8 pairs of 2 convolution blocks. As in the `Interleaved` and `Alternating` networks, each convolution block has kernel size 9x9 and 32 output features, followed by the final, two-feature 2D convolutional layer, resulting in 1,256,006 parameters. This network captures the frequency-only baseline methods from [2, 14, 22] for the purpose of our analysis.

We also implement an image space network `Image`, trained by optimizing:

$$\theta_i^* = \arg\min_{\theta_i} \sum_{m=1}^{|\mathcal{D}|} \mathcal{L}\left(I_m, g\left(\mathscr{F}^{-1}\left(\tilde{F}_m\right); \theta_i\right)\right). \tag{12}$$

This network's architecture is exactly identical to that of `Frequency` and also contains 1,256,006 parameters, but it operates on image space data. By applying convolutions to image space representations, the network captures image-based baseline methods [1, 13, 15, 19, 24, 26, 32, 35, 37, 38, 39, 40, 46] for the purpose of our analysis.

We initialize all convolution weights using the He normal initializer [17] and use the Adam optimizer [23] (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 is available at https://github.com/nalinimsingh/interlacer.

5. Experiments

5.1. Data

We demonstrate the advantages of the joint methods on a large collection of T_1-weighted brain MRI images from patients aged 55-90 collected as part of the Alzheimer's Disease Neuroimaging Initiative (ADNI) [33]. For training and evaluation, we select the central 2D axial image of each volume. To simulate acquired data, we apply the 2D Fourier transform to each image. After simulating the artifacts described in Eqs. 1, 2, and 5, we normalize each input and output training pair by dividing by the maximum value in the corrupted image. We split the dataset into 4,115 training images, 2,061 validation images, and 96 test images such that that no subjects are shared across the training, validation, and test sets. Preliminary experiments and hyperparameters are evaluated on the validation dataset; the test set is only used for computing the performance statistics.

We also demonstrate the advantages of joint learning on real proton-density knee MRI frequency space data from the single-coil FastMRI Dataset [47]. We train separate networks for signals acquired with and without fat suppression. After applying a standard undersampling pattern described below, 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 only use it for the final evaluation by comparing the network's output to the high quality fully-sampled multi-coil images provided as part of the FastMRI dataset.

5.2. Experimental Setup

Brain Undersampled Reconstruction To simulate undersampling as described in Eq. 1, we set the sampling frequency γ_s to be 33%, 25%, or 10% (equivalent to an acceleration factor of 3, 4, or 10, respectively). The selected line indices S_y are sampled at random, 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 acquisition scheme degrades the low frequency structure of the corresponding images with varying degrees of severity unlike previous work [13, 39], where the center rows of the Fourier plane are sampled more densely than the rest of the image. We analyze the more common center-dense undersampling scheme on the FastMRI dataset later in this section. The more challenging undersampling pattern studied here measures how well different layer architectures perform with non-traditional acquisition schemes, for example, when reconstructing data obtained with a scan-specific acquisition pattern [5]. The ground truth data used in this experiment has conjugate symmetry in frequency space, so in the hypothetical case of γ_s=50% with our random sampling scheme it is possible, but not guaranteed, that all of the data required to perfectly reconstruct the image is present in the input. This is impossible for the acceleration factors we consider.

Brain Motion Correction. To simulate motion artifacts during image acquisition as described in Eq. 2, we sample three motion parameters at various lines in frequency space: a horizontal translation Δx, vertical translation Δy, and rotation ϕ. We vary the fraction γ_m of the total number of lines at which motion occurs to be either 0.01, 0.03, or 0.05. 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 $[-20\text{px}, 20\text{px}]$. Rotation parameter values are drawn uniformly from the range $[-15°, 15°]$. 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 combined frequency space data at the end of this process 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.

Brain Denoising. To simulate noisy acquisitions as described in Eq. 5, we sample pixelwise independent noise from a zero-mean Gaussian with standard deviation γ_n as 5000, 10000, or 15000.

Knee Undersampled Reconstruction. Beyond the diverse and aggressive corruption models considered in the brain MRI experiments, we analyze the performance of the four architectures on real-world, scanner-acquired knee MRI frequency signals. In particular, we apply the FastMRI 4x and 8x undersampling schemes to single-coil knee MRI data. The 4x undersampling scheme acquires all of the central 8% of lines and uniformly samples lines outside of the central region such that 25% of all lines are sampled in total. The 8x undersampling scheme acquires all of the central 4% of lines and uniformly samples lines outside of the central region such that 12.5% of all lines are sampled in total.

5.3. Training Loss and Evaluation Metrics

For all experiments, we compare the reconstructed images to the ground truth using absolute (L1) error on the real and imaginary parts of the image space reconstruction as a metric of reconstruction quality. We also perform an

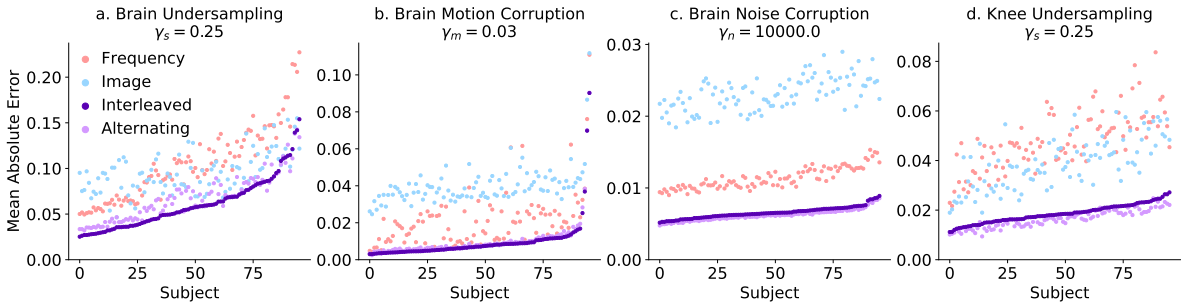

Figure 3. Subjectwise mean absolute error comparison for all tasks at a particular corruption level. Subjects are sorted by performance of the `Interleaved` network. For all tasks, networks combining frequency and image space convolutions outperform single-domain networks. Plots for other corruption levels are provided in the Supplementary Material and show similar trends.

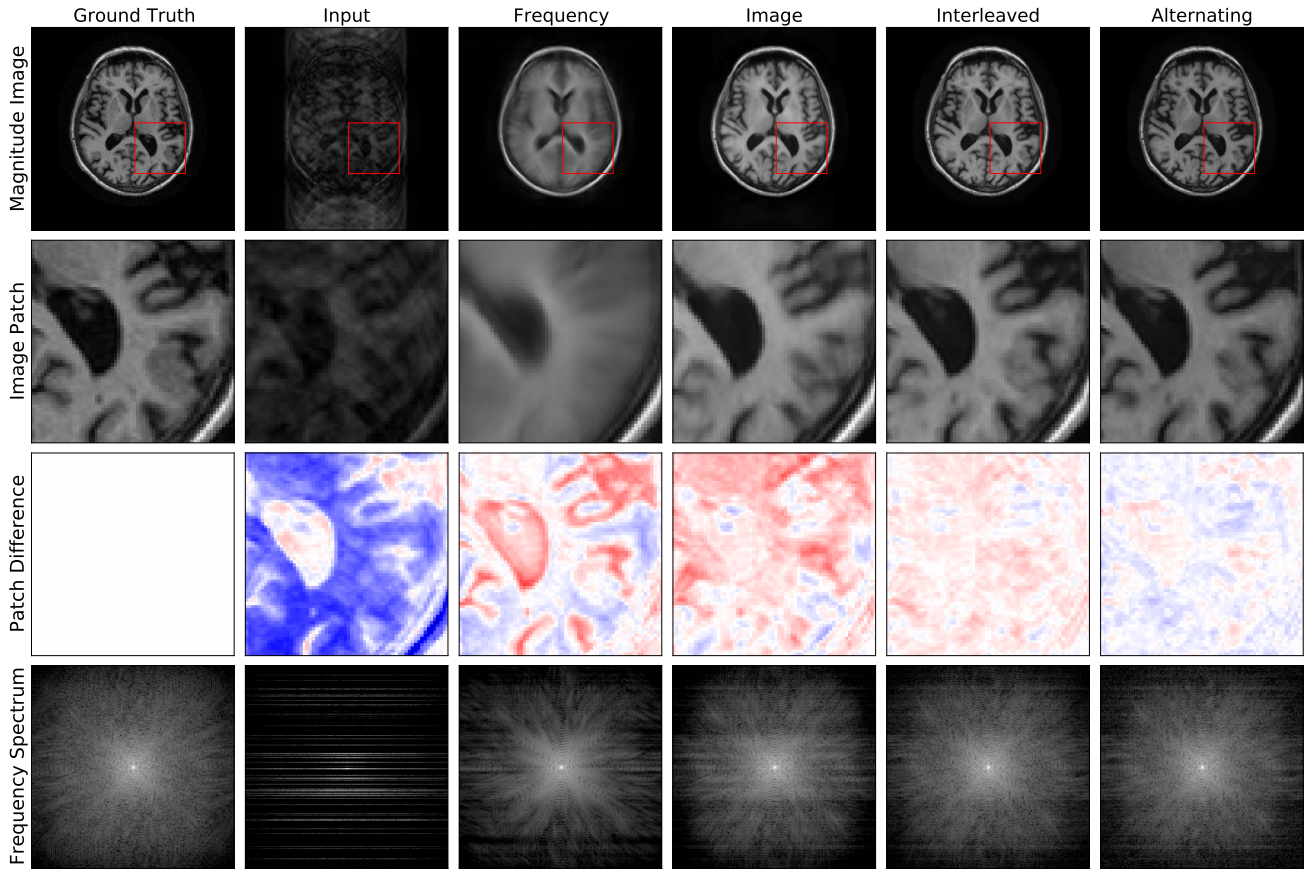

Figure 4. Example reconstructions from 4x undersampled data, zoomed-in image patches, difference patches between reconstructions and ground truth images, and frequency space reconstructions. The `Interleaved` and `Alternating` architectures provide more accurate and detailed reconstructions of the ground truth images, eliminating more 'ringing' and blurring artifacts.

extensive evaluation of networks trained with a variety of loss functions and assessed via different evaluation metrics. We train each of the four network architectures with seven loss functions on the 4x undersampled FastMRI knee data: image space L1 error, frequency space L1 error, a joint L1 metric summing image and frequency L1 errors, SSIM [43], multiscale SSIM [44], PSNR [18], and LPIPS [48]. We evaluate all networks on each metric except LPIPS, which we qualitatively found to poorly correspond to reconstruction quality. The joint L1 metric weights the frequency space L1 error by 0.1 relative to the image space L1 error. The SSIM and multiscale SSIM scores are computed with a window size of 7×7 and constants $k_1 = 0.01$, $k_2 = 0.03$, as is standard for the FastMRI challenge. For all SSIM metrics

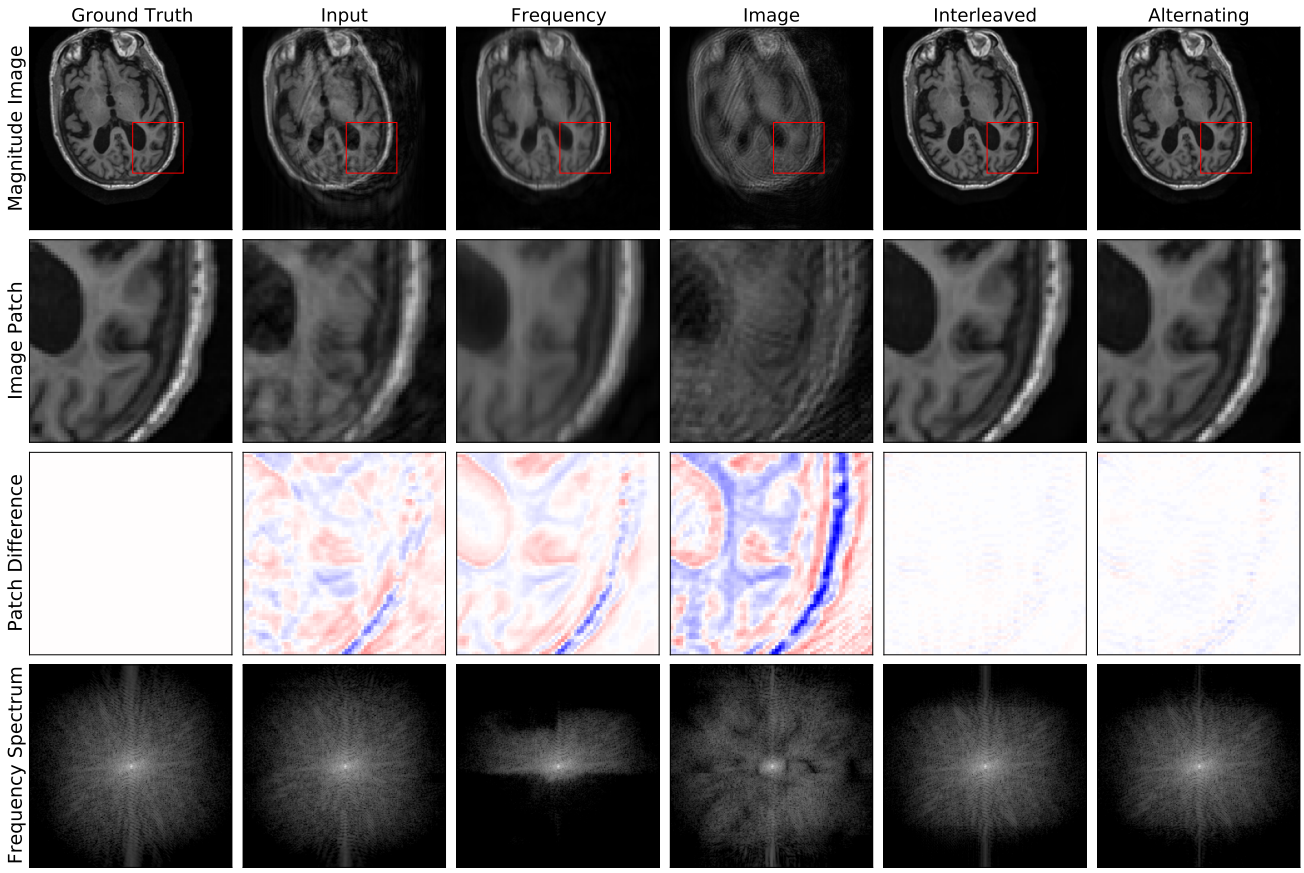

Figure 5. Example reconstructions with motion induced at 3% of scanning lines. The `Interleaved` and `Alternating` architectures more accurately eliminate the 'shadow' of the moved brain and the induced blurring compared to the single-domain networks.

and PSNR, we use the maximum value within a single *slice* as the maximum possible pixel value. This differs from the FastMRI evaluation stratgy, which uses the maximum value within a volume as the maximum possible pixel value. As a result, our reported SSIM, multiscale SSIM, and PSNR values underestimate the performance relative to the FastMRI evaluation metrics. For these reasons, our reported SSIM and PSNR values can be used for only rough comparison with other methods on the FastMRI leaderboard. Since the reported values are computed consistently across the methods, they are sufficient for evaluating performance differences between various fundamental neural network layer architectures in our experiments.

5.4. Results

Fig. 3 reports the performance statistics of all experiments described in Section 5.2. The `Interleaved` and `Alternating` architectures outperform the baseline architectures for nearly every subject. Across all tasks and nearly all subjects, the `Interleaved` and `Alternating` architectures are quite similar in numerical performance.

Sample image reconstructions are shown in Figs. 4-6. Qualitatively, for each task, the `Frequency` network provides a smoothed version of a high-quality image reconstruction. The `Image` network provides a reconstruction which effectively removes the 'background' effects of aliasing, motion corruption, and noise, but has limited success in effectively correcting these artifacts within the image. In contrast, the `Interleaved` and `Alternating` networks provide sharp, 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.

Our results on the knee MRI reconstruction task show that joint learning is beneficial in encouraging networks to reconstruct high-quality outputs along a manifold of realistic images preferred by several reasonable loss functions. Qualitatively, `Interleaved` and `Alternating` networks trained with SSIM-based losses (both SSIM and multi-scale SSIM) provide the highest-quality, sharpest reconstructions. Fig. 7 provides a qualitative comparison

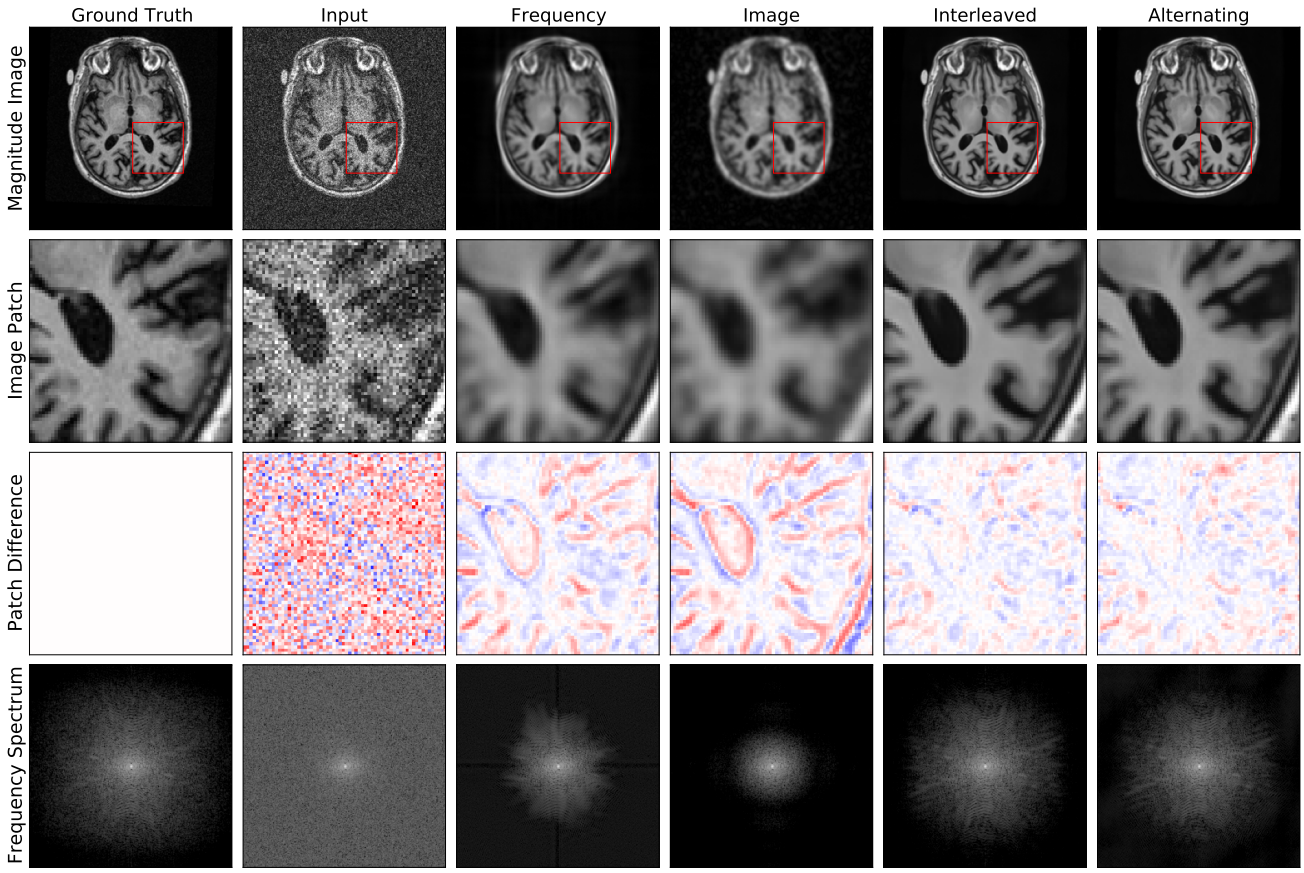

Figure 6. Example reconstructions with noise of standard deviation of 10^4. The `Interleaved` and `Alternating` architectures provide reconstructions which remove the pixelated noise effect without over-smoothing, in contrast to the single-domain networks.

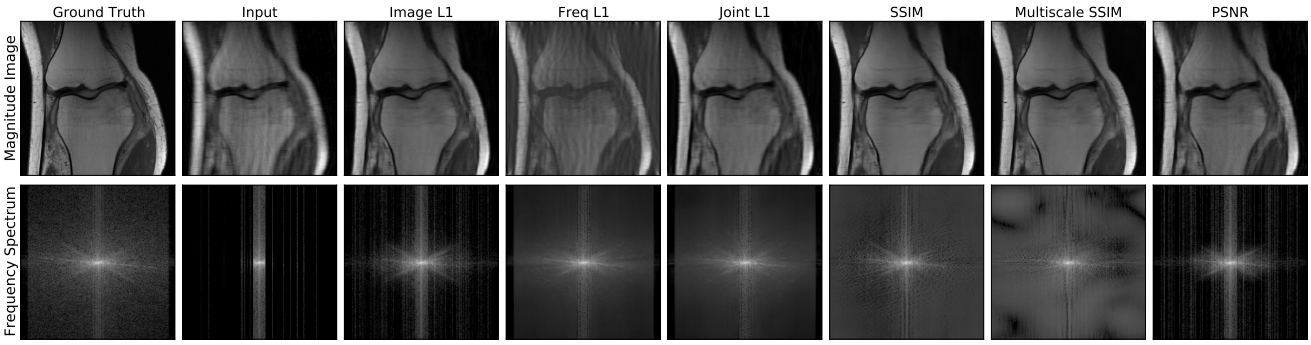

Figure 7. Example `Interleaved` architecture knee reconstructions with 4x undersampling, for networks trained with varying losses.

of sample reconstructions from `Interleaved` networks trained with different losses. A more extensive qualitative comparison is available in the Supplementary Material.

Quantitatively, for every loss function and evaluation metric, the `Interleaved` architecture performs better or equal to all other architectures (Fig. 8). `Alternating` networks outperform the baseline architectures across in all cases except networks trained with a PSNR or Frequency L1 loss, or when trained with a multiscale SSIM loss and evaluated in terms of Frequency L1 and Joint L1 metrics, which both incorporate a frequency space term. Results shown are for images without fat suppression; results for images with fat suppression are in the Supplementary Material and follow similar trends. Together, these results suggest that the success of joint learning is not specific to a certain loss landscape.

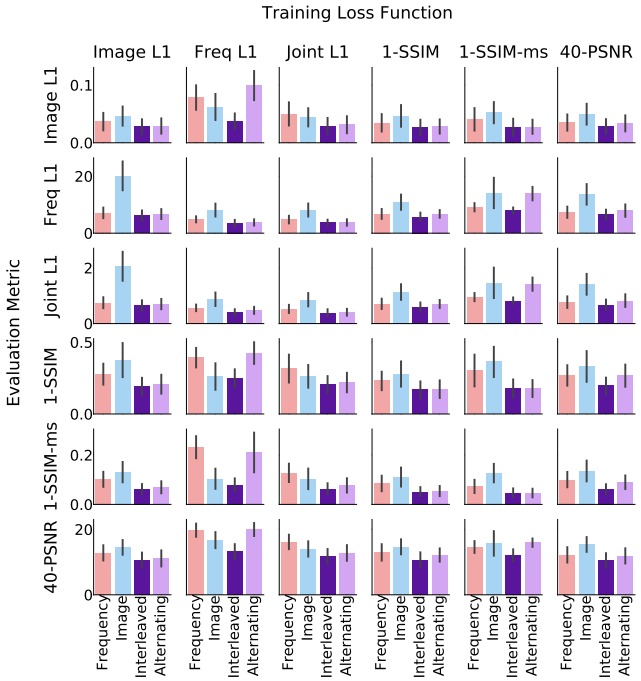

Figure 8. Image reconstruction evaluation metrics (rows) for networks trained with varying loss functions (columns). Metrics are shown such that lower values are better. Across nearly every training loss function and metric, the `Interleaved` network performs best. In almost every case, the `Alternating` 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.

6. Conclusions

We demonstrate the advantages of joint image and frequency space learning strategies for correcting corrupted Fourier imaging data. These approaches provide both correction of artifacts native to the frequency space and reconstruction of coherent structures in the image space. We demonstrate that these strategies perform better than pure frequency or image space baselines on a set of image reconstruction tasks under multiple diverse, challenging data corruption mechanisms. In addition, we demonstrate that joint learning strategies outperform pure frequency and image space baselines on real undersampled knee MRI frequency space signals, and when training via a wide variety of loss functions. These results point to joint layers as a useful starting point when designing neural network architectures for correcting frequency space artifacts.

In the future, we aim to build on the success and flexibility of joint layer architectures by incorporating task-specific techniques. One such technique is a data consistency constraint that has been effective for undersampled reconstruction [13, 39]. We hope to develop additional strategies for more varied applications such as motion correction, where direct consistency with acquired data is not necessarily desirable. 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. The combination of these advances promises to enable significantly improved reconstruction and analysis of Fourier imaging data in the face of widely-varying acquisition challenges and downstream applications.

References

[1] 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. 2, 4

[2] 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. 2, 4

[3] C Shyam Anand and Jyotinder S Sahambi. Wavelet domain non-linear filtering for mri denoising. *Magnetic Resonance Imaging*, 28(6):842–861, 2010. 3

[4] 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. 1

[5] 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. 5

[6] 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. 3

[7] Ariel Benou, Ronel Veksler, Alon Friedman, and T Riklin Raviv. Ensemble of expert deep neural networks for spatio-temporal denoising of contrast-enhanced mri sequences. *Medical Image Analysis*, 42:145–159, 2017. 3

[8] 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. 3

[9] 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. 2

[10] Ben A Duffy, Wenlu Zhang, Haoteng Tang, Lu Zhao, Meng Law, Arthur W Toga, and Hosung Kim. Retrospective correction of motion artifact affected structural mri images using deep learning of simulated motion. 2018. 3

[11] 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. 2

[12] 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. 2

[13] 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. 2, 4, 5, 9

[14] Yoseob Han, Leonard Sunwoo, and Jong Chul Ye. k-space deep learning for accelerated mri. *IEEE Transactions on Medical Imaging*, 2019. 2, 4

[15] 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. 3, 4

[16] 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. 3

[17] 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. 4

[18] Quan Huynh-Thu and Mohammed Ghanbari. Scope of validity of psnr in image/video quality assessment. *Electronics letters*, 44(13):800–801, 2008. 6

[19] 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. 2, 4

[20] 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. 3

[21] 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. 3

[22] 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. 4

[23] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Machine Learning*. 4

[24] 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. 3, 4

[25] Paul C Lauterbur. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *Nature*, 242(5394):190–191, 1973. 1

[26] 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. 2, 4

[27] R Leitgeb, M Wojtkowski, A Kowalczyk, CK Hitzenberger, M Sticker, and AF Fercher. Spectral measurement of absorption by spectroscopic frequency-domain optical coherence tomography. *Optics Letters*, 25(11):820–822, 2000. 1

[28] Michael Lustig, David L Donoho, Juan M Santos, and John M Pauly. Compressed sensing mri. *IEEE Signal Processing Magazine*, 25(2):72–82, 2008. 1

[29] 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. 2

[30] Albert Macovski. Noise in mri. *Magnetic Resonance in Medicine*, 36(3):494–497, 1996. 1

[31] 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. 3

[32] 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. 3, 4

[33] 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. 5

[34] Robert D Nowak. Wavelet-based rician noise removal for magnetic resonance imaging. *IEEE Transactions on Image Processing*, 8(10):1408–1419, 1999. 3

[35] 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. 3, 4

[36] 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. 2

[37] Patrick Putzky and Max Welling. Invert to learn to invert. In *Advances in Neural Information Processing Systems*, pages 446–456, 2019. 2, 4

[38] 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. 2, 4

[39] 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. 2, 4, 5, 9

[40] 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. 2, 4

[41] Lei Tian, Xiao Li, Kannan Ramchandran, and Laura Waller. Multiplexed coded illumination for fourier ptychography with an led array microscope. *Biomedical Optics Express*, 5(7):2376–2389, 2014. 1

[42] 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. 3

[43] 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. 6

[44] 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. 6

[45] CA Willey. Synthetic aperture radars–a paradigm for technology evolution. *IEEE Tranactions on Aerospace and Electronic Systems*, 21:440–443, 1985. 1

[46] 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. 2, 4

[47] 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. 5

[48] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 6

[49] 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. 2

[50] 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. 2

Supplementary Material

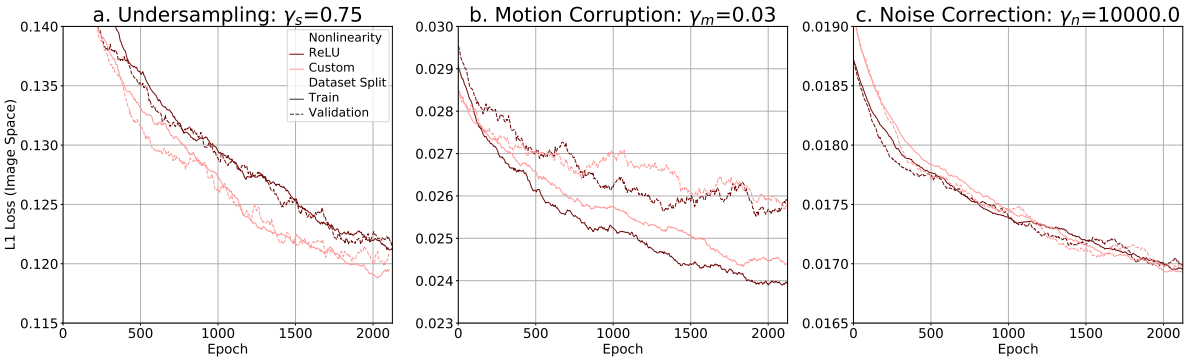

Figure 9. Comparison of frequency-space convolutional networks with ReLUs and with the custom nonlinearity described in Section 3.2. On every task, the network with the custom nonlinearity outperforms or performs on par to the network with ReLU nonlinearities on the validation set.

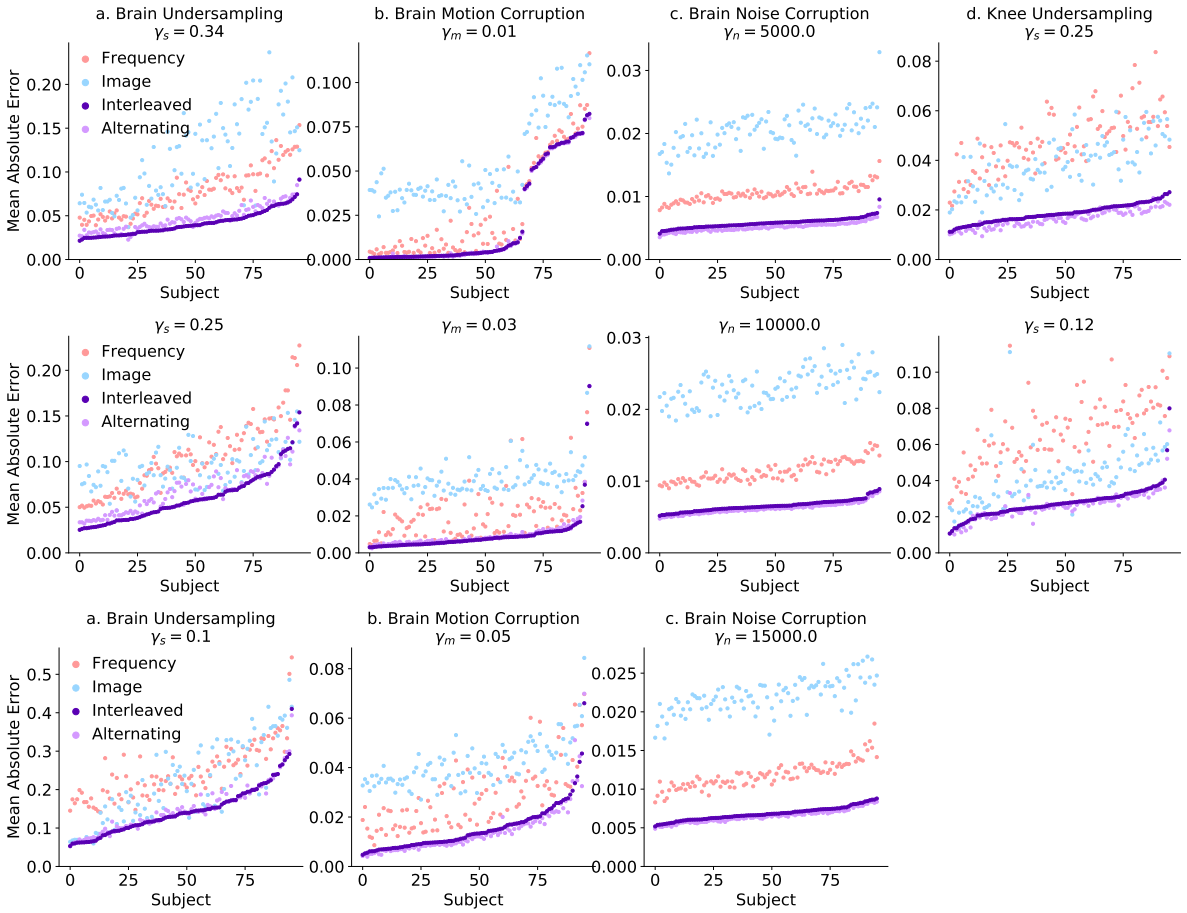

Figure 10. Subjectwise comparison of different architectures across multiple tasks with different architectures. Across all tasks, corruption levels, and nearly all subjects, the joint networks perform the best.

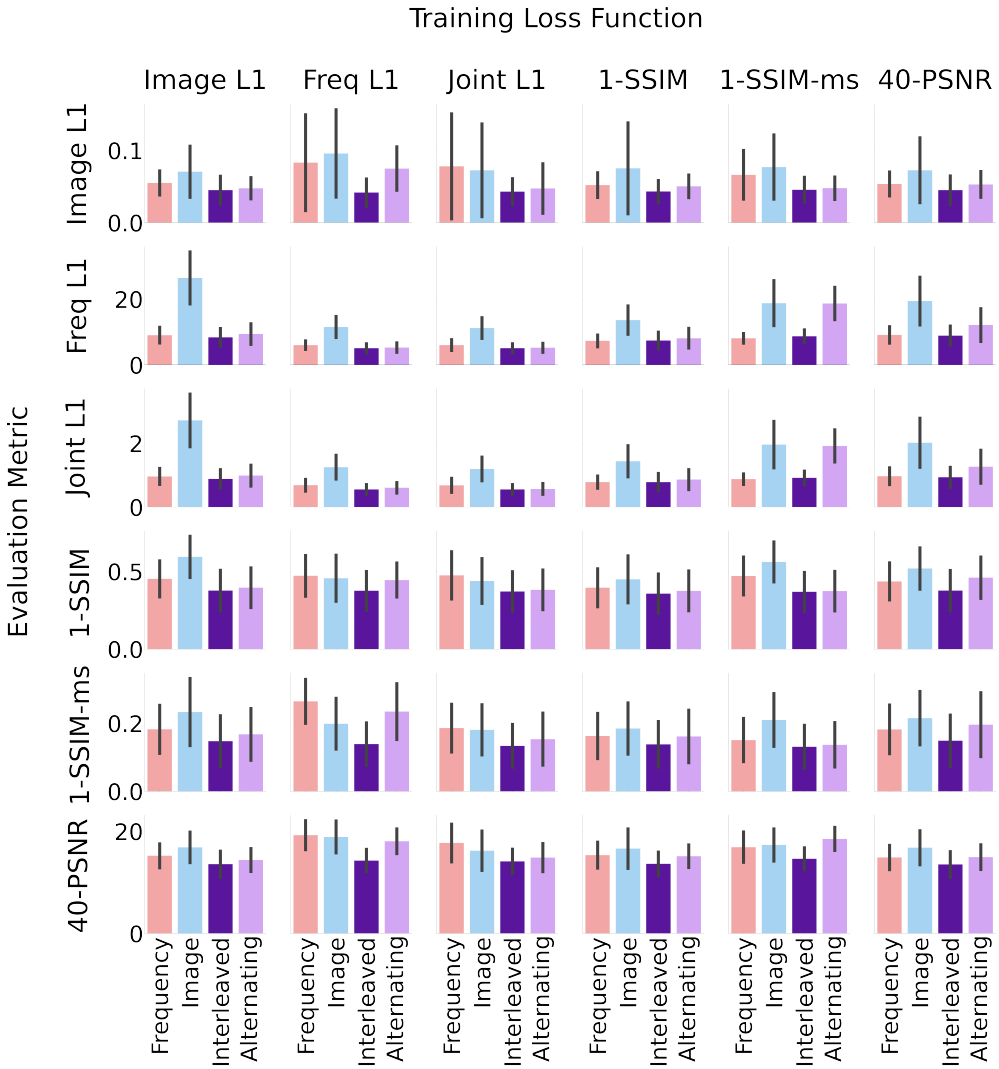

Figure 11. Image reconstruction evaluation metrics (rows) for networks trained with varying loss functions (columns) on images acquired with fat suppression. Metrics are shown such that lower values are better. Across nearly every training loss function and metric, the `Interleaved` network performs 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 12. Image reconstruction and detailed patches 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.

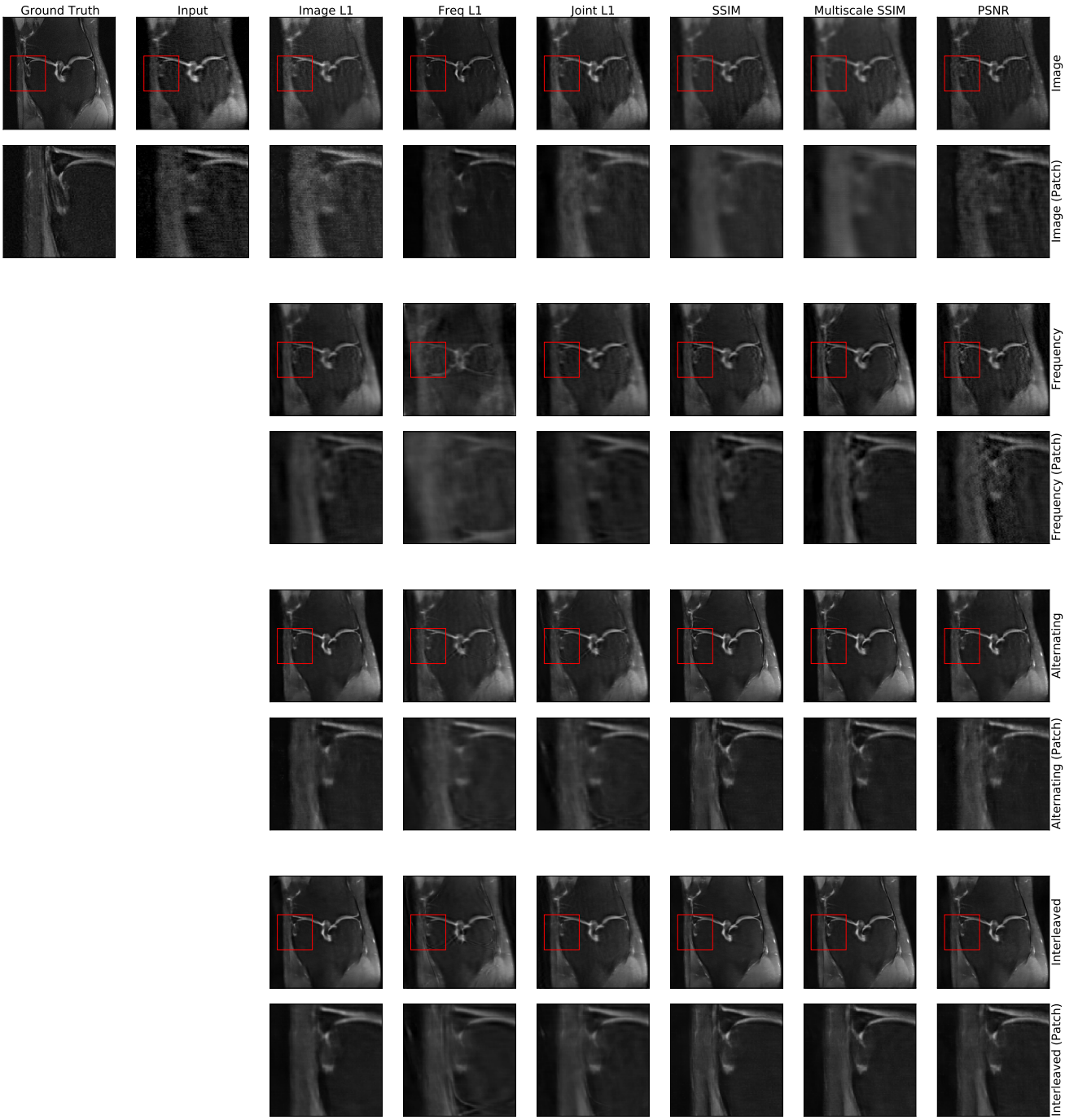

Figure 13. Image reconstruction and detailed patches for all architectures (rows) and loss functions (columns) on FastMRI images with fat suppression. The `Interleaved` and `Alternating` networks provide the sharpest reconstructions for all loss functions. Amongst these, both SSIM-based loss functions most sharply reconstruct high-frequency structures within the zoomed-in patch.