Subsampled MRI Reimplementation

Corey Zumar
EECS 5th Year MS
24247266
czumar@berkeley.edu

Alex Kot
BioE MTM
3033126202
akot@berkeley.edu

Abstract

We present a reimplementation of a deep learning method for accelerating magnetic resonance imaging (MRI) by reducing k-space data sampling. This method leverages a deep convolutional network based on the U-Net autoencoder for biological image segmentation. Our reimplementation shows acceptable reconstruction with only 29% of the k-space data. Extended analysis reveals the potential to reconstruct images using only 12.3% of the k-space data, as well as generalizability of the method to areas beyond the brain.

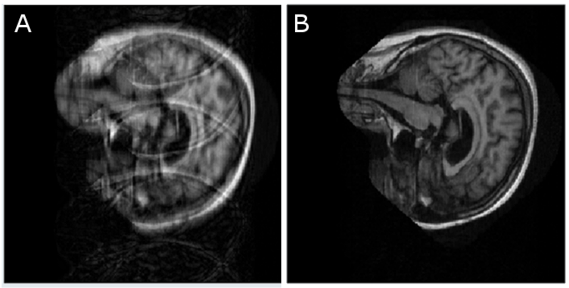

Figure 1. Example result of our technique on sagittal brain MRI. (A) subsampled image: 4-step, 4% low-frequency. (B) Reconstruction

1. Introduction

Nuclear magnetic resonance imaging (MRI) is a technique that uses strong magnetic fields and radio waves to generate images of the body based on signals from protons. It is used to non-invasively provide physicians with visual information of a patient's anatomy. MRI is an expensive technique, it can be difficult to apply to children due to its sensitivity to movement, and its long scan duration can prove troublesome for some patients with claustrophobia. If a system could take undersampled MRI data and produce medically acceptable images, then the MRI scan time could be reduced, decreasing the procedure's cost and allowing more access for claustrophobic patients.

MRI techniques collect raw data, known as k-space data, and produce images through complex data processing and inverse Fourier transforms. The raw k-space data is known to be low-quality with many missing entries, motivating research surrounding image reconstruction [7]. Current MRI techniques must use intrascan data to reconstruct medically acceptable images, with state of the art techniques beginning to leverage inter-scan data to improve MRI acquisition.

Producing medically acceptable images from highly undersampled (25-30% of the current standard) MR data was the main challenge faced by Hyun et al [6]. In their paper, they demonstrated that the use of subsampling in conjunction with deep learning methods can produce MR images comparable to those of standard reconstruction techniques, reducing data collection and processing overhead. Aside from the main research question, Hyun et al. described challenges in selecting subsampling strategies and working with the high-dimensional data. Here, we describe our reimplementation of Hyun et al.'s method, as described in their ArXiv submission (Figure 1). We also extend their analysis to provide a more thorough evaluation of their proposed technique. [6].

2. Related Work

While we are not aware of an existing open source implementation of the proposed method, several areas of related work are worth discussing.

2.1. Deep Methods for MRI Reconstruction

We have found an existing implementation of a similar deep learning model [14, 13, 12]. One of the main differences between this model and that of Hyun et al. is its dependence on compressive sensing MRI. As described in Hyun et al.'s paper, compressive sensing is not ideal for this application, for it does not preserve fine details that may be clinically important. Another similar deep learning approach is tailored for q-space MRI, which is used for diffusion MRI scans and does not cover the same clinical niche.

Additionally, this related work used a simple, feed-forward neural network with three fully-connected layers; Hyun et al.'s approach should offer a large improvement in reconstruction effectiveness due to its deeper, more nuanced architecture and the impressive performance statistics associated with its core CNN unit [4, 11].

2.2. Networks for Biomedical Image Processing

The task of reconstructing high resolution MRIs from undersampled inputs belongs to a broader set of biomedical image processing tasks that are increasingly leveraging deep learning to achieve better results. One of the more impactful technologies that precipitated the work of Hyun et al. is a deep convolutional autoencoder called U-Net [11]. In contrast with traditional autoencoding methods, U-Net's structure is inherently biphasic. In the first phase, inputs are projected into successively lower-dimensional spaces via convolutional layers; this is where autoencoders generally stop and emit an output. Building upon this structure, the additional second phase of the U-Net architecture performs successive upsampling and augmentation via skip connections, ultimately emitting an output of equivalent dimensionality to the input image. The work of Hyun et al. tailors the U-Net architecture to achieve better performance in the domain of MRI segmentation, as is described in (3.2).

2.3. Other Methods for MRI Reconstruction

High quality image reconstruction from messy, raw data can also be accomplished by non-deep-learning systems. Because the MRI technique relies heavily on reconstruction, there is an established body of work on reconstructing images with increasingly less raw data. Described strategies include compressed sensing, parallel imaging, and MRI fingerprinting, all of which require significant computational resources when deployed [5]. Another way to decrease total scan time is to combine data collection for different contrasts (scans of the same anatomy but measuring different things), sharing some information [10]. These non-deep-learning reconstruction/acceleration techniques do not take advantage of the same types of information leveraged by the CNN approach. Therefore, there is potential to combine these techniques with deep learning methods such as Hyun et al.'s, further reducing scan times.

3. Approach

3.1. Subsampling Method

The MR images were Fourier transformed into k-space data, then n-skip subsampled. After subsampling, the images were 'folded' with anterior-posterior uncertainty.

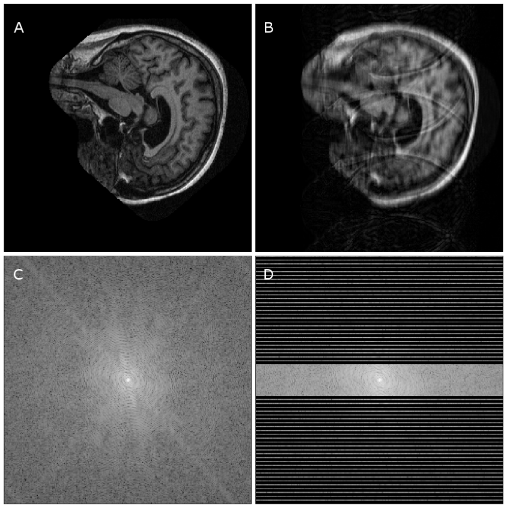

Figure 2. Subsampling visualization using 4-step, 4% low frequency subsampling. (A) Original image. (B) subsampled image. (C) k-space representation of original image. (D) k-space representation of subsampled image.

In order to remove this uncertainty, low frequency data was sampled from the center of k-space. Augmentation with this low frequency data, which comprised 4% of the total k-space data, helped to achieve separability (Figure 2).

3.2. Reconstruction Approach

Given a subsampled MR image, a deep neural network was used to construct an approximate representation of the corresponding original image. This network was a modified version of U-Net [11], which is a deep convolutional auto-encoder augmented with skip connections. U-Net was explicitly designed for biomedical image segmentation and is the state-of-the-art for this task. Notably, the distinction between U-Net and the architecture used for MRI reconstruction is that the latter uses upsampling layers (Referred to by Hyun et al. as "average unpooling layers") in place of U-Net's deconvolutional layers. The reconstruction network architecture diagram is shown in Figure 3. Network inputs and outputs were 256 x 256 images, where each pixel was an integer in the range [0, 255]. In practice, the reconstruction network, referred to as f_{net}, was implemented in Keras and trained for 2000 epochs on 1400 input images across 8 Tesla K80 GPUs.

3.3. k-space Correction

After obtaining a partially reconstructed output image from f_{net}, we applied the Fourier transform to the image to obtain its k-space representation, which we used to "fill in the blanks" of the original subsampled k-space data. This process is referred to as k-space correction and is denoted f_{corr}. k-space correction preserves the original data and

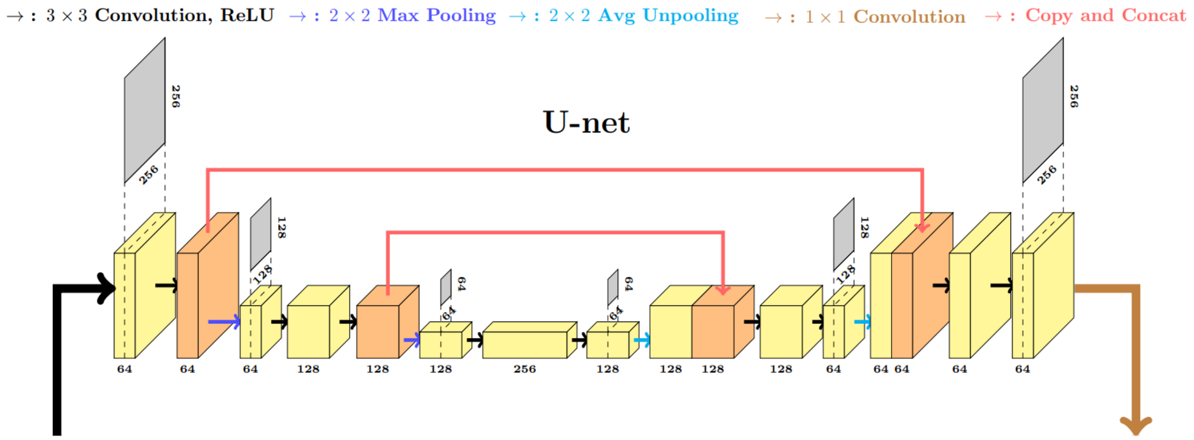

Figure 3. U-Net architecture [6]

improves the reconstructed image. In all, the reconstruction function is defined as $f : x \mapsto y$ is:

$$f = |\mathcal{F}^{-1}| \circ f_{corr} \circ \mathcal{F} \circ f_{net} \circ |\mathcal{F}^{-1}|$$

where $\mathcal{F}$ denotes the Fourier transform operation.

4. Implementation Details

We selected languages and libraries for the reimplementation of the proposed method as follows:

1. **Language**: All components of the reimplementation were written in Python.

2. **MR Image Processing**: In order to load and process samples, we used Nibabel [2] and NumPy's [1] support for relevant transformations, such as the Fourier transformation.

3. **Reconstruction Network**: While the original, closed-source implementation described by Hyun, et al. was indicated to have used Tensorflow, we selected Keras due to the ease of prototyping and debugging networks written using the framework.

Specifics regarding the network design and training process, as well as information about experimental setup, are detailed in the following subsections.

4.1. Dataset

We obtained a training set of 1,400 MR image slices from the Open Access Series of Imaging Studies (OASIS) project [9]. These were sagittal slices taken from men and women with or without clinically diagnosed dementia. After training, we used a test set of 400 MR image slices sampled from the same OASIS dataset.

4.2. Network Design

While we eventually produced an image reconstruction network implementation that we believe to be identical to the one outlined in the original paper, there were several difficulties associated with achieving this implementation that are worth discussing:

1. **Image input representation**: The proposed method for image reconstruction requires that inputs and intermediate image representations be transformed multiple times between k-space and human-interpretable image space. Additionally, there are several representations for grayscale image data that produce meaningful renderings of MRI samples. After experimentation with several image representations, we settled on the following formats that helped produce results consistent with those outlined in Hyun, et al:

   (a) **k-space data**: Data in k-space were represented as NumPy arrays of 128-bit complex values. Value ranges were not modified from the initial range obtained by projecting images into k-space via Fourier transform.

   (b) **Image space data**: Images were represented as NumPy arrays of integers in the range $[0, 255]$.

2. **Average Unpooling**: After completion of the autoencoding (dimensionality reduction) process performed by the first half of the network architecture, the latter half of the network produces an output that is dimensionally identical to the network input. This processing of expanding dimensions in the latter half of the network hinged on several ”average unpooling” layers described in the paper. However, the term ”average unpooling” is unorthodox and, to our knowledge, is not explicitly defined in the research sphere. Fortunately, with the aid of community input, we were able

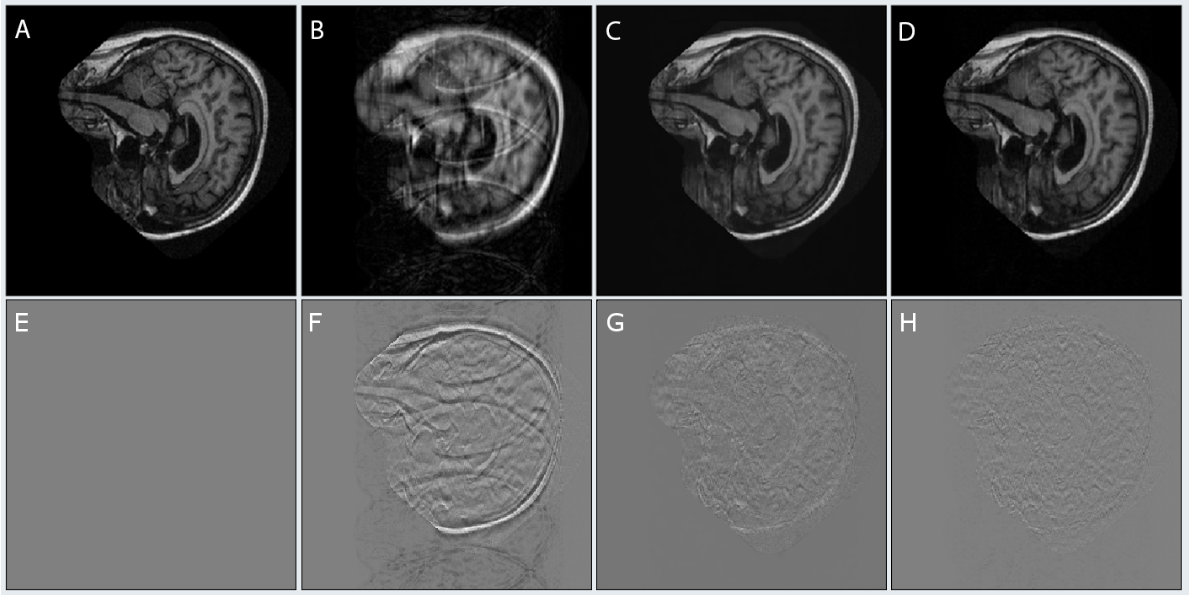

Figure 4. Reconstruction visualization using 4-step, 4% low-frequency subsampling. (A) Original image. (B) subsampled image. (C) U-Net result. (D) Corrected result. (E-H) Difference of image and image A.

to determine that ”average unpooling” is a synonym for ”upsampling.”

3. **Loss Calculation**: The loss function for training the reconstruction network described in the original paper was a variant of mean squared error (MSE). However, the representation specifics for the network outputs and labels were not clearly defined. As a result, we made our best attempt to infer the representation from the scale of the experimental losses reported in the original paper, which were on the order of thousandths. Ultimately, we arrived at the following loss function:

$$MSE = \frac{1}{n} * \frac{1}{256^2} \sum_{i=1}^{n} (\frac{R_i}{255} - \frac{L_i}{255})^2$$

where $R_i \in [0, 255]^{256\times256}$ represents a reconstructed and corrected image slice, while $L_i \in [0, 255]^{256\times256}$ represents the ground truth image slice. Pixel values were normalized to fall within the range $[0, 1]$, accounting for the interior division by a factor of 255. Additionally, error was represented as ”average error per pixel”, accounting for the exterior division by a factor of 256^2.

4.3. Network Training

The training set, T, for the reconstruction network consisted of 1400 slices. The training set can be mathematically represented as

$$T = \{\, x_i \mid x \in [0, 255]^{256\times256}, i \in [1, 1400] \,\}$$

Each MR image is a collection of 128 such slices. We sampled the interior-most 35 slices from consecutive images in the OASIS brain dataset until the cardinality of the training set reached 1400 slices, an identical size to the training set described in the original paper. Interior slices were selected because they generally contain the most interesting and developed anatomical structures. Accordingly, this set building strategy excluded those slices with small sections of anatomy at the lateral edges of the head.

Network training occurred on a single Amazon `p2.8xlarge` instance. This instance is equipped with eight Tesla K80 GPUs, and the training process was concurrently distributed across all of them. The training process occurred for 2000 epochs with a mini-batch size of 256, divided such that a sub-batch of 32 image slices was distributed to each GPU per iteration of stochastic gradient descent.

These attributes differ slightly from the training setup described in the paper. Namely, the training process described in the original paper leveraged a system with four NVIDIA GTX-1080 GPUs, using a distributed mini-batch size totaling 32 image slices. While the original paper indicated that the training process took 6 hours with its described setup, our training process took approximately 12 hours. The discrepancy in training time may be attributable to differences in CPU and GPU processing capabilities associated with the selected machines.

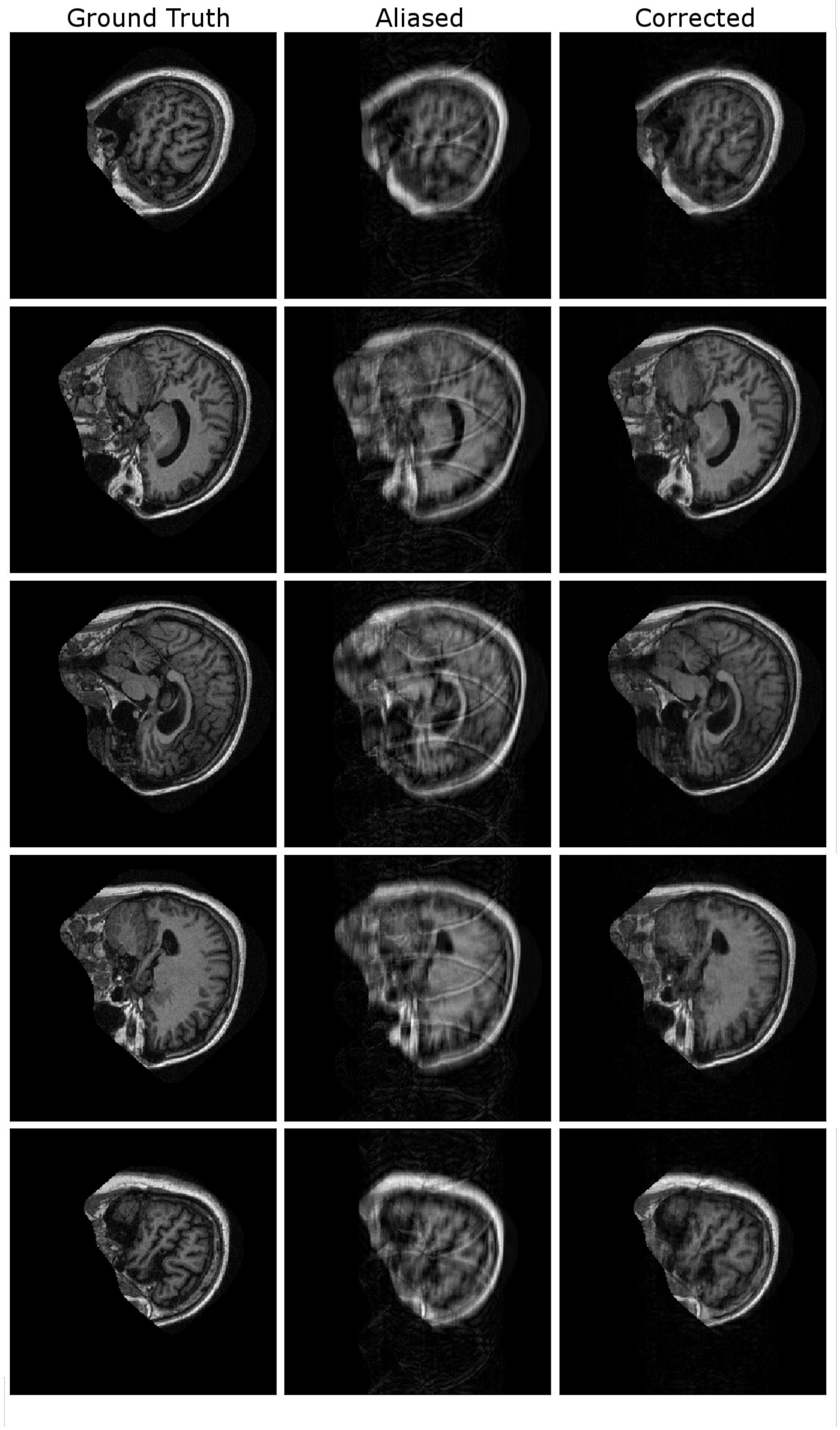

Figure 5. Reconstruction results of 5 MR images. The columns show ground truth, aliased, and corrected images respectively.

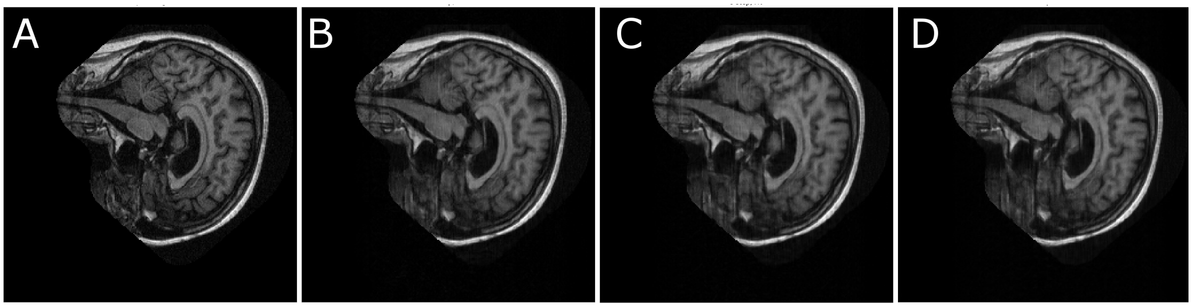

Figure 6. Reconstruction visualization using n-step, 4% low-frequency subsampling. (A) Original image. (B) 4-step. (C) 8-step. (D) 12-step.

5. Replication Experiments

The Reconstruction network was first evaluated on sagittal brain images from the OASIS dataset [9]. The subsampling parameters were set to match that of Hyun et al. (4-step; 4% low-frequency) [6]. Visually, our results were very similar to those of Hyun et al. As compared to the ground truth image, the subsampled image had significant noise and clear folds (Figure 4B, F). The U-Net reconstruction removed the folding artifacts and restored much of the image clarity (Figure 4C, G). k-space correction slightly improved the image (Figure 4D, H). The quantitative results followed this trend and were similar to those reported by Hyun et al. (Table 1). Reconstructed and aliased (subsampled) images were evaluated by quantitative comparison against original, fully-sampled label images. The metrics used to produce the quantitative results were MSE, as defined in (4.2), and the Structural Similarity Index (SSIM). SSIM attempts to more accurately measure perceptive loss than MSE, and a higher SSIM value indicates greater similarity between images. Ultimately, the network was able to reconstruct acceptable images across a range of anatomical structures present in the sagittal brain images (Figure 5).

	Aliased	Reconstructed
MSE	0.003929 ± 0.0015	0.001380 ± 0.0004
SSIM	0.7380 ± 0.026	0.8661 ± 0.026

Table 1. Quantitative results of 400 OASIS test images using 4-step, 4% low-frequency subsampling

6. Extension Experiments

6.1. Increasing Substep

We explored the network's performance with 8- and 12-step subsampling, adding 4% low-frequency data in each case (Table 2, Figure 6). There was approximately a 45% difference between the MSE associated with the 4-step and 8-step experiments and approximately a 5% difference between the 8-step and 12-step experiment. The qualitative results mirrored this result, as there were noticeable differences in clarity and detail between Figure 6B and Figure 6C, with smaller perceptible differences between Figures 6C and 6D.

Substep	Aliased MSE	Reconstructed MSE
4	0.003929 ± 0.0015	0.001380 ± 0.0004
8	0.004669 ± 0.0016	0.002001 ± 0.0006
12	0.004709 ± 0.0017	0.001911 ± 0.0005

Table 2. MSE results for 400 OASIS test images using n-step, 4% low-frequency subsampling

6.2. Training and Testing Subsample Mismatch

As part of our extended evaluation, we sought to determine whether or not the effectiveness of the reconstruction network was dependent on the degree to which its training inputs were subsampled. We qualitatively analyzed the results of inputting 4- and 8-step subsampled images to a network trained on 12-step subsampled images, as well as 8-step subsampled images to a network trained on 4-step subsampled images. None of the resulting images (Figure 7) approached the qualitative clarity and level of detail present in Figure 4D, implying that the network's learned reconstruction did not generalize to images of higher or lower quality than the training data.

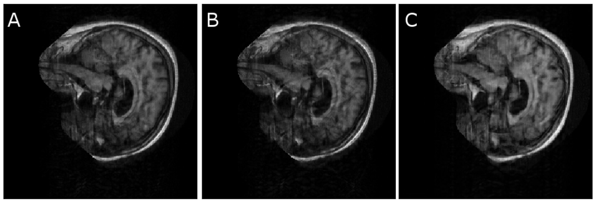

Figure 7. Example result of mismatch subsampling with significant artifacts. (A) 4-step input, 8-step network. (B) 4-step input, 12-step network. (C) 8-step input, 4-step network.

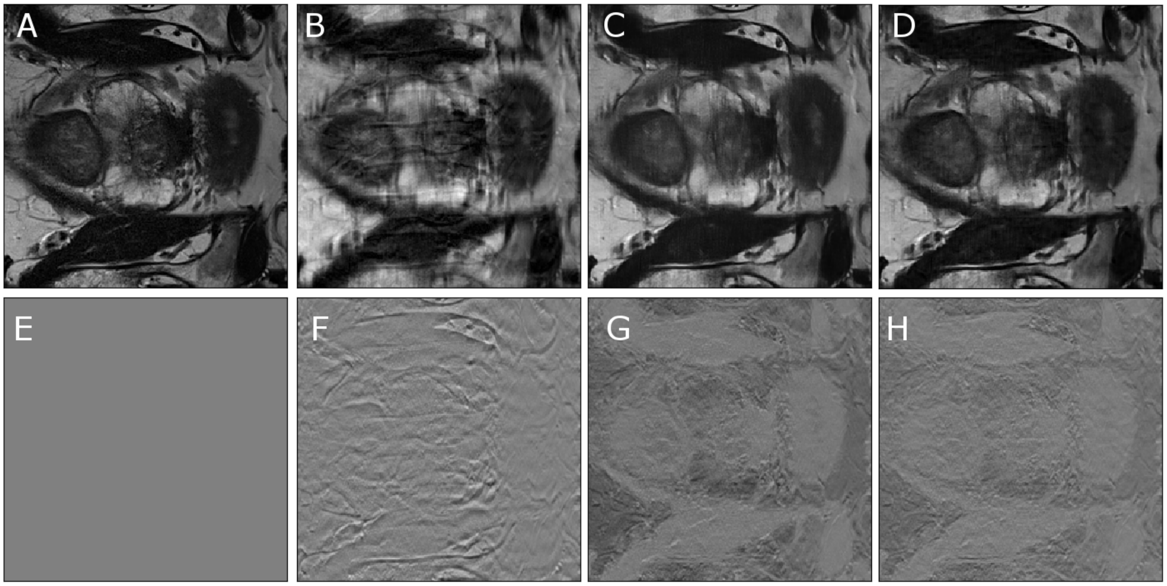

Figure 8. Prostate reconstruction visualization using 4-step, 4% low-frequency subsampling. (A) Original image. (B) subsampled image. (C) U-Net result. (D) Corrected result. (E-H) Difference of image and image A.

6.3. MAE Loss

The reconstruction network outlined by Hyun, et al. used mean squared error (MSE) for its loss function, as described in (4.3). While we found MSE to be a performant loss function for this use case, we wondered whether or not mean absolute error (MAE) would produce performance improvements. MAE is conventionally known to perform well with sparse inputs. Due to the fact that uncropped, grayscale s may be classified as sparse, we hypothesized that MAE had the potential to improve reconstruction performance.

Accordingly, the reconstruction network was retrained with inputs constructed using a sub-step of 4 and a low-frequency subsampling rate of 4%, as well as the following loss function:

$$\text{MAE} = \frac{1}{n} * \frac{1}{256^2} \sum_{i=1}^{n} \left| \frac{R_i}{255} - \frac{L_i}{255} \right|$$

where $R_i \in [0, 255]^{256 \times 256}$ represents a reconstructed and corrected image slice, while $L_i \in [0, 225]^{256 \times 256}$ represents the ground truth image slice. The test set for this experiment was identical to the test set used in (5).

For consistency, we used the MSE formulation described in (4.3), as well as SSIM, when evaluating test set performance. MAE training had a very small effect on the MSE (4% improvement) and SSIM, as shown in (Table 3).

	Aliased	MAE Reconstructed
MSE	0.003929 ± 0.0015	0.001322 ± 0.0004
SSIM	0.7380 ± 0.026	0.8678 ± 0.026

Table 3. Quantitative results of 400 OASIS test images using 4-step, 4% low-frequency subsampling. Reconstruction network trained with MAE loss

6.4. Prostate Images

In addition to exploring the proposed reconstruction technique using sagittal plane images of the brain, we also evaluated the method on a dataset of prostate images.

6.4.1 Dataset

We used 1400 images from 70 patients included in the ”Prostate X Challenge” dataset [8, 3] for this extension experiment. This dataset contains images from patients with and without disease affecting the prostate.

6.4.2 Experimental Setup

We constructed a training set of 1000 images from the prostate dataset. Each image slice was cropped to fit within a 256 × 256 pixel window. We constructed a test set of 400 images using the same methodology. As in the case of the replication experiments (5), MSE and SSIM were used as the evaluation metrics for comparison with ground truth images.

	Aliased	Reconstructed
MSE	0.009643 ± 0.0062	0.002150 ± 0.001
SSIM	0.7398 ± 0.038	0.8718 ± 0.019

Table 4. Quantitative results for 400 prostate test images using 4-step, 4% low-frequency subsampling

6.4.3 Results

As expected, MSE and SSIM evaluation metrics associated with the reconstructed image were substantially better than those associated with the aliased image (Table 4). This establishes the effectiveness of Hyun, et al.'s method on general MRI reconstruction tasks that extend beyond brain imaging.

7. Conclusion

In their paper, Hyun et al. provided a great starting point for subsampled MRI reconstruction. We were able to reproduce all of their qualitative and quantitative results, save for their optimistic training time. The modified U-Net architecture was effective for this task, and the described 4-step, 4% low frequency subsampling strategy worked well. However, there were a few limitations of Hyun et al.'s approach. The network did not reconstruct areas with very fine detail, such as the folds in the cerebellum in Figure 4. Also, it is not clear how the network performs with rare or inconspicuous pathologies.

Despite these limitations, this approach has great potential to impact clinical practice. A few potential improvements include harnessing the 3-dimensional structure of the data, optimizing the subsampling/data acquisition method, and experimenting with the use of LSTM network architectures.

8. Accessing the Open Source Code

All of the code and documentation associated with the reimplementation is publicly available on GitHub: `github.com/Corey-Zumar/MRI-Reconstruction`.

References

[1] NumPy. http://www.numpy.org/, 2017.

[2] M. Brett, M. Hanke, B. Cipollini, M.-A. Ct, C. Markiewicz, S. Gerhard, E. Larson, G. R. Lee, Y. Halchenko, E. Kastman, cindeem, F. C. Morency, moloney, J. Millman, A. Rokem, jaeilepp, A. Gramfort, J. J. van den Bosch, K. Subramaniam, N. Nichols, embaker, bpinsard, chaselgrove, N. N. Oosterhof, S. St-Jean, B. Amirbekian, I. Nimmo-Smith, S. Ghosh, G. Varoquaux, and E. Garyfallidis. nibabel: 2.1.0, Aug. 2016.

[3] K. Clark, B. Vendt, K. Smith, J. Freymann, J. Kirby, P. Koppel, S. Moore, S. Phillips, D. Maffitt, M. Pringle, L. Tarbox, and F. Prior. The cancer imaging archive (TCIA: Maintaining and operating a public informatino repository, 2013.

[4] V. Golkov, A. Dosovitskity, J. Sperl, M. Menzel, M. Czisch, P. Sämann, T. Brox, and D. Cremers. a-space deep learning: Twelve-fold shorter and model-free diffusion MRI scans. *IEEE Transactions on Medical Imaging*, 35(5):1344–51, 2016.

[5] K. Hollingsworth. Reducing acquisition time in clinical MRI by data undersampling and compressed sending reconstruction. *Phys Med Biol*, 60:R297–322, 2015.

[6] C. Hyun, H. Kim, M. Lee, S. Lee, and J. Seo. Deep learning for undersampled MRI reconstruction. *ArXiv preprint*, ArXiv:1709.02576, 2017.

[7] T. Jhamb, V. Rejathalal, and V. Govindan. A review on image reconstruction through MRI k-space data. *I J Image Graphics and Signal Processing*, 1:42–59, 2015.

[8] G. Litjens, O. Debats, J. Barentsz, N. Karssemeijer, and H. Huisman. Prostatex challenge data. *The Cancer Imaging Archive*, 2017.

[9] D. Marcus, T. Wang, J. Parker, J. Csernansky, J. Morris, and R. Buckner. Open access series of imaging studies (OASIS): Cross-sectional MRI data in young, middle aged nondemented, and demented older adults. *Journal of Cognitive Neuroscience*, 19:1498–1507, 2007.

[10] R. Mekle, E. Wu, and A. Laine. MRI scan time reduction through k-space data sharing in combo acquisitions with a spin echo sequence. *IEEE Transactions on Medical Imaging*, 22(7):806–23, 2003.

[11] O. Ronnenberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. *ArXiv preprint*, ArXiv:1505.04597, 2015.

[12] J. Schlemper. Deep-MRI reconstruction. *GitHub repository*, https://github.com/js3611/Deep-MRI-Reconstruction, 2017.

[13] J. Schlemper, J. Caballero, J. Hajnal, A. Price, and D. Rueckert. A deep cascade of convolutional neural networks for dynamic MR image reconstruction. *ArXiv preprint*, ArXiv:1704.02422, 2017.

[14] J. Schlemper, J. Caballero, J. Hajnal, A. Price, and D. Rueckert. A deep cascade of convolutional neural networks for MR image reconstruction. *ArXiv preprint*, ArXiv:1703.00555, 2017.