
ML-Driven Fast-MR: CNN-Guided Sub-Sampling and Reconstruction of Clinical Brain MR Data

Daniela F. de Albuquerque*

Department of Electrical and Computer Engineering
Duke University
Durham, NC 27708
dfd4@duke.edu

Abstract

Magnetic Resonance (MR) is one of the most commonly used imaging modalities in clinical medicine, with its main advantages being high soft-tissue contrast and no radiation exposure. However, it suffers from long-acquisition times leading to high susceptibility to motion artifacts, increased overall costs and reduced patient comfort. Trying to mitigate such problems by shortening acquisition time has long been a goal for researchers in the field. In this work, we propose utilizing deep learning to find an optimal k-space sub-sampling routine that will allow us to achieve high-fidelity reconstructions from under-sampled inputs. More specifically, we jointly trained a k-space sub-sampling layer and a deep convolutional neural network (CNN) to both search for an optimal sub-sampling scheme and reconstruct high-fidelity brain MR images. To force our sub-sampling layer to yield a binary output and choose only a limited number of k-space columns, we used an annealing factor. Our proposed method outperforms a naive equispaced k-space sub-sampling configuration. Additionally, it approaches performance observed with an informed k-space sub-sampling configuration. This gap in performance when compared to an informed sub-sampling control is likely due to inefficient annealing and overall random initialization of our sub-sampling layer. These results demonstrate that jointly optimizing k-space sub-sampling trajectory and reconstruction in a CNN framework might be a viable option to reducing acquisition time and, with that, making clinical MR cheaper and more readily accessible to a wider range of patients.

1 Introduction

Since the invention of the first MR scanner circa 1977, proton MR has become one of the most utilized imaging modalities in clinical medicine due to its superior soft-tissue contrast in comparison to its counterparts CT, X-Ray and ultrasound (US). This explains its many applications, for example, in cardiology to assess myocardial perfusion and the risk of acute coronary syndrome and in neurology to discern dead tissue from recoverable penumbra regions in stroke patients and study brain activity with high spatial and temporal resolution. Additionally, since it does not utilize ionizing radiation, MR is also amongst the few safe imaging modalities available to children, pregnant women and other at-risk patient populations.

Despite all of these benefits, a considerable drawback of MR imaging lies on its often very long acquisition times, which contribute to increased procedure costs, reduced patient comfort and compliance and increased overall susceptibility to motion artifacts. Acquisition time is dependent on the number of phase encoding k-space trajectories sampled. However, drastically under-sampling

*MD-PhD Candidate, Medical Scientist Training Program, Duke University School of Medicine

k-space (i.e., below Nyquist frequency) leads to well-known aliasing artifacts, which are not easily eliminated with simple reconstruction techniques.

Overcoming this problem and obtaining high fidelity reconstructions from drastically under-sampled k-space data has long been a topic of research in the field of MR imaging. One of the most popular approaches to tackle this problem is compressed-sensing MR (CS-MR). In a nutshell, CS-MR aims to acquire only the most important coefficients in a sparse transform domain so as reconstruct an image of interest with high fidelity and acceptable SNR. Importantly, this approach is limited in that it requires a priori knowledge about transform sparsity, as well as pseudo-random under-sampling and non-linear reconstruction [1].

2 Related Work

Deep learning models have proven very successful in several computer vision tasks such as image recognition [2] and classification [3]. This success has led to increased interest in applying such models to automate and improve performance of several tasks involving large-scale biomedical imaging data. In the field of MR imaging, for example, CNNs have been used to perform reconstruction and cortical parcellation of structural human brain scans in large scale and with reduced compute time [4].

In fact, CNNs have already been successfully used to reconstruct high-fidelity MR data from sub-Nyquist sampled inputs. For example, in [5] this was done using a fixed uniform sub-sampling scheme with an added small number of center frequencies. Although this approach is interesting, it is limited in that sub-sampling routine is not optimizable and relies on a priori knowledge about the importance of center frequencies in achieving high-fidelity reconstructions. Authors in [6] used instead a compressed-sensing like approach implemented using a variational model which was trained offline but could be applied online to unseen data. Yet another group [7] utilized an adversarial training scheme between a k-space reconstruction network and an evaluator to choose, at inference time, which phase-encoding trajectory should be acquired next by the scanner.

Interestingly, physical 'hardware' parameters of imaging systems (e.g., illumination scheme, lens aperture) can be optimized using deep learning models by incorporating such parameters in the network's computational graph. For example in [8], the authors trained a network capable of selecting an ideal illumination scheme to identify malaria-infected cells. That is, knowledge gained by a CNN can be used to optimize and inform imaging system design and operation itself. This has enormous potential applications to almost all commonly used medical imaging modalities, including proton MR.

Here, we propose utilizing a Unet-type architecture along with an optimizable k-space sub-sampling layer to perform high-fidelity reconstruction of multicoil brain MR data from drastically under-sampled sets. This approach exploits proven strengths of deep learning models in MR reconstruction tasks while allowing for the sub-sampling scheme to be jointly optimized, potentially increasing prior observed performance on this task. Finally, trajectory used to sample phase-encoding lines can be viewed as a real 'physical' operation performed by the scanner at acquisition time. Therefore, knowledge from such a model can be used to design optimal k-space sub-sampling trajectories, which would lead to a drastic reduction in MR acquisition time.

3 Methods

3.1 Data

Data consisted of 2,286 fully-sampled T1 multicoil brain scans acquired in 1.5T and 3T clinical grade MR scanners. These were collected at multiple sites and are a part of the larger publicly available FastMRI database, created as a collaboration between NYU Radiology and Facebook AI Research [9]. The primary goal of this collaboration is to test whether machine learning can aid in the reconstruction of medical images and, more specifically, if it can be used to reconstruct clinical MR images with high fidelity from under-sampled data.

More specifically, this database contains both singlecoil fully-sampled knee MR images and multicoil brain MR data, containing both the above-mentioned T1 scans as well as scans acquired with T2 and FLAIR sequences. Results reported herein were achieved using sets containing only 20% randomly chosen scans out of the 2,286 available per simulation.

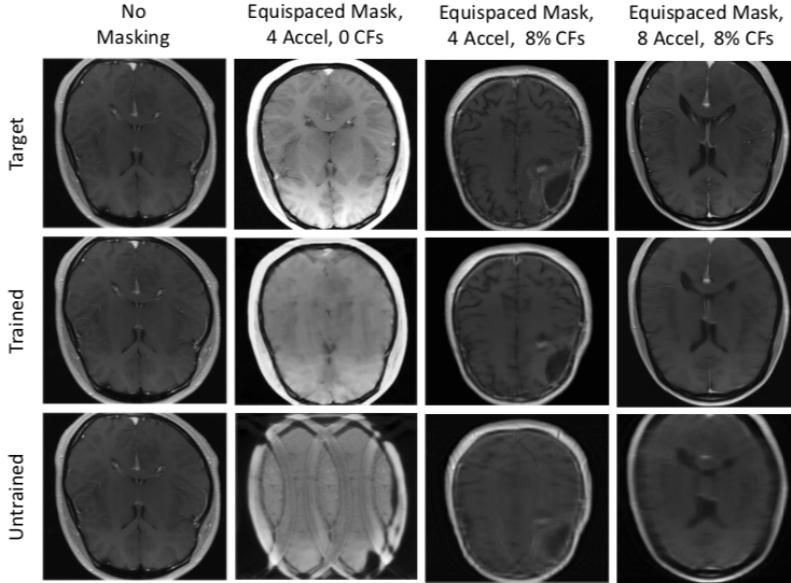


Figure 1: Target vs. Trained and Untrained reconstructions for control simulations. None of these utilized the physical_layer. Column 2 represents naive sub-sampling and Columns 3 & 4 represent informed sub-sampling controls.

3.2 Pre-processing and Network

Fully-sampled multicoil k-space data was first center-cropped to achieve a 260x260 shape per slice. It was then instance normalized and inverse Fourier transformed to image space, where information from different coils was combined using a root sum of squares (RSS) approach. Combined coil data was then Fourier transformed back into k-space and fed into an optimizable sub-sampling layer.

The optimizable layer assigns a trainable weight to each column of the input k-space data. It then utilizes simulated annealing [10] to force convergence of final mask weights to either 0 or 1, which is a natural constraint given the all-or-none choice of sampling or not a given phase-encoding trajectory. More specifically, annealing is achieved by first multiplying the column weights by a temperature factor (which is linearly increased with each epoch) and then passing the result to a Softmax function, yielding a mask that can then be applied to our fully-sampled input k-space data. Corresponding masked image is then obtained by applying an inverse Fourier transform to the masked k-space data and taking its absolute value. The resulting masked slice is fed to the CNN described later in this section.

For the control simulations, masking was instead applied as a pre-processing step using a fixed equispaced mask with (informed sub-sampling) or without (naive sub-sampling) the addition of 8% center frequencies. For these controls, all pre-processing steps highlighted above were the same, but final masked k-space data was instead converted to its corresponding masked image and directly fed to the CNN described below (i.e., the physical layer is bypassed in the control simulations).

Briefly, CNN implementation was based on the popular Unet architecture first proposed by Olaf Ronneberger, Phillip Fischer and Thomas Brox [11]. It has 4 down and up-sampling blocks. Each block consists of 2 convolutional layers (kernel-size 3, stride 1, 32 filters), each followed by an instance normalization layer, a leaky ReLU and a dropout layer. Skip connections were added between matching up and down sampling blocks. An average pooling layer with kernel-size 2 and stride of 2 was also added after each down-sampling block.

Loss utilized is a simple L1 norm between predicted reconstruction and target image with an added regularization term. This regularization consists of a constant penalty multiplied by the mean square error between the sum of the weights in the optimizable mask (after Softmax step) and the desired number of sampled columns. Finally, desired number of sampled columns was computed by dividing

total number of k-space columns (in the fully-sampled data) by the acceleration factor (e.g., 4) and rounding the result to the nearest integer.

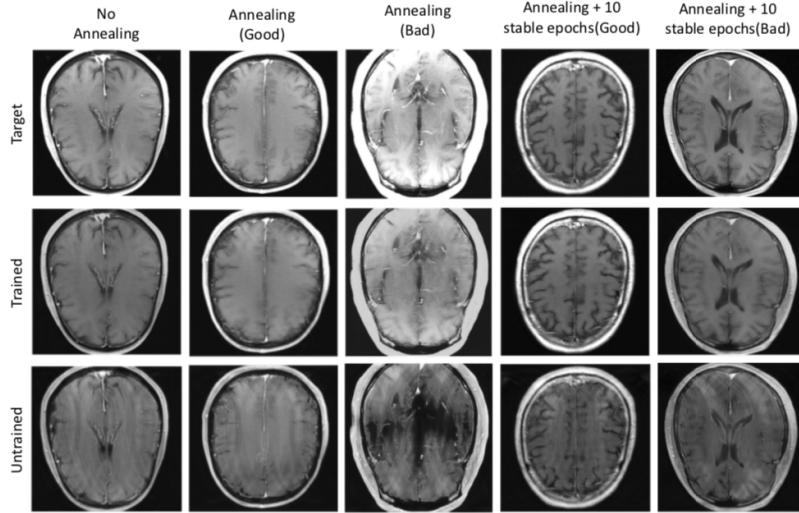


Figure 2: Target vs. Trained and Untrained reconstructions for experimental simulations. All of these utilized physical_layer+Unet and acceleration 4. 'Good' and 'Bad' labels refer to good and poor reconstruction examples. Last 2 columns refer to simulation with 50 epochs of annealing + 10 epochs of stable temperature at the end of training. Maximum temperature reached was the same in both simulations with annealing.

3.3 Training

The model described above was implemented using PyTorch and trained on an NVIDIA GeForce GTX 1650 GPU. We ran a total of 4 control simulations: no-masking with Unet only, equispaced masking with acceleration 4 and without any added center frequencies (naive sub-sampling) and equispaced masking with accelerations 4 and 8 and 8% added center frequencies (informed sub-sampling). All of these models were trained for 50 epochs, with learning rate 0.001, learning-rate step-size 40, batch size 16 and drop-out probability 0.5.

The experimental simulations were as follows: 1) model with physical layer but without annealing, 2) model with physical layer and annealing (to a maximum temperature of 6) and acceleration 4, 3) model with physical layer and annealing (maximum temperature of 6), acceleration 4 and 10 extra epochs of stable temperature at the end of training. First and second experimental simulations were trained for 50 epochs while the third was trained for 60, to account for the added 10 stable temperature epochs. The same learning rate, learning-rate step-size, batch-size and drop-out probabilities highlighted above were used for all of these. The regularization penalty was set to 0.5 and physical_layer weights were initialized using values drawn from an uniform distribution with zero mean and standard deviation of 1. Temperature parameter was initialized as 1 and (for the simulations with annealing) linearly increased to the maximum temperature given above. All hyper-parameters were tuned using a separate validation set. Unfortunately, I did not have enough computational resources to repeat simulations and gain additional insight into performance variation and stability.

3.4 Metrics

We used three primary metrics to evaluate model performance, namely normalized mean square error (NMSE), peak-signal to noise ratio (PSNR) and structural similarity (SSIM), all computed between predicted reconstruction and target images.

We utilized the standard definitions of PSNR and SSIM, as follows:

$$PSNR(x, y) = \frac{10 \log_{10}(\max(\max(x), \max(y))^2)}{|x - y|^2} \quad (1)$$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1) + (2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (2)$$

where μ_x and μ_y are the averages and σ_x^2 and σ_y^2 are the variances of x and y, and σ_{xy} is the covariance between x and y. C_1 and C_2 represent factors used to stabilize division by a weak denominator.

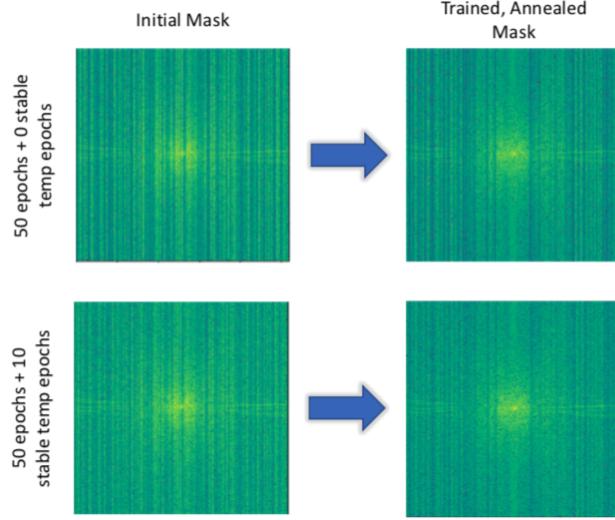


Figure 3: Initial and Trained Masks for Annealing Simulations. Both simulations were performed using the same seed so as to make performance comparison more straightforward.

4 Results

Non-surprisingly, control without masking and using Unet only performed very well (SSIM = 0.99, PSNR = 61.7, NMSE = $2.84e^{-5}$) whereas naive sub-sampling control performed poorly (SSIM = 0.69, PSNR = 22.1, NMSE = 0.03). Adding 8% center frequencies increased performance for controls using equispaced masking and accelerations equalling 4 and 8 (Table 1). From a qualitative standpoint, with naive sub-sampling scheme and Unet only, original masked images had severe aliasing and final trained versions looked much improved, albeit still with notable aliasing artifacts (Figure 1, column 2). Using an informed sub-sampling scheme allowed Unet model to reconstruct images without the original aliasing artifacts, but still with worse off fine details (e.g., less defined tumor border) - Figure 1, columns 3 & 4.

Model with physical_layer + Unet, 4 acceleration, and annealing (but no added stable temperature epochs) outperformed the naive sub-sampling control by a big margin. In fact, its performance (SSIM = 0.84, PSNR = 29.6, NSME = 0.005) came somewhat close to the one observed for informed sub-sampling control with matching acceleration factor (SSIM = 0.91, PSNR = 35.1, NMSE = 0.007). Finally, adding the 10 extra epochs with stable temperature at the end of training did improve overall performance (SSIM = 0.86, PSNR = 30.3, NMSE = 0.005), without compromising desired binary nature of k-space masking (Figure 3). As expected, model with physical_layer + Unet but no annealing constraint had superior performance in comparison to the simulations using annealing (SSIM = 0.90, PSNR = 31.9, NMSE = 0.003). Qualitatively, sharper looking reconstructions were achieved with physical_layer + Unet but no annealing (Figure 2, column 1). Simulations with

annealing yielded overall good reconstructions (Figure 2, columns 2 & 4) with minimal to no residual aliasing. However, in both cases, some reconstructions were of poorer quality (Figure 2, columns 3 & 5).

Table 1: Validation Metrics for Control and Experimental Simulations at the End of Training

Simulation	Accel	Anneal	+Epochs	SSIM	PSNR	NMSE
No-Mask	0	No	0	0.99	61.7	$2.84e^{-5}$
Naive-subsampling	4	No	0	0.69	22.1	0.030
Informed-subsampling	4	No	0	0.91	35.1	0.010
Informed-sub-sampling	8	No	0	0.87	31.7	0.015
Physical_layer+Unet	4	No	0	0.90	31.9	0.003
Physical_layer+Unet	4	Yes	0	0.84	29.6	0.005
Physical_layer+Unet	4	Yes	10	0.86	30.3	0.005

5 Discussion

In this work we designed and implemented a deep learning model capable of both finding an optimal k-space sub-sampling mask and reconstructing high-quality images from significantly under-sampled inputs. Specifically, we masked fully-sampled k-space data using a weighted mask and then utilized annealing and an added regularization term to force network convergence to a binary mask, with only a limited number of selected k-space columns.

As highlighted above, model with added physical layer and annealing easily outperformed our naive sub-sampling control. Although not very surprising, this result is still interesting because, similarly to naive control, our model was not given any a priori knowledge about the importance of center frequencies. In fact, as shown in Figure 3, after only 50 epochs of training, our model was able to learn the importance of center-frequencies in achieving high-fidelity reconstructions and correctly assigned higher weights to center columns. This is in agreement with our knowledge of center-frequencies carrying more information.

Simulations with annealing came close to, but ultimately did not match, performance observed for informed sub-sampling control with equivalent acceleration factor. We suspect this gap in performance was due to both inherent instability of annealing method utilized and random initialization of mask weights used. If temperature is increased too fast or too high, mask binarization occurs before model has had enough time to correct for possible poor initialization choices (i.e., assigning initial low weights to center columns) and overall result converges only to a sub-optimal local minima. On the other hand, if temperature is not increased fast enough or high enough, model ends up failing to yield a binary mask.

In this context, simulation with 50 epochs of annealing followed by 10 epochs with stable temperature is informative. It shows that allowing model to have some more epochs of training to compensate for poorly initialized weights improves performance without compromising final binary nature of mask. Ultimately, it might be possible to match performance seen with informed controls by running simulations for longer (i.e., reducing temperature increase rate) and adding a small number of stable temperature epochs either at the very end of training or after a certain number of epochs with annealing. If possible, this would be a very interesting result since model would have arrived at an optimal sub-sampling scheme despite disadvantage of starting off without a priori knowledge given to controls. Another possibility to address this issue is instead embedding knowledge about the importance of center-frequencies into our model by biasing mask so as to have higher weights assigned to center columns at initialization. Although current work does not allow us to confirm this, we speculate this model would easily match and potentially even outperform our informed controls.

Acknowledgments

I would like to acknowledge Dr. Horstmeyer and Amey Chaware for their useful suggestions and guidance in this project.

References

- [1] S. Geethanath et al., "Compressed Sensing MRI: A Review," *Crit Rev Biomed Eng*, vol. 41, no. 3, pp. 183–204, 2013, doi: 10.1615/CritRevBiomedEng.2014008058.
- [2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *arXiv:1512.03385* [cs], Dec. 2015, Accessed: Apr. 28, 2020. [Online]. Available: <http://arxiv.org/abs/1512.03385>.
- [3] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017, doi: 10.1145/3065386.
- [4] L. Henschel, S. Conjeti, S. Estrada, K. Diers, B. Fischl, and M. Reuter, "FastSurfer – A fast and accurate deep learning based neuroimaging pipeline," *arXiv:1910.03866* [cs, eess, q-bio], Mar. 2020, Accessed: Mar. 31, 2020. [Online]. Available: <http://arxiv.org/abs/1910.03866>.
- [5] C. M. Hyun, H. P. Kim, S. M. Lee, S. Lee, and J. K. Seo, "Deep learning for undersampled MRI reconstruction," *Phys. Med. Biol.*, vol. 63, no. 13, p. 135007, Jun. 2018, doi: 10.1088/1361-6560/aac71a.
- [6] K. Hammernik et al., "Learning a variational network for reconstruction of accelerated MRI data: Learning a Variational Network for Reconstruction of Accelerated MRI Data," *Magn. Reson. Med.*, vol. 79, no. 6, pp. 3055–3071, Jun. 2018, doi: 10.1002/mrm.26977.
- [7] Z. Zhang, A. Romero, M. J. Muckley, P. Vincent, L. Yang, and M. Drozdzal, "Reducing Uncertainty in Undersampled MRI Reconstruction with Active Acquisition," *arXiv:1902.03051* [cs], Feb. 2019, Accessed: Apr. 04, 2020. [Online]. Available: <http://arxiv.org/abs/1902.03051>.
- [8] R. Horstmeyer, R. Y. Chen, B. Kappes, and B. Judkewitz, "Convolutional neural networks that teach microscopes how to image," *arXiv:1709.07223* [physics], Sep. 2017, Accessed: Apr. 28, 2020. [Online]. Available: <http://arxiv.org/abs/1709.07223>.
- [9] J. Zbontar et al., "fastMRI: An Open Dataset and Benchmarks for Accelerated MRI," *arXiv:1811.08839* [physics, stat], Dec. 2019, Accessed: Apr. 04, 2020. [Online]. Available: <http://arxiv.org/abs/1811.08839>.
- [10] A. Chakrabarti, "Learning Sensor Multiplexing Design through Back-propagation," *NIPS 2016*, p.9.
- [11] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *arXiv:1505.04597* [cs], May 2015, Accessed: Apr. 04, 2020. [Online]. Available: <http://arxiv.org/abs/1505.04597>.