CS 7643: Fast MRI

Enrico Zerilli
Georgia Institute of Technology
ezerilli@gatech.edu

Robert Bartel
Georgia Institute of Technology
rbartel3@gatech.edu

Sudipto Lodh
Georgia Institute of Technology
slodh3@gatech.edu

Abstract

Magnetic Resonance Imaging is a fundamental diagnostic tool for a large spectrum of disorders, including oncological, musculoskeletal and neurological diseases. Unfortunately, its major weakness is its long acquisition time, which easily exceeds 30 minutes. This poses several limitations to its applicability, ranging from its high costs to common motion artifacts. Thus, accelerating its acquisition speed is critical to fully take advantage of its considerable benefits, mainly the absence of harmful X-rays and its excellent soft tissue contrast. Current techniques use parallel imaging through multiple coils and compressed sensing to acquire fewer measurements. However, more accurate reconstruction can be achieved via the use of Deep Learning on large datasets, that includes k-space data, like the NYU fastMRI dataset [8]. In this project, we show how self-supervision can be beneficially used with comparable results to the supervised case, without the need to acquired prohibitively expensive fully-sampled data.

1. Introduction/Background/Motivation

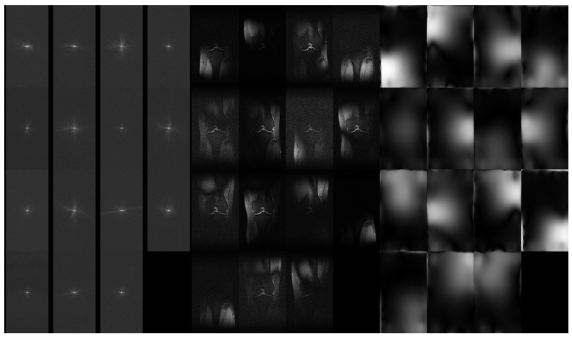
(5 points) What did you try to do? What problem did you try to solve? Articulate your objectives using absolutely no jargon. (5 points) How is it done today, and what are the limits of current practice? (5 points) Who cares? If you are successful, what difference will it make? (5 points) What data did you use? Provide details about your data, specifically choose the most important aspects of your data mentioned here. You don't have to choose all of them, just the most relevant.

Medical diagnosis of a wide range of disorders makes, nowadays, extensive use of imaging techniques such as X-Rays, Ultrasounds, Computed Tomography (CT) or MRI. In particular, *Magnetic Resonance Imaging* has become a powerful and flexible technique that provides accurate details and excellent soft tissue contrast, especially for neurological, musculoskeletal and oncological diseases [8]. MRI consists in emitting a sequence of magnetic fields, varying in time and space, and measuring the resonant electromagnetic response through one or multiple receiver coils for all different frequencies. This allows to fully-sample the interested region in Fourier space (also known as *k-space*) up to a maximum frequency. The spatially-resolved image m can then be reconstructed simply by inverse Discrete Fourier Transform (DFT) of the *k-space* image y:

$$\hat{m} = \mathcal{F}^{-1}(y) \tag{1}$$

Nonetheless, the use of MRI as a diagnostic tool is greatly limited by a number of relevant factors, first of all its long acquisition time, which easily exceeds 30 minutes in most of the cases. As a direct consequence, MRI has low patient throughput and high costs [8]. Also, patients are not comfortable in staying perfectly still for such a long time and often cause artifacts from motion, which highly impact the quality of the diagnosis. This makes MRI less suited to several diseases, for which CT is preferable due to its shorter acquisition time and its minor costs. Therefore, increasing acquisition speed has been a major goal of research related to MRI in the last decades. Indeed, a shorter acquisition time would allow to decrease its costs, to enlarge the spectrum of disorders for which MRI can be used, to avoid harmful X-rays to patients through the use of CT, to diminish artifacts from patient motion and improve thus the quality of the acquired images. Unfortunately, two simple accelerations, like sampling up to a lower maximum frequency or undersampling, either decrease the spatial resolution of the reconstructed image or introduce *aliasing* artifacts by violation of the Nyquist-Shannon sampling theorem[1], and are thus not applicable.

[1] the sampling frequency f_s should be at least twice as the maximum frequency in the spectrum f_{max} to avoid aliasing: $f_s > 2f_{max}$.



(a) k-space data from 15 coils (b) Individual coil spatial images from fully sampled data (c) Coil sensitivity map magnitudes given by ESPIRiT

Figure 1. Multi-coil MRI reconstruction [8].

Since the advent of MRI in the 1970s, the introduction of *parallel imaging* has improved acquisition times by using multiple coils to acquire several views of the imaged volume in *k-space*, each modulated by a complex-valued position-dependant coil sensitivity to the MR signal [8]. The *k-space* image y_i measured by coil i of n_c is:

$$y_i = \mathcal{F}(S_i m) + noise = g_i * \mathcal{F}(m) + noise \qquad (2)$$

where S_i is the coil sensitivity and g_i its Fourier transform, which can be estimated either by separate low-resolution scans or directly from the *k-space* measurements by a fully-sampled small central (low spatial frequencies) region [8]. This is shown in Figure 1. In the case of a fully-sampled *k-space*, the above represents a set of linear equations that is overdetermined by a factor n_c and can thus be pseudo-inverted to reconstruct m, as long as the system is full-rank. On the other hand, this overdetermination can be exploited to accelerate acquisition by undersampling the *k-space*. As long as there are more measurements than image voxels to be reconstructed, a least-squares solution can be used to get an n_c-fold acceleration. In practice, different factors, such as a tight resolution, spread out the coil sensitivities, thus lowering the rank of the linear system, which leads to the need of regularization and to smaller typical acceleration factors.

Compressed sensing (CS) provides a framework to reconstruct an image whose measured (undersampled) *k-space* y matches closely its Fourier transform, subject to sparsity constraints. As such, the reconstructed image best approximates the ground-truth image that would be reconstructed from fully-sampled data. In the single-coil case, this optimization problem can be expressed in its Lagrangian dual form as:

$$\min_m \frac{1}{2}||\mathcal{P}(\mathcal{F}(m)) - y||_2^2 + \lambda R(m) \qquad (3)$$

	Volumes		Slices	
Coil	Single	Multi	Single	Multi
training	973	973	34,742	34,742
validation	199	199	7,135	7,135
test	118	108	,4032	3,903
challenge	104	92	3,810	3,305

Table 1. Volumes and slices per task and set.

where, $\mathcal{P}$ is a projection mask that performs undersampling and R is a convex regularizer typically chosen to be either the L1-norm of the image, or the L1-norm of a wavelet representation of the image (which is typically sparse), or a total-variation (TV) penalty penalizing spatial gradients [8]. In the multi-coil case, which provides increased *Signal Noise Ratio* (SNR) over extended fields of view and can take advantage of *parallel imaging* speedup, the CS formulation given by 3 becomes:

$$\min_m \frac{1}{2}\sum_{i=1}^{n_c} ||\mathcal{P}(\mathcal{F}(S_i m)) - y_i||_2^2 + \lambda R(m) \qquad (4)$$

which can be fused with *parallel imaging* in the ESPIRiT approach [6]. Unfortunately, CS has the major drawback of introducing compression artifacts and oversmoothing (lost of details fundamental to radiologists for diagnosis), while being computational expensive. Moreover, CS frequently makes use of hand-crafted features, that are not those representing the actual data.

New emerging approaches attempt to improve these baselines through the use of Deep Learning models. Specifically, this has been enormously facilitated by the publication of the NYU fastMRI dataset [3], whose primary goal is to test whether machine learning can aid in the reconstruction of medical images [2, 8]. This dataset contains a large amount of anonymized MRI scans of knees and brains, providing: raw multi-coil *k-space* data; emulated single-coil (ESC) *k-space* data, that is linear combinations of multi-coil raw data fitted to the ground-truth root-sum-of-squares (RSS) in a least-squares sense; ground-truth images reconstructed from fully-sampled multi-coil acquisitions through a simple RSS of the inverse DFT and, for single-coil, also as inverse DFT of the ESC data; additional DICOM images representing a larger variety of machines and protocols, but for which raw data is not available. The dataset thus allows both single-coil and multi-coil reconstruction tasks from undersampled data. Data is provided separately for each task and already splitted in training and validation sets (containing fully-sampled scans), as well as test and challenge sets (containing undersampled data). Knee multi-coil raw data contains 6970 scans, acquired

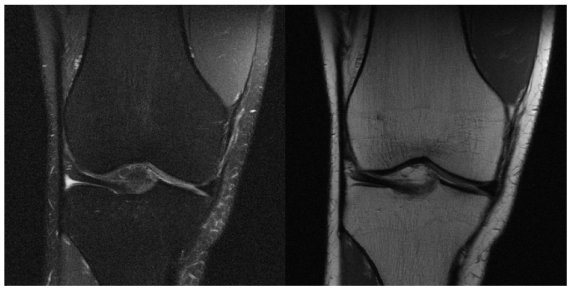


Figure 2. A PD knee image with fat suppression (a) and without (b). Being a soft tissue, fat as a high response in MRI, but makes details difficult to see [8].

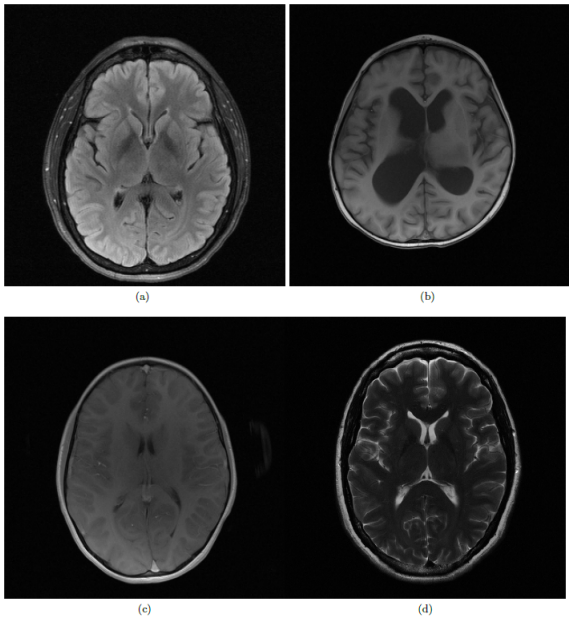


Figure 3. Axial brain slices with FLAIR (a), T1 weighted (b), T1 POST (c) and T2 weighted (d). Each protocol allows to highlight different structures. [8]

using 15 coils, matrix size 320x320, slice thickness 3mm and two protocols: a coronal proton-density weighting (PD) with (798 scans) and without (796 scans) fat suppression. Brain multi-coil raw data contains 1594 scans and includes axial T1 weighted (some of which with contrast agent, T1 POST), T2 weighted and FLAIR protocols. In the brain scans, only axial scans up to the orbital rim has been used to ensure data de-identification. Table 1 summarizes the number of volumes and slices per task in the provided sets (only in the training and validation sets the same slices are shared between the single-coil and multi-coil frameworks), while Figure 2 and 3 show slices of knee and brain for the different protocols used.

In the context of this project, due to dataset size constraints w.r.t. the available hard-drives, **we are going to limit our study to the case of single-coil knee data, for which sizes are tractable**: training ∼88 GB, validation ∼19 GB, test ∼7 GB, challenge ∼1.5 GB. Unfortunately, the multi-coil data both for knee and brain is just too big for us to use (just the training sets are about 1 TB each). Also, DICOM data will not bee used, since it does not provide with raw *k-space* data and it is intended to be eventually used as additional training data.

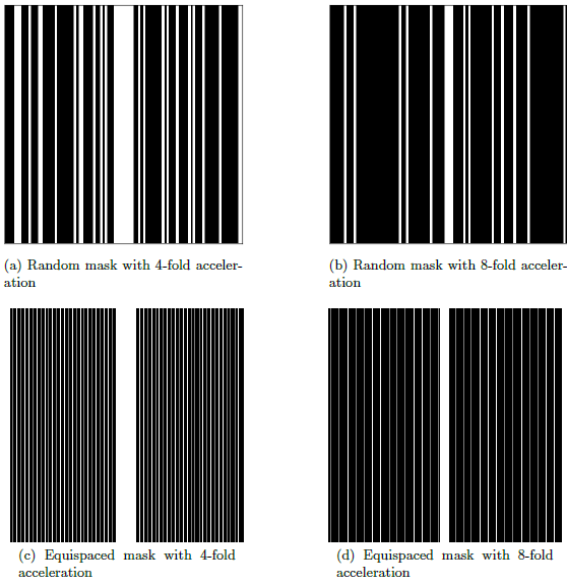
2. Approach

(10 points) What did you do exactly? How did you solve the problem? Why did you think it would be successful? Is anything new in your approach? (5 points) What problems did you anticipate? What problems did you encounter? Did the very first thing you tried work? Important: Mention any code repositories (with citations) or other sources that you used, and specifically what changes you made to them for your project.

NYU fastMRI provides a large dataset and a code base [3, 4] that can be leveraged to experiment with Deep Neural Networks (DNNs) in the context of MRI acceleration through subsampling [8]. These models are often trained using supervised methods that are at the moment the current *de facto* choice, when it comes to applying these models in the context of MRI. The simplest of these image reconstruction models are U-Nets [5] and ResNets [1]. More advanced hybrid techniques involving *parallel imaging* data models with DNN-base priors are current state-of-the-art and some of them are even being adopted in the clinic, but we will not explore those.

Rather, our approach consists first in **fine-tuning the U-Net model** provided in the fastMRI code base [4], that will become our **baseline**. However, supervised learning methods require a prohibitively expensive process of annotating the collected data with ground-truth outputs. This is especially true in the context of MRI, where collecting a large amount of fully-sampled *k-space* data is quite costly and *k-space* acquisitions are naturally subsampled via *parallel imaging*. Therefore, we will concentrate our attention on self-supervised methods and attempt to implement a **self-supervision hold-out method** [7], which have yet to be assessed on a large dataset like fastMRI. Finally, we are going to examine a second self-supervision method, involving ... to complete ...

In all our experiments, the undersampling is performed by masking *k-space* lines from a fully-sampled acquisition to all slices in a volume along the phase encoding direction, in order to simulate an acceleration via subsampling that is



(a) Random mask with 4-fold acceleration

(b) Random mask with 8-fold acceleration

(c) Equispaced mask with 4-fold acceleration

(d) Equispaced mask with 8-fold acceleration

Figure 4. Examples of undersampling masks. [8]

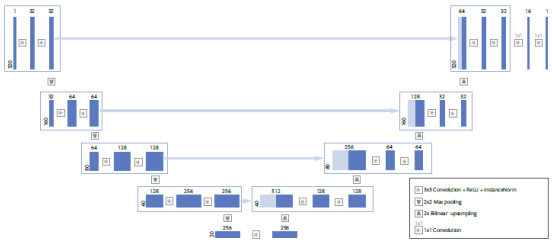


Figure 5. Single-coil baseline U-Net model. [8]

actually realizable in real-world scenarios. For obvious reasons related to keeping most of the low-frequency information contained in a slice, undersampling masks are produced by first cropping a small center region of lowest frequencies in the *k-space*, representing either 8% (4-fold acceleration) or 4% (8-fold acceleration) of all *k-space* lines. The remaining higher frequency *k-space* lines are included uniformly at random for the knee and equidistant for the brain, until the desired acceleration factor is achieved [8]. Examples of canonical undersampling masks are shown in Figure 4.

2.1. Fine-tuning the baseline U-Net model

Sudipto talks about the U-Net model here, paragraph 6.3 of the fastMRI paper [8] and U-Net paper [5].

2.2. Self-Supervised hold-out ResNet

Robert talks about self-supervised on held-out data. To cite U-Net paper [7]

2.3. Self-Supervised ResNet other

Finally, we will
Enrico talks about a second self-supervised method

3. Experiments and Results

(10 points) How did you measure success? What experiments were used? What were the results, both quantitative and qualitative? Did you succeed? Did you fail? Why? Justify your reasons with arguments supported by evidence and data. Important: This section should be rigorous and thorough. Present detailed information about decision you made, why you made them, and any evidence/experimentation to back them up. This is especially true if you leveraged existing architectures, pre-trained models, and code (i.e. do not just show results of fine-tuning a pre-trained model without any analysis, claims/evidence, and conclusions, as that tends to not make a strong project).

In order to measure success, we are going to use in our experiments the same evaluation metrics that are currently being used in the literature, given a reconstructed slice $\hat{v}$ and the ground-truth v, with means μ_v and $\mu_{\hat{v}}$, variances σ_v^2 and $\sigma_{\hat{v}}^2$ and covariance $\sigma_{\hat{v}v}$:

- *Normalized Mean Square Error* (NMSE):

$$NMSE(\hat{v}, v) = \frac{||\hat{v} - v||_2^2}{||v||_2^2} \tag{5}$$

- *Peak Signal-to-Noise Ratio* (PSNR)

$$PSNR(\hat{v}, v) = 10 \log_{10} \frac{max(v)^2}{MSE(\hat{v}, v)} \tag{6}$$

- *Structural Similarity* (SSIM)

$$SSIM(\hat{v}, v) = \frac{(2\mu_{\hat{v}}\mu_v + c_1)(2\sigma_{\hat{v}v} + c_2)}{(\mu_{\hat{v}}^2 + \mu_v^2 + c_1)(\sigma_{\hat{v}}^2 + \sigma_v^2 + c_2)} \tag{7}$$

Although, NMSE is the main evaluation metrics in the current literature, it tends to favor much more smoothness rather than sharpness, which in MRI may be problematic due to the loss of fundamental details for diagnosis. Thus, we also report PSNR, which gives an idea of the ratio of powers between the maximum image intensity and the noise, and SSIM, which measures similarity between the images' structures [8]. Nonetheless, these metrics do not necessarily ensure the level of details in the MRI reconstruction required for proper diagnosis, which would need sharp and detailed images. Therefore, we will be training using L1-loss (faster to compute due to not having any division operator) to encourage pixel-to-pixel identity between reconstructed images and ground-truth:

$$L_1(\hat{v}, v) = ||\hat{v} - v||_1 \tag{8}$$

We will also take advantage of the professional evaluation of a 2nd year (post-graduation) resident in Radiology in one of the best hospitals in Italy, similar to what is done in [7].

3.1. Fine-tuning the baseline U-Net model

We ran the base experiment by trainingthe model on singlecoil for knee dataset , download fromhttps://fastmri.med.nyu.edu/.The dataset contains the kspaces images from the MRI coils which grouped in volume for image for an individual scan.The training is performed with an acceleration of 4 , with mask type of random applied on the k-space input image (to simulate the under sampled input date) . The start of art model used is UNet with 1 input channel/output channel ,32 numbers of top level channels to the UNet , 4 poolinglayers , RMS prop learning rate of 0.001 and 40 step size to decrease the learning rate by 0.1. The complete training on 32 core CPU and 64 GB RAM took almost 3 days to complete.The model is trained in 50 epochs using the singlecoil_train, singlecoil_test and singlecoil_val dataset.

We get the following graphs for training.

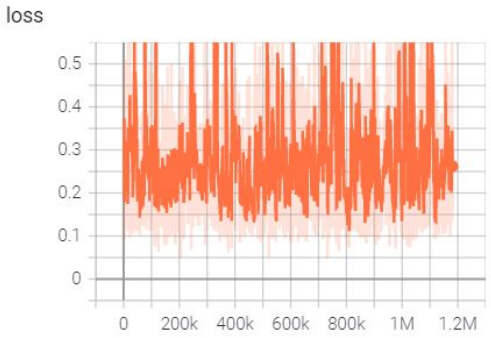


Figure 6. Training Loss

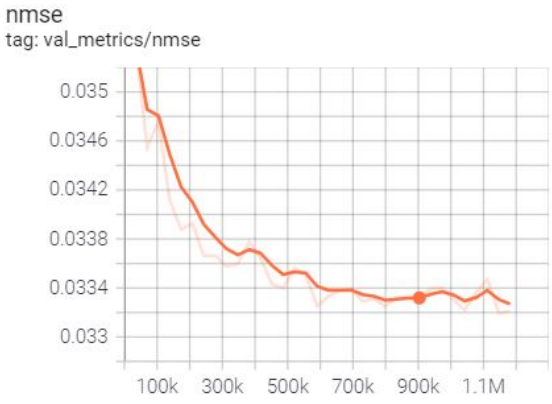


Figure 7. Normalized Mean Square Error over the Validation Set

The Loss used during training is L1 Loss

We can see from the Graph that the Normalized mean square error decrease and them starts saturating after 900K steps and then.

Similarly we see that NMSE error also follows the same pattern and becomes stable after 1000K steps.

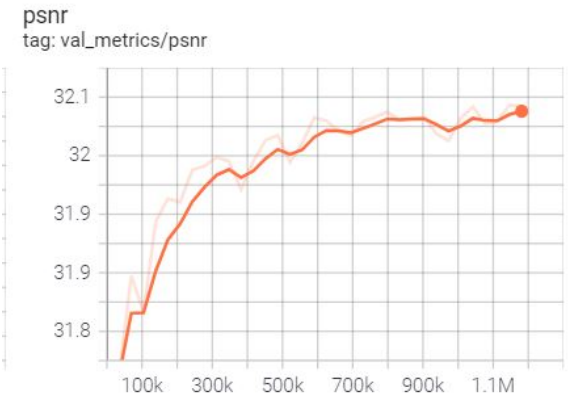


Figure 8. Peak Signal to Noise Radio over the Validation Set

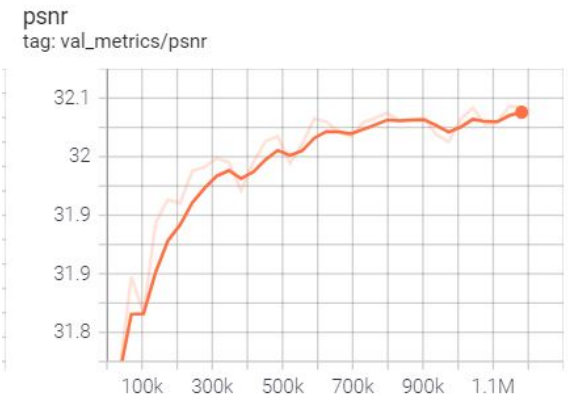


Figure 9. Structural Similarity over the Validation Set

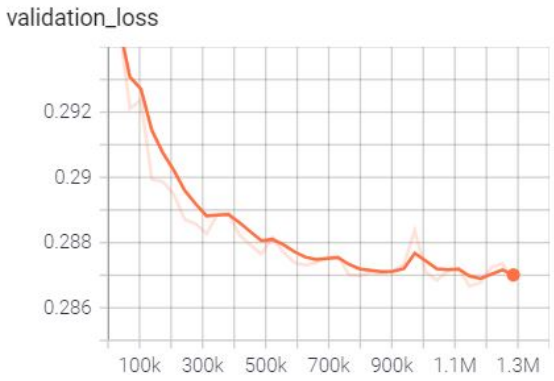
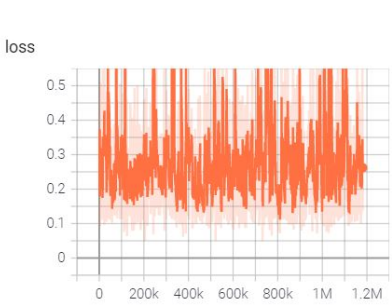


Figure 10. Validation Loss

The Peak Signal to Noise Ratio and the Structural Similarity also keeps improving in the training and becomes saturated 1000K steps.

If we view the reconstruction image and ground reality image from the validation set we see significant improvement with respect to noise reduction.

Please find below a sample set of the reconstructed and ground reality(target) images and also the validation error



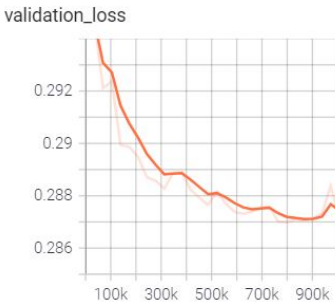


Figure 11. bla bla

image.

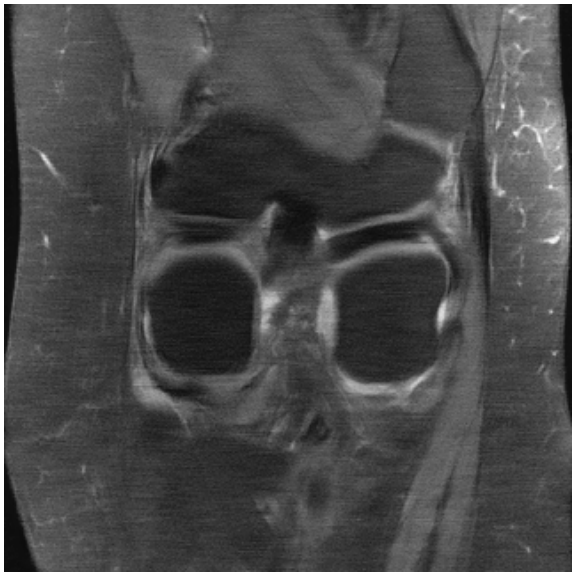


Figure 12. Reconstructed Image

3.2. Self-Supervised hold-out ResNet

Robert talks about the SSL model experiments and results here.

3.3. Self-Supervised ResNet other

Enrico talks about the SSL2 model experiments and results here.

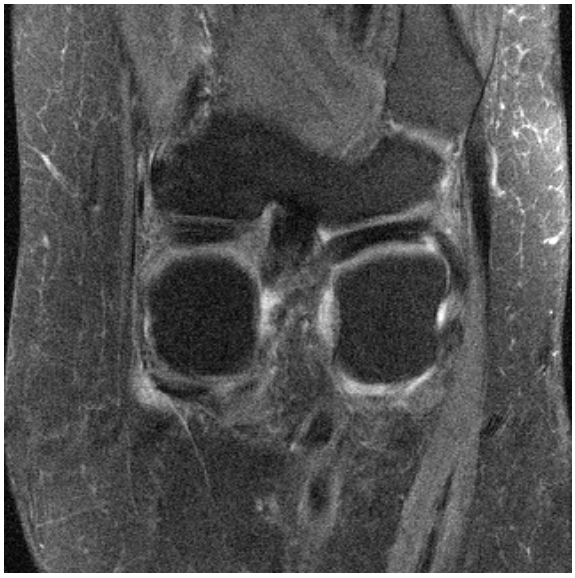


Figure 13. Ground Reality Image (target)

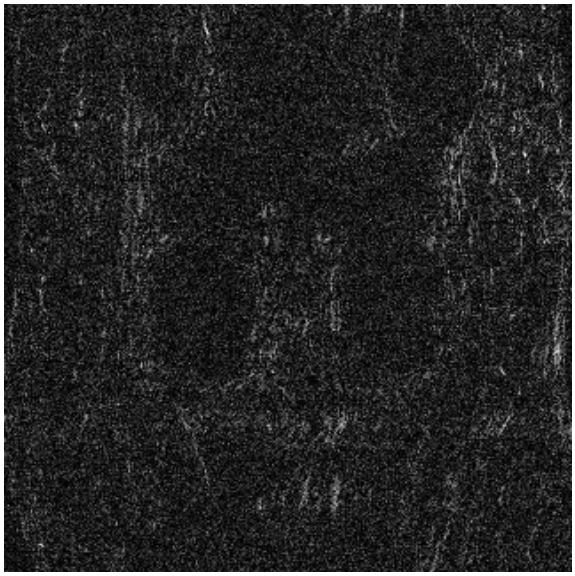


Figure 14. Error Image

4. Other Sections

You are welcome to introduce additional sections or subsections, if required, to address the following questions in detail.
(5 points) Appropriate use of figures / tables / visualizations. Are the ideas presented with appropriate illustration? Are the results presented clearly; are the important differences illustrated?
(5 points) Overall clarity. Is the manuscript self-contained? Can a peer who has also taken Deep Learning understand all of the points addressed above? Is sufficient detail provided?

(5 points) Finally, points will be distributed based on your understanding of how your project relates to Deep Learning. Here are some questions to think about:
What was the structure of your problem? How did the structure of your model reflect the structure of your problem?
What parts of your model had learned parameters (e.g., convolution layers) and what parts did not (e.g., post-processing classifier probabilities into decisions)?
What representations of input and output did the neural network expect? How was the data pre/post-processed?
What was the loss function?
Did the model overfit? How well did the approach generalize?
What hyperparameters did the model have? How were they chosen? How did they affect performance?
What optimizer was used?
What Deep Learning framework did you use?
What existing code or models did you start with and what did those starting points provide? Briefly discuss potential future work that the research community could focus on to make improvements in the direction of your project's topic.

Student Name	Contributed Aspects	Details
Enrico Zerilli	Project Setup, Abstract, Introduction and Self-Supervised method other	Scraped the dataset for this project and trained the CNN of the encoder. Implemented attention mechanism to improve results.
Robert Bartel	Self-Supervised Hold-out ResNet	Trained the LSTM of the encoder and analyzed the results. Analyzed effect of number of nodes in hidden state. Implemented Convolutional LSTM.
Sudipto Lodh	Fine-tuning of the supervised U-Net model	Trained the LSTM of the encoder and analyzed the results. Analyzed effect of number of nodes in hidden state. Implemented Convolutional LSTM.

Table 2. Contributions of team members.

5. Work Division

Please add a section on the delegation of work among team members at the end of the report, in the form of a table and paragraph description. This and references do **NOT** count towards your page limit. An example has been provided in Table 2.

References

[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *CoRR*, abs/1512.03385, 2015. 3

[2] Florian Knoll, Jure Zbontar, Anuroop Sriram, Matthew J. Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, James Pinkerton, Duo Wang, Nafissa Yakubova, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastmri: A publicly available raw k-space and DICOM dataset of knee images for accelerated MR image reconstruction using machine learning. *Radiology: Artificial Intelligence*, 2(1):e190007, 2020. 2

[3] NYU School of Medicine and Facebook AI Research (FAIR). The fastMRI dataset. https://fastmri.med.nyu.edu/. Accessed: 2021-04-25. 2, 3

[4] NYU School of Medicine and Facebook AI Research (FAIR). The fastMRI repository. https://github.com/facebookresearch/fastMRI. Accessed: 2021-04-25. 3

[5] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *CoRR*, abs/1505.04597, 2015. 3, 4

[6] M. Uecker, P. Lai, M. J. Murphy, P. Virtue, M. Elad, J. M. Pauly, S. S. Vasanawala, and M. Lustig. ESPIRiT–an eigenvalue approach to autocalibrating parallel mri: where sense meets grappa. *Magnetic Resonance in Medicine*, page 990–1001, 2014. 2

[7] Burhaneddin Yaman, Seyed Amir Hossein Hosseini, Steen Moeller, Jutta Ellermann, Kâmil Uğurbil, and Mehmet Akçakaya. Self-supervised learning of physics-guided reconstruction neural networks without fully sampled reference data. *Magnetic Resonance in Medicine*, 84(6):3172–3191, Jul 2020. 3, 4, 5

[8] Jure Zbontar, Florian Knoll, Anuroop Sriram, Tullie Murrell, Zhengnan Huang, Matthew J. Muckley, Aaron Defazio, Ruben Stern, Patricia Johnson, Mary Bruno, Marc Parente, Krzysztof J. Geras, Joe Katsnelson, Hersh Chandarana, Zizhao Zhang, Michal Drozdzal, Adriana Romero, Michael Rabbat, Pascal Vincent, Nafissa Yakubova, James Pinkerton, Duo Wang, Erich Owens, C. Lawrence Zitnick, Michael P. Recht, Daniel K. Sodickson, and Yvonne W. Lui. fastMRI: An open dataset and benchmarks for accelerated MRI. 2018. 1, 2, 3, 4, 5

Figure 15. Example of a short caption, which should be centered.