Task-Based Assessment for Neural Networks: Evaluating Undersampled MRI Reconstructions based on Signal Detection

Joshua D. Herman & Rachel E. Roca, Advised by Dr. Angel R. Pineda
Mathematics Department, Manhattan College, The Bronx, NY

Manhattan College 1853

Mathematics of Medical Imaging (MoMI) at Manhattan College

NIH

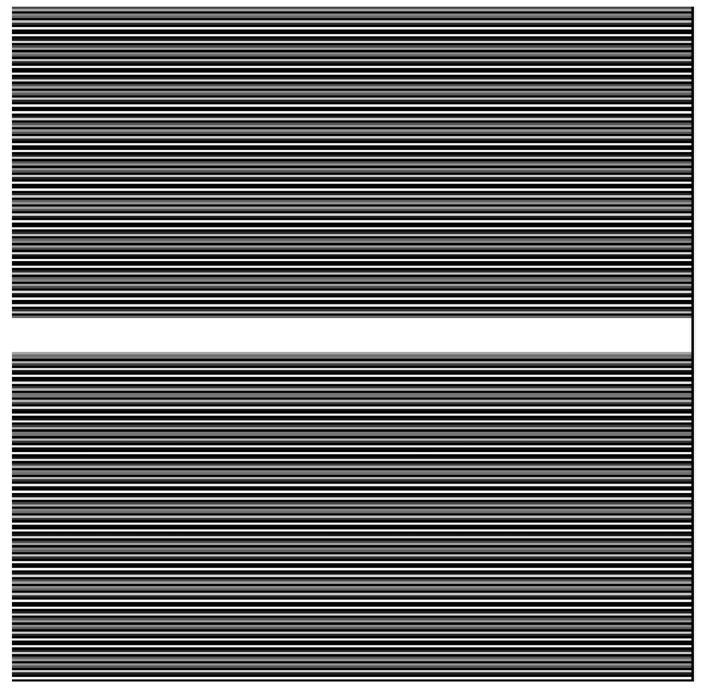
Introduction & Objective

- Magnetic Resonance Imaging (MRI) is widely used for noninvasively diagnosing many medical conditions, but it is time consuming to acquire the data.
- Much research has been done to accelerate MRI by undersampling the data, and recently neural networks have become an active research area.
- Subtle features are lost when MRI undersampling is increased, even after neural network reconstruction, and it is hard to quantify MR image quality. Most metrics are not tied directly to clinical usage.
- We built a U-Net [1] neural network that is evaluated on common metrics and on two 2AFC task-based metrics based on the clinical task of tumor signal detection.

Background: Terminology

- A **neural network** is a function filled with many tunable parameters that allows for a flexible set of output behavior based on parameter choices. This function is used to perform a specified task. The parameters can be tuned by applying an optimization algorithm to minimize a loss function representative of task performance
- **Masking** is the technique used to simulate **undersampling** of our MR images . Undersampling is specified with respect to **k-space**, the Fourier domain, and thus the process in MR imaging involves skipping a certain set of spacial frequencies when scanning anatomy. To simulate undersampling, we apply a sampling mask to our Fourier domain MR Image that removes some frequencies while keeping others
- **Observer** is a person or mathematical model that performs a specific task and has its performance reported. In this research we have a human and ideal linear model observer applied to a **2AFC** task.
- **2AFC task** or Two-Alternative Forced Choice task is a task in which there are two options, and the observer must try to pick the correct choice. In this research, the task is correctly picking the image that contains a tumor, from a pair of two images

Methods: Data Acquisition



- We are using four masks: 2x, 3x, 4x, and 5x
- A **kx** mask keeps the middle 16 frequency bands while keeping every k bands outside of the middle.

Figure 1: A 3x mask keeps the middle 16 frequency bands while keeping every 3 bands outside of the middle. Shown in white is what is kept, while what is shown is black is removed.

Methods: Neural Network Reconstruction

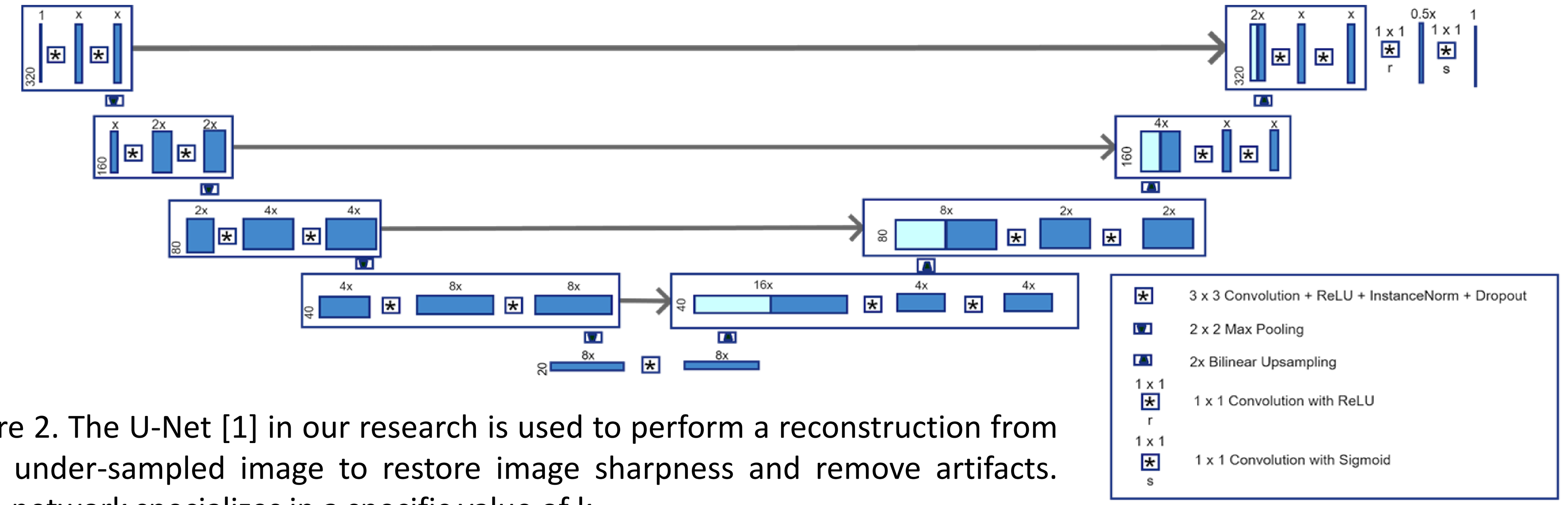


Figure 2. The U-Net [1] in our research is used to perform a reconstruction from a kx under-sampled image to restore image sharpness and remove artifacts. Each network specializes in a specific value of k.

- Constant hyperparameters: RMSProp with 150 epochs of batch size 16, x = 64 channels, dropout rate = .1, ReLU activation functions, with a sigmoid at the end, loss function was 1-structural similarity (SSIM)[3].
- The U-Net was trained on 500 images with each level of undersampling and used to reconstruct 50 testing images: with and without artificial tumors planted at 4 different locations in each image.
- 5-fold cross validation was used on the set of 500 images to generate scores for SSIM and NRMSE.
- The fastMRI Dataset [1] was used to generate the 550 images used in our study, each of which are 320 x 320 x 1 pixels.
- We worked on a Linux workstation using a Quadro P5000 16 GB CUDA GPU for training and reconstructing the images using the Tensorflow/Keras packages on Python.

Methods: Observer Studies

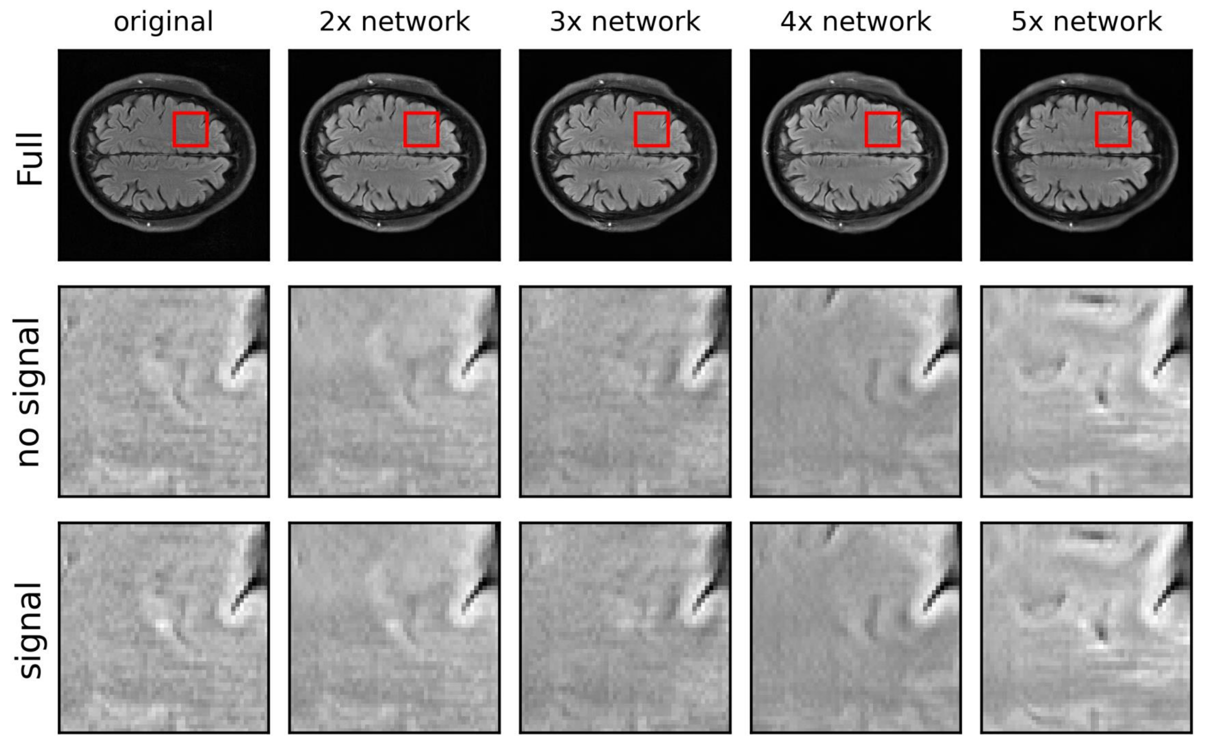


Figure 3. Shows single MR image in 1x-5x U-Net Reconstructions. The bottom two rows show the region boxed in red, with and without signal.

- For the observer studies [2] small tumor signals were added in 4 different locations in a set of 50 images, to create a set of 200 smaller images with a tumor in the center. The third row of images in Figure 3 have tumors in their centers.
- We are using a human observer [2] to perform the 2AFC task, which is just a person who repeats the task 200 times and reports on the percent correct. We used 4 of these observers and recorded mean and standard deviation across these four.
- We are also using an Ideal Linear Model Observer which approximates ideal machine performance. It specifically is a Channelized Hotelling Observer with Laguerre-Gauss Channels (LG- AUC) [4]. We report on the AUC of this model, which gives an estimate of percent correct.

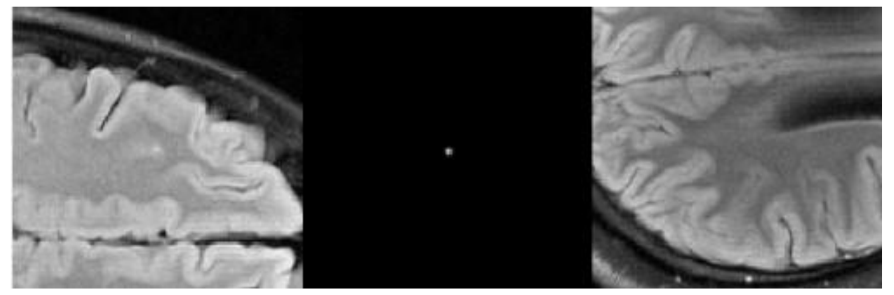


Figure 4. Sample 2AFC trial: The right and left images are our choices, while the middle image is the isolated signal. An observer is presented with 200 of these and must chose the anatomical image containing the signal in its center.

Results

Unet 64 channels 0.1 dropout	SSIM	NRMSE	LG AUC	Human Observer
Full recon	-	-	98.2/0.7	96.3/1.8
2x	0.907/0.004	0.142/0.014	95.7/1.1	**93.5/1.6**
3x	**0.905/0.006**	**0.137/0.007**	**94/1.4**	83.9/2.8
4x	0.831/0.008	0.179/0.011	86.7/2.1	74.5/2.1
5x	0.807/0.021	0.196/0.013	83/2.7	60.3/5.3

Table 1. Results for all under-sampling rates, in the form mean/standard deviation.

- A small change in metric performance occurs in SSIM from 2x to 3x undersampling rates, while a large drop occurs from 3x to 4x. Thus 3x is a reasonable undersampling rate based on SSIM
- The same reasoning leads to us determining that 3x is reasonable for NRMSE and LG-AUC, while 2x is reasonable for the human observer.

Conclusion

- A human observer may require a more conservative undersampling rate than that deemed reasonable by an ideal observer or more standard metrics.

Future Work

- Use a mathematical model to approximate human observers
- Use a task where the tumor is in one of several possible parts of the image as opposed to only the center
- Using a neural network model that includes MRI Physics

References

1. J Zbontar, et al, “fastMRI: An Open Dataset and Benchmarks for Accelerated MRI”, arXiv: 1811.08839v1 (2018)
2. CK Abbey, HH Barrett, “Human-and model-observer performance in ramp-spectrum noise: effects of regularization and object variability”, JOSA A, 18: 1237-1242 (1994)
3. Z Wang, et al, “Image quality assessment: from error visibility to structural Similarity”, IEEE TMI, 13: 600-612 (2004)
4. AR Pineda, "Laguerre-Gauss and sparse difference-of-Gaussians observer models for signal detection using constrained reconstruction in magnetic resonance imaging", Proc. SPIE 10952:53-58 (2019)

Acknowledgments

We thank Dr. Angel Pineda for his mentorship, Dr. Sajan Goud Lingala, Dr. Krishna Nayak, and Dr. Craig Abbey for their guidance, and Michael Rozycki for his technical support. We acknowledge support from the National Institute of Health under award number R15-EB029172.