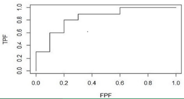
Task-Based Assessment for Neural Networks: Evaluating Undersampled MRI Reconstructions based on Signal Detection

Joshua D Herman[1] , Rachel E Roca[1] , Alexandra G O'Neill[1]
Sajan G Lingala[2] , and Angel R Pineda[1]

[1] Mathematics Department, Manhattan College, Riverdale, NY, United States
[2] Roy J. Carver Department of Biomedical Engineering, University of Iowa, Iowa City, IA, United States

Manhattan College Mathematics Seminar

March 23, 2021



Mathematics of Medical Imaging (MoMI) at Manhattan College

Introduction

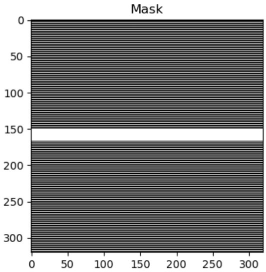
- Magnetic Resonance (MR) Imaging is time consuming
- Omitting frequencies during scanning (under-sampling) speeds up data acquisition
- Neural Networks have recently been applied to reconstruction
- Standard metrics of image quality do not incorporate the task
- In this research a task-specific evaluation method is introduced

Goals

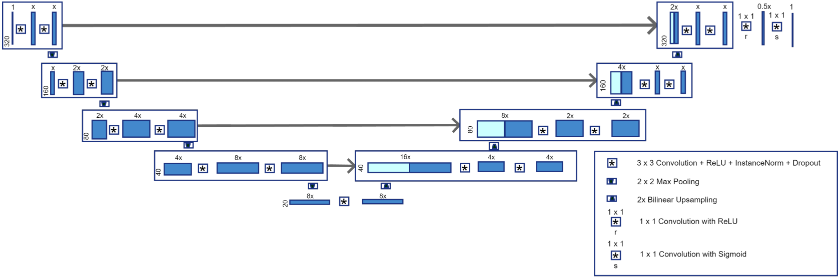
- Introduce 2AFC-task based evaluation of network reconstruction
- Pick ideal undersampling factor using new and standard metrics

Undersampling

- The FastMRI FLAIR (Fluid Attenuation Inversion Recovery) dataset is used to generate a set of fully sampled images
- The images are 320 by 320 pixels
- MR data in k-space (Fourier Transform of Image Space)
- Keep 16 middle k-space lines and sample every k outside of middle
- e.g. 3x sampling mask



Reconstruction



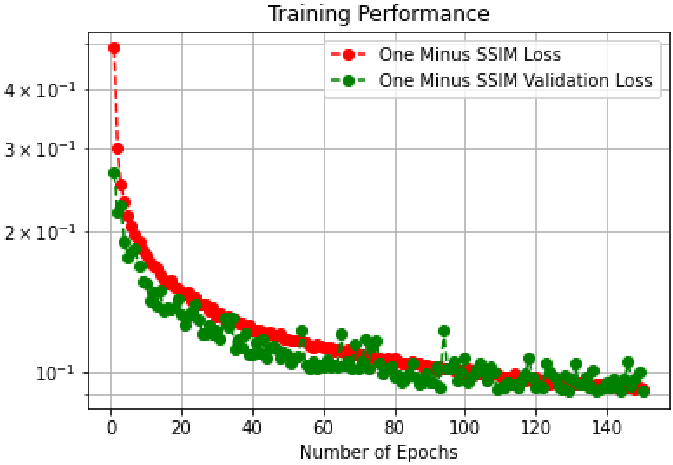
Unet with x channels

- 1-SSIM is used as the loss function
- RMSProp is used as the optimization algorithm

Computational Methods

- Used the Kakos Center's Linux workstation Athena
- Used a Quadro P5000 16 GB CUDA GPU.
- Used TensorFlow in Python to code the neural networks

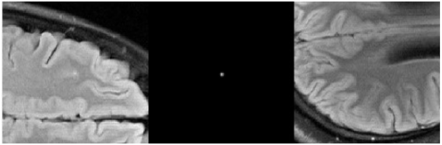
Training



Example Convergence Plot: 2x, split 2

2AFC task, observers and metrics

- Task: Pick correct image that has artificially placed tumor in center
- Sample 2AFC image



- Human observer study: given measure of human performance (200 trials per condition)
- Channelized Hotelling observer with Laguerre Gauss Channels: estimate of ideal machine performance
- NRMSE as general measure of pixel error
- SSIM as measure of image structural similarity

Cross Validation Results

Unet 64 channels 0.1 dropout	SSIM	NRMSE
2x	0.907/0.004	0.142/0.014
3x	**0.905/0.006**	**0.137/0.007**
4x	0.831/0.008	0.179/0.011
5x	0.807/0.021	0.196/0.013

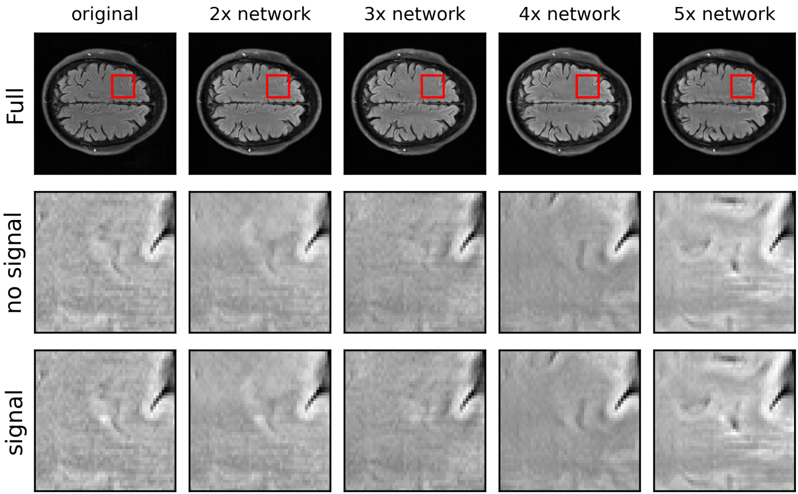
- Table containing 5-fold cross validation SSIM and NRMSE scores
- Format is mean/standard deviation of the 5 splits
- A reasonable undersampling factor would be 3x, since large performance drop from 3x to 4x

Testing Results

Unet 64 channels 0.1 dropout	LG AUC	Human Observer
Full recon	98.2/0.7	96.3/1.8
2x network	95.7/1.1	**93.5/1.6**
3x network	**94/1.4**	83.9/2.8
4x network	86.7/2.1	74.5/2.1
5x network	83/2.7	60.3/5.3

- Table containing Single Run LG AUC and Human Observer PC
- Format is mean/standard deviation
- Human observer means and standard deviations are calculated across the 4 participants, while LG AUC using bootstrap
- Based on LG AUC, a reasonable undersampling factor would be 3x
- Based on Human Observer, a reasonable undersampling factor would be 2x

Visual Comparison

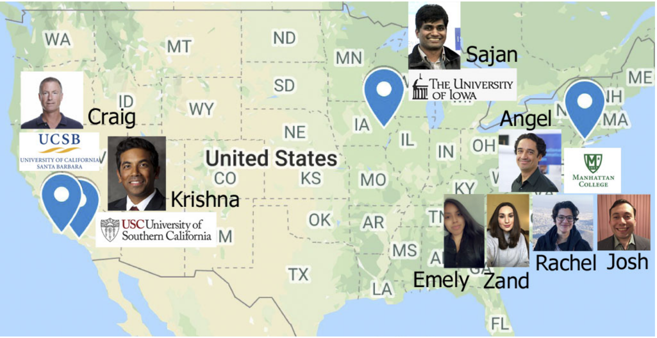


Picture 22 out of 0 -199

Conclusion and Future Work

- SSIM , NRMSE and Laguerre-Gauss metrics predict 3x undersampling factor to be best compromise between speed and performance
- Human observer predicts 2x undersampling factor as best compromise
- Preferred undersampling factor depends on the observer e.g. Human Observer PC may predict the undersampling better for a human
- Future work may include
  - model human observer
  - variable tumor location detection
  - larger dataset
  - MSE Loss for Unet
  - incorporate the physics of MRI to the neural network

Acknowledgments



Collaboration Map

We acknowledge support from NIH R15-EB029172 and thank Dr. Krishna S. Nayak and Dr. Craig K. Abbey for their thoughtful feedback.

Appendix

- Cross-validation done using crossvalmag111320/unet_run_cross_val_mag_110220.py