

# Modeling human observer detection in undersampled magnetic resonance imaging reconstruction with total variation and wavelet sparsity regularization

Alexandra G. O'Neill,<sup>a</sup> Emely L. Valdez,<sup>a</sup> Sajan Goud Lingala<sup>b</sup>,  
and Angel R. Pineda<sup>a,\*</sup>

<sup>a</sup>Manhattan College, Department of Mathematics, New York City, New York, United States

<sup>b</sup>University of Iowa, Roy J. Carver Department of Biomedical Engineering,  
Iowa City, Iowa, United States

## Abstract

**Purpose:** Task-based assessment of image quality in undersampled magnetic resonance imaging provides a way of evaluating the impact of regularization on task performance. In this work, we evaluated the effect of total variation (TV) and wavelet regularization on human detection of signals with a varying background and validated a model observer in predicting human performance.

**Approach:** Human observer studies used two-alternative forced choice (2-AFC) trials with a small signal known exactly task but with varying backgrounds for fluid-attenuated inversion recovery images reconstructed from undersampled multi-coil data. We used a 3.48 undersampling factor with TV and a wavelet sparsity constraints. The sparse difference-of-Gaussians (S-DOG) observer with internal noise was used to model human observer detection. The internal noise for the S-DOG was chosen to match the average percent correct (PC) in 2-AFC studies for four observers using no regularization. That S-DOG model was used to predict the PC of human observers for a range of regularization parameters.

**Results:** We observed a trend that the human observer detection performance remained fairly constant for a broad range of values in the regularization parameter before decreasing at large values. A similar result was found for the normalized ensemble root mean squared error. Without changing the internal noise, the model observer tracked the performance of the human observers as the regularization was increased but overestimated the PC for large amounts of regularization for TV and wavelet sparsity, as well as the combination of both parameters.

**Conclusions:** For the task we studied, the S-DOG observer was able to reasonably predict human performance with both TV and wavelet sparsity regularizers over a broad range of regularization parameters. We observed a trend that task performance remained fairly constant for a range of regularization parameters before decreasing for large amounts of regularization.

© 2023 Society of Photo-Optical Instrumentation Engineers (SPIE) [DOI: [10.1117/1.JMI.10.1.015502](https://doi.org/10.1117/1.JMI.10.1.015502)]

**Keywords:** model observers; magnetic resonance imaging; constrained reconstruction; image quality assessment.

Paper 22180GR received Jul. 14, 2022; accepted for publication Feb. 6, 2023; published online Feb. 25, 2023.

## 1 Introduction

Task-based assessment of image quality<sup>1,2</sup> for reconstructed magnetic resonance imaging (MRI) images is critical to the development and validation of accelerated reconstruction techniques. This approach has been extensively used in other imaging modalities<sup>3-12</sup> There has been research using task-based techniques in MRI for parallel imaging<sup>13</sup> and compressed sensing<sup>14</sup> but most methods utilized for assessment of image quality in MRI are typically root mean square error

---

\*Address all correspondence to Angel R. Pineda, [angel.pineda@manhattan.edu](mailto:angel.pineda@manhattan.edu)

(RMSE) and structural similarity index (SSIM). These are measures of pixel value differences between the original and reconstructed image.<sup>15,16</sup> While these measures do give an indication of similarity between images, neither RMSE nor SSIM take into account the specific task for which the image will be used. As a result, images with the same RMSE and SSIM could produce different performance in these tasks. Observer models are an alternative way to assess image quality by taking into account human visual principles as well as the task for which the image will be used.<sup>17-19</sup>

Previous results related to evaluating MRI reconstruction of undersampled images using an approximation to the ideal linear observer<sup>20-22</sup> showed that the area under the ROC curve for the channelized Hotelling observer with Laguerre Gauss channels showed only a small improvement with regularization. However, both SSIM and RMSE showed a large improvement. Our preliminary results using human observer studies<sup>23,24</sup> suggest a similar result in that there was no large improvement with regularization for this task. The purpose of this work is to evaluate constrained reconstruction of undersampled MRI data using TV, wavelet, and a combination of these constraints based on human observer performance in detecting a small signal in a two-alternative forced choice (2-AFC) task. A model observer sparse difference-of-Gaussians (S-DOG) is used to model human observer performance in this detection task. Along with optimizing reconstruction, a goal of this work is also to reduce the number of future human observer studies needed in undersampled MRI reconstruction using model observers to evaluate image quality.

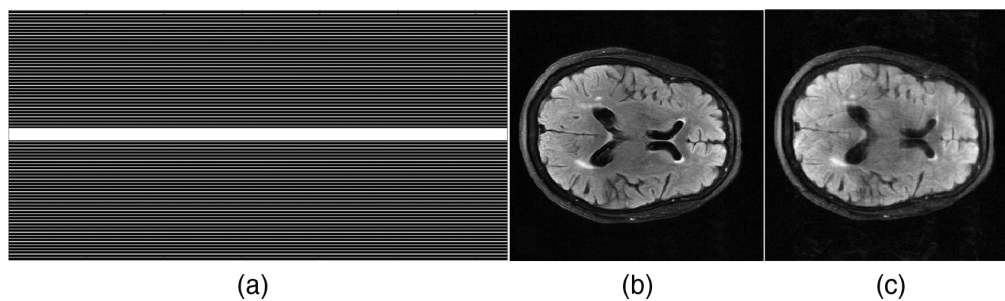
## 2 Methods

### 2.1 Undersampled Acquisition in MRI

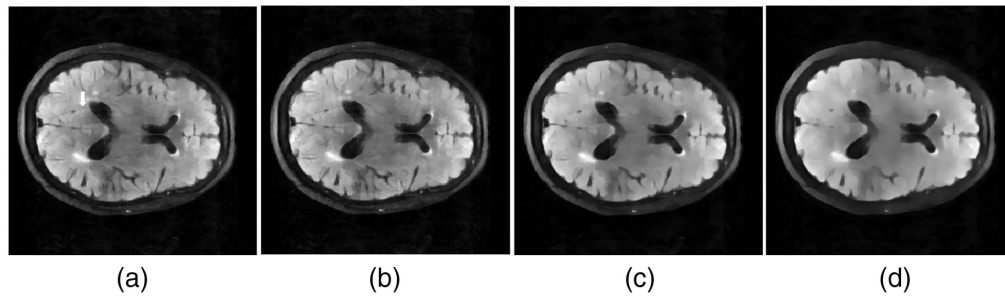
For this study, we consider one-dimensional undersampling of fluid-attenuated inversion recovery images (Fig. 1) with an acquisition that samples every fourth phase encoding line plus fully sampling the middle 16 kspace lines resulting in an effective undersampling factor of 3.48. Data used in the preparation of this article were obtained from the New York University/NYU Langone Health (NYU) fastMRI initiative database.<sup>25</sup> The average white matter signal to noise ratio (SNR) for the fully sampled multicoil sensitivity encoding (SENSE) reconstructions ( $R = 1$ ) was measured by estimating the standard deviation from components of the image with only noise and assuming a Rayleigh distribution<sup>26</sup> and the mean signal across a homogeneous region of white matter image in the reconstructed slices. This led to an average white matter SNR of 83 in the 50 slices used for the observer studies. The NYU fastMRI investigators provided data but did not participate in analysis or writing of this report.

### 2.2 Constrained Reconstruction from Multi-coil Data

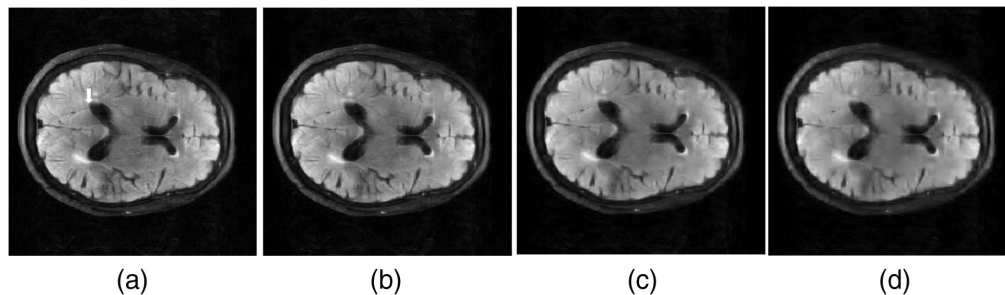
Constrained reconstruction minimizes a data agreement functional with additional constraints. For this work we consider a total variation (TV) and wavelet constraint<sup>16</sup> and multicoil parallel



**Fig. 1** (a) Sampling mask for 4× undersampling, (b) fully sampled image, and (c) undersampled constrained reconstruction with no TV regularization, aliasing in the vertical axis is visible in the undersampled image. The acquisition had a 2× oversampling in the horizontal direction.



**Fig. 2** Sample undersampled images reconstructed with TV regularization, (a)  $\alpha_{TV} = 0.01$ , (b)  $\alpha_{TV} = 0.02$ , (c)  $\alpha_{TV} = 0.05$ , and (d)  $\alpha_{TV} = 0.1$ . As the regularization increases, there are reduced aliasing artifacts but also reduced resolution. The arrow in panel (a) shows the location of one of the undersampling artifacts.



**Fig. 3** Sample undersampled images reconstructed with wavelet sparsity: (a)  $\alpha_W = 0.01$ , (b)  $\alpha_W = 0.02$ , (c)  $\alpha_W = 0.05$ , and (d)  $\alpha_W = 0.1$ . As the regularization increases, there are reduced aliasing artifacts but also reduced resolution but with different texture than in TV regularization. The arrow in image A shows the location of one of the undersampling artifacts.

imaging using SENSE<sup>27</sup> with the coil sensitivities estimated using the sum of squares method, which leads to real estimates of real-valued objects.

The following functional was minimized when reconstructing these images

$$\|H(f) - g\|_2^2 + \alpha_\Psi \|\Psi(f)\|_1 + \alpha_{TV} \|\nabla f\|_1, \quad (1)$$

where  $H(f)$  is the undersampled Fourier operator acting on the discretized object  $f$ , the Fourier data are  $\mathbf{g}$ ,  $\Psi$  is the Daubechies 4 wavelet transform,  $\|\nabla f\|_1$  is the  $L_1$  norm of the gradient which is the TV operator and  $\alpha_\Psi$  and  $\alpha_{TV}$  are the regularization parameters for the wavelet and TV constraints respectively.<sup>16</sup> Sample images with different amounts of TV and wavelet regularization are shown in Figs. 2 and 3, respectively. The reconstruction of the images was done using the Berkeley Advanced Reconstruction Toolbox toolbox.<sup>28</sup> The ensemble RMSE (ERMSE) for the reconstructed images was computed using the fully sampled unregularized reconstruction normalized to [0,1] as the reference image. The underampled reconstructions were normalized to [0,1] before computing the RMSE. The ERMSE was computed from 50 slices from five volumes.

### 2.3 Two-Alternative Forced Choice Experiments

In each individual trial of the 2AFC experiment we presented three  $128 \times 128$  pixel images: one image of an anatomical background with the signal, the signal, and one image of an anatomical background without the signal. The signal image was always in the center, and the location (left or right) of the anatomical image with the signal was randomly chosen for each trial. All images are scaled to [0,1] before being displayed using the 8-bit grayscale colormap in MATLAB. Because the observers were not radiologists, the images were windowed and leveled for them. Our choice of windowing leads to a different number of grayscale values representing the signals

depending on the backgrounds and regularization. In all cases, a reasonable number of grayscale values was used to represent the signal contrast with an average of 32 grayscale values for images with no regularization down to an average of 8 grayscale values for images with the most regularization ( $\alpha_{TV} = 0.1$ ,  $\alpha_W = 0.1$ ). An example trial is shown in Fig. 4. The signal location is always in the middle of the anatomical image, which makes this task a signal known exactly with varying backgrounds and the human observer only determines whether or not it is present in the image.

## 2.4 Experimental Procedure

For each experimental condition, 200 2AFC trials were carried out by four observers. As training, all observers repeated an initial set of 200 trials until the performance plateaued. Based on performance in the training trials, the signal amplitude was chosen so that the mean percent correct (PC) for the observers would be close to 90% when no regularization was used.

The observer studies were done using a Barco MDRC 2321 monitor in a dark room. The resolution of the Barco monitor was 0.294 mm and the observers were ~50 cm from the screen. A marker at 50 cm was provided for reference.

For each individual trial, a set of images like Fig. 4 appeared on the screen and the observer chose which image they believed contained the signal. The observers received feedback on whether they had identified the signal correctly after each trial for possible improvement between trials and took breaks between sets of 200 trials to avoid fatigue.

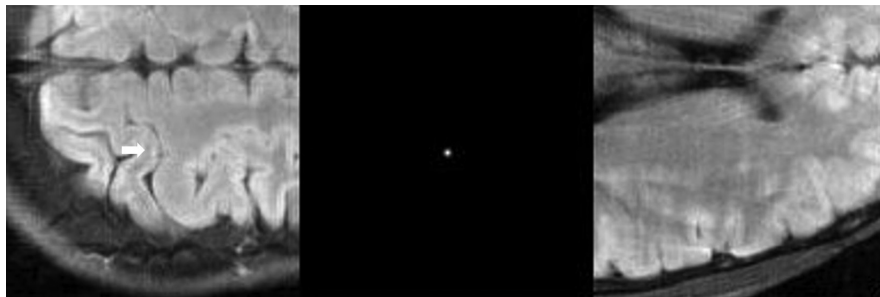
For each observer, the images used in each 2-AFC trial were randomly paired from 200 backgrounds with the signal and 200 background images. For each level of regularization, the standard deviation was computed using the PC of each of the four observers. There are several sources of variability in this estimate of the PC due to the multiple reader–multiple case (MRMC) experimental design. Our estimate of the standard deviation based on the PC of the individual observers is a summary of the variability of the PC values from the readers but not based on an unbiased estimate.<sup>29</sup>

## 2.5 Model Observers

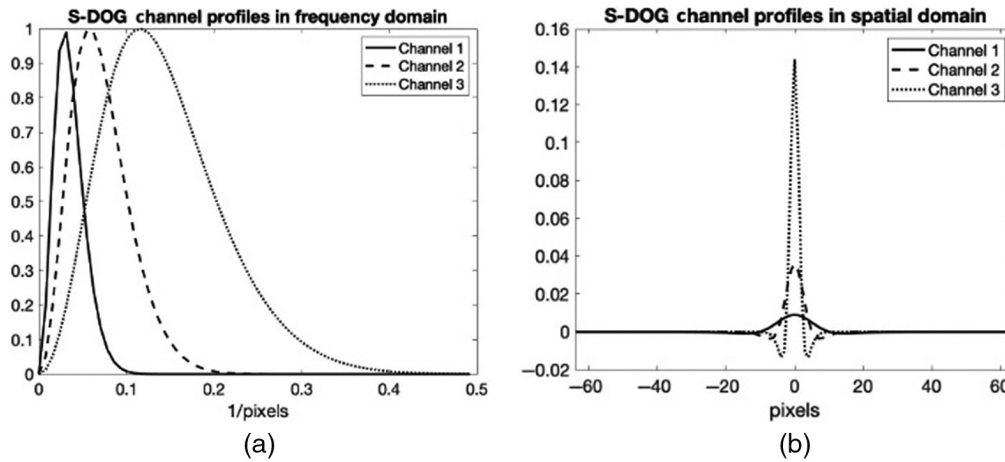
The model that we used in our study was the S-DOG channelized Hotelling Observer.<sup>17</sup> The channel data  $\mathbf{g}_c$  are obtained by taking the inner product of the image with the channels centered at the signal. The elements  $g_{c,j}$  of the vector  $\mathbf{g}_c$  are given as

$$g_{c,j} = C_j^t f, \quad (2)$$

where  $C_j$  is the  $j$ 'th channel in the spatial domain and  $f$  is the object in the spatial domain. We used the S-DOG channels of the form



**Fig. 4** Sample 2AFC trial with signal in the left image. The arrow present in this image was not used in the observer study and is there to help the reader identify the signal. In order to see differences in performances due to regularization, the images used were subtle. These images were generated with no regularization but the artifacts are more visible in the image to the right.



**Fig. 5** (a) The cross section of S-DOG channels matched for human performance with no TV regularization in the frequency domain and (b) spatial domain.

$$C_j(k) = \exp\left[-\frac{1}{2}\left(\frac{k}{Q * \sigma_j}\right)^2\right] - \exp\left[-\frac{1}{2}\left(\frac{k}{\sigma_j}\right)^2\right], \tag{3}$$

in the frequency domain where  $k$  is the distance from the zero frequency,  $Q$  is the multiplicative factor of the bandwidth,  $\sigma_j = \sigma_0 * \alpha^j$ , where  $j$  denotes the  $j$ 'th channel and  $\sigma_j$  denotes the standard deviation of each channel. The parameters that were used for the S-DOG were  $Q = 2$ ,  $\alpha = 2$ , and  $\sigma_0 = 0.015$  which were used by Abbey<sup>17</sup> (Fig. 5).

The test statistic ( $\lambda$ ) for the S-DOG observer for each image was computed using

$$\lambda = [(K_c + K_{int})^{-1} s_c]^t (g_c + n_{int}), \tag{4}$$

where  $K_c$  is the sample covariance matrix of the data in the channel domain,  $K_{int}$  is the covariance matrix of the channel internal noise,  $s_c$  is the mean signal difference in the channel domain,  $g_c$  is the image in the channel domain and  $n_{int}$  is a noise vector drawn from the distribution  $N(0, K_{int})$ .

The covariance for the internal noise is

$$K_{int} = \epsilon \text{diag}(K_c), \tag{5}$$

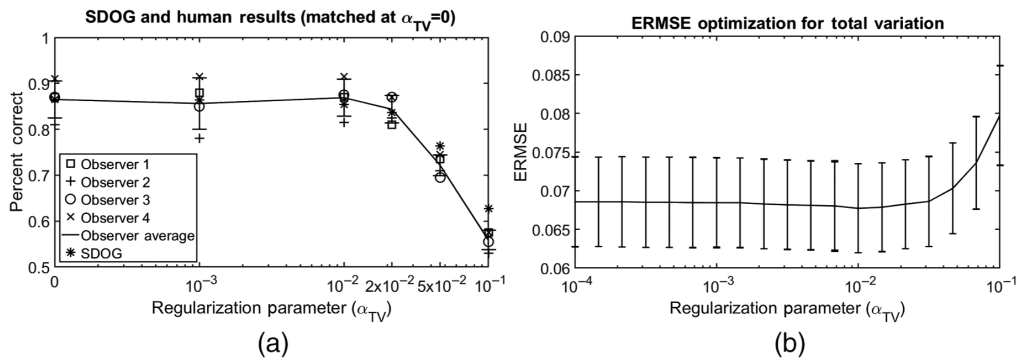
where  $\epsilon$  is the internal noise constant and  $\text{diag}(K_c)$  is a diagonal matrix with the diagonal elements of  $K_c$ .

The channel covariance matrix  $K_c$  is computed by taking the average of the channel covariance matrix with and without the signal. The mean signal difference ( $s_c$ ) is computed by subtracting the sample average of the channel outputs without the signal from the average channel outputs with the signal. The internal noise was determined by varying the noise constant  $\epsilon$  values until the performance matched the average human performance for the unregularized reconstruction. This calibrated model with the images with no regularization was used to predict all other experimental conditions with regularization.

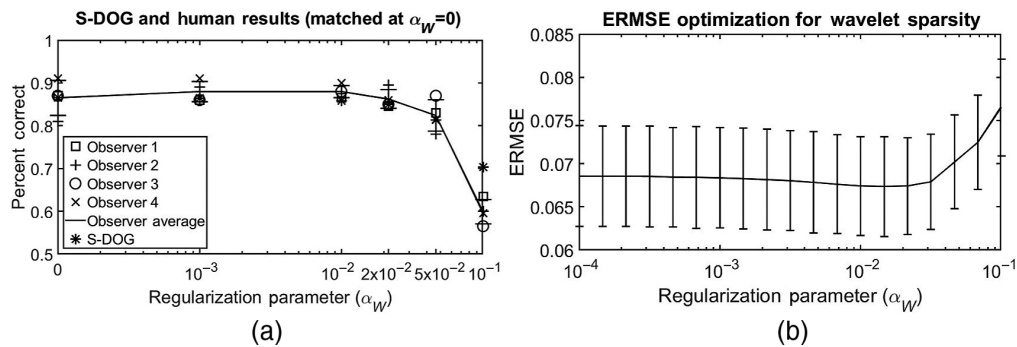
We utilized six regularization parameter values for each of the three regularization conditions that were generated with 4x undersampling. For TV and wavelet regularization, the parameters used were: 0, 0.001, 0.002,  $2 * 10^{-1}$ ,  $5 * 10^{-1}$ , and 0.1. These values were also used when both types of regularization were combined.

### 3 Results and Discussion

The S-DOG overestimated human performance without the addition of internal noise. However, the pattern of performance as TV increased was the same as both the human observers. Once the internal noise was added, tracking the human performance can be seen with the S-DOG in Fig. 6.



**Fig. 6** (a) S-DOG observer with internal noise matched to average human observer for images with no TV regularization. The S-DOG observer tracked the human observer data for the other regularization parameters. The S-DOG slightly overestimates for the largest regularization parameters. (b) Normalized ERMSE for varying TV regularization. The ERMSE metric leads to similar conclusions as the 2-AFC detection performance.



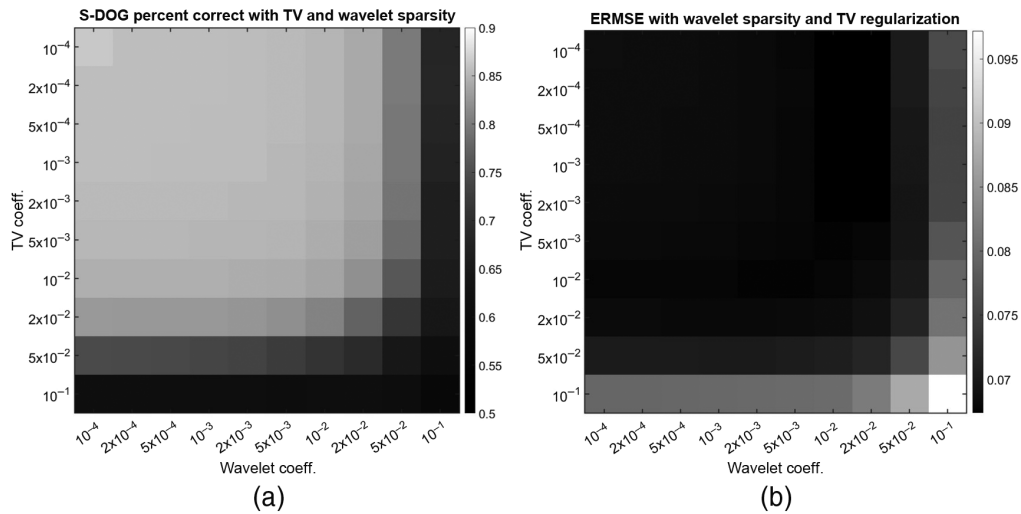
**Fig. 7** (a) S-DOG observer with internal noise matched to average human observer for images with no wavelet sparsity. The S-DOG observer tracked the human observer data for the other regularization parameters. The S-DOG slightly overestimates for the largest regularization parameter. (b) Normalized ERMSE for varying wavelet sparsity. The ERMSE metric leads to similar conclusions as the 2-AFC detection performance.

The S-DOG closely tracks human performance for small amounts of regularization (which would be used in practice) and over-estimates the human performance for large amounts of regularization.

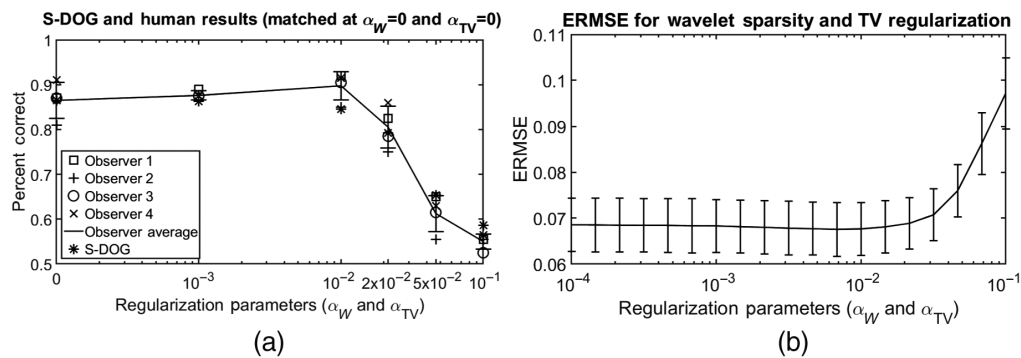
Similarly, the S-DOG is also able to track human observer performance for varying wavelet regularization but again overestimates performance at the largest values of regularization as seen in Fig. 7. Wavelet regularization was not found to meaningfully improve human performance in this task. Lastly, the combination of wavelet and TV regularization does not lead to a large improvement in human performance for this detection task. Figure 8 shows the prediction for performance based on the S-DOG before doing the observer study where the two types of regularization were combined. In Fig. 9, the predictions were validated using an observer study. The same S-DOG model was used in this plot as in Figs. 6–8. The ERMSE showed a similar behavior for all these types of regularization. In all cases, the internal noise was chosen so that the model observer matched the average human observer performance with no regularization.

It is difficult to quantify task performance based on subjective evaluation of image quality by looking at the images. Figure 10 shows sample images with a signal for different levels of regularization. Previous work showed that the MSE and SSIM may change in a meaningful way with regularization<sup>22</sup> but the task performance and visual assessment remain similar.

This study considered undersampling at a high SNR. The effect of regularization on detection performance for with varying backgrounds could be different at lower SNR regimes. Some slight improvement was seen in the context of ramp-spectrum noise for human observers<sup>17</sup> and in ideal observer performance in undersampled MRI.<sup>22</sup> Recent work studying the effect of denoising using



**Fig. 8** (a) The S-DOG observer predicts that even with a combination of wavelet and TV regularization, the average human observer performance remains fairly constant for a range of regularization parameters and that it degrades at high levels of regularization. (b) The ERMSE behaves leads to similar conclusions as the S-DOG observer.



**Fig. 9** (a) The S-DOG observer predicts that even with a combination of wavelet and TV regularization, the average human observer performance remains fairly constant for a range of regularization parameters and that it degrades at high levels of regularization. (b) The ERMSE behaves leads to similar conclusions as the S-DOG observer.

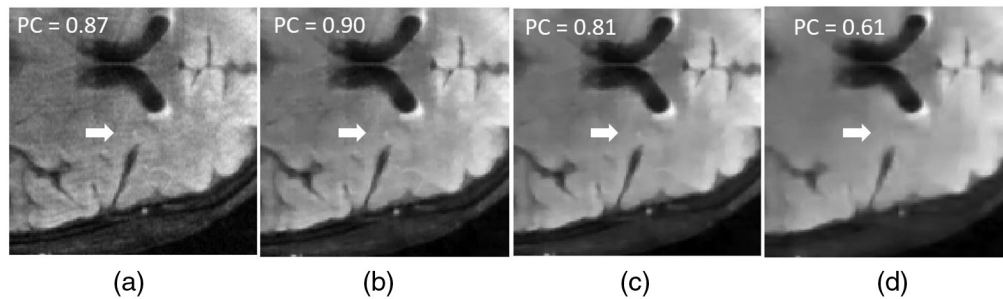
neural networks on detection performance in single positron emission computed tomography showed that denoising could decrease detection performance even if it improved RMSE and SSIM.<sup>30</sup> In another study using simulated images showed similar results.<sup>31</sup> Exploring applications where regularization helps in a detection-based perspective would be useful to better understand the regimes where metrics like ERMSE and detection performance agree and disagree.

The visibility of the signal varies depending on the background. Figure 11 shows the variability in the sample images for a single regularization value.

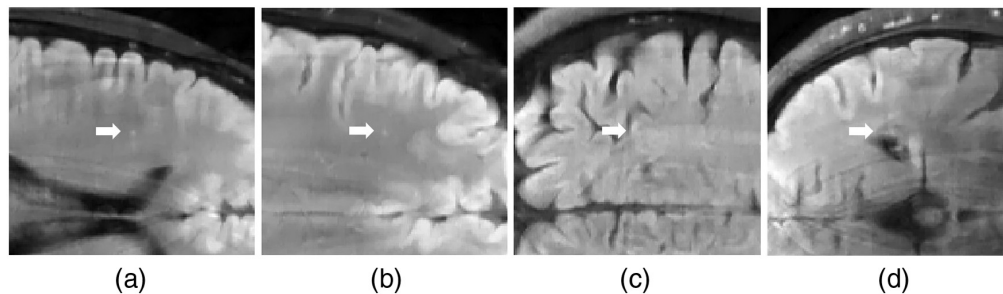
A limitation of this study is the experimental design which paired the images in the 2-AFC study randomly for each of the observers. This leads to complicated correlations within the cases and readers since readers may share the signal image in a 2-AFC trial but with different background images, for example. To our knowledge, current approaches for unbiased estimates of variance do not include such an experimental design.<sup>32,33</sup>

## 4 Conclusion

To our knowledge, this is the first application of a model observer for tracking human observer detection performance in undersampled MRI reconstruction. The S-DOG model observer was



**Fig. 10** Sample subimages with lesions undersampled images reconstructed with TV and wavelet regularization, (a)  $\alpha_{TV} = 0.0$ ,  $\alpha_W = 0.0$ , (b)  $\alpha_{TV} = 0.01$ ,  $\alpha_W = 0.01$ , (c)  $\alpha_{TV} = 0.02$ ,  $\alpha_W = 0.02$ , and (d)  $\alpha_{TV} = 0.05$ ,  $\alpha_W = 0.05$ . As the regularization increases, there are reduced aliasing artifacts but also reduced resolution. Image quality and task performance (as measured by the average PC shown in the images) is difficult to assess from the sample images.



**Fig. 11** Sample subimages with lesions undersampled images reconstructed with TV and wavelet regularization ( $\alpha_{TV} = 0.01$ ,  $\alpha_W = 0.01$ ). Even though the signals are always present, subjectively we see that the signal is easier to detect in panels (a) and (b) but harder to see in panels (c) and (d). This variability in visibility due to background variability for the same reconstruction is one of the reasons why it is important to average over multiple backgrounds to estimate performance.

able to track human performance for TV and wavelet constraints, as well as the combination of both TV and wavelet. One of the results of this study is that model observers are able to track human observer performance as the regularization changes. We also observed a trend that commonly used regularization parameters of TV, wavelet, and combination TV and wavelet constraints led to a plateau in detection performance for low levels of regularization before causing a drop in performance for large levels of regularization. This suggests that within that plateau, the regularization parameter could be chosen by some other criteria without affecting task performance.

## Disclosures

No conflicts of interest, financial or otherwise, are declared by the authors.

## Acknowledgments

This work was supported by the National Institute of Biomedical Imaging and Bioengineering of the National Institutes of Health (Grant No. R15-EB029172, the Manhattan College Faculty Development Grant and the Kakos Center for Scientific Computing. The authors thank Dr. Krishna S. Nayak at the University of Southern California and Dr. Craig K. Abbey at University of California, Santa Barbara for their time and guidance. The authors would also like to thank the reviewers for their thoughtful comments which improved the manuscript and identified areas of future work.



## Data, Materials and Code Availability

The code used for placing the signals in the raw k-space data, and running the observer studies and model observers can be found at: <https://github.com/MoMI-Manhattan-College/MRI-Signal-Detection>.

## References

1. H. H. Barrett, "Objective assessment of image quality: effects of quantum noise and object variability," *J. Opt. Soc. Am. A* **7**, 1266–1278 (1990).
2. H. H. Barrett and K. J. Myers, *Foundations of Image Science*, John Wiley and Sons, Hoboken, NJ (2004).
3. L. Yu et al., "Correlation between a 2D channelized hotelling observer and human observers in a low-contrast detection task with multislice reading in CT," *Med. Phys.* **44**(8), 3990–3999 (2017).
4. C. P. Favazza et al., "Use of a channelized Hotelling observer to assess CT image quality and optimize dose reduction for iteratively reconstructed images," *J. Med. Imaging* **4**(3), 031213 (2017).
5. H.-W. Tseng, J. Fan, and M. A. Kupinski, "Design of a practical model-observer-based image quality assessment method for x-ray computed tomography imaging systems," *J. Med. Imaging* **3**(3), 035503 (2016).
6. G. J. Gang et al., "Analysis of Fourier-domain task-based detectability index in tomosynthesis and cone-beam ct in relation to human observer performance," *Med. Phys.* **38**(4), 1754–1768 (2011).
7. H. C. Gifford et al., "Evaluation of multiclass model observers in PET LROC studies," *IEEE Trans. Nucl. Sci.* **54**(1), 116–123 (2007).
8. J. G. Brankov, "Evaluation of the channelized hotelling observer with an internal-noise model in a train-test paradigm for cardiac SPECT defect detection," *Phys. Med. Biol.* **58**, 7159–7182 (2013).
9. J. Baek, A. R. Pineda, and N. J. Pelc, "To bin or not to bin? The effect of CT system limiting resolution on noise and detectability," *Phys. Med. Biol.* **58**, 1433 (2013).
10. A. K. Jha, E. Clarkson, and M. A. Kupinski, "An ideal-observer framework to investigate signal detectability in diffuse optical imaging," *Biomed. Opt. Express* **4**, 2107–2123 (2013).
11. A. R. Pineda et al., "Information content of data types in time-domain optical tomography," *J. Opt. Soc. Am. A* **23**, 2989–2996 (2006).
12. A. Pineda et al., "Optimization of a tomosynthesis system for the detection of lung nodules," *Med. Phys.* **33**, 1372–1379 (2006).
13. Y. Jiang, D. Huo, and D. L. Wilson, "Methods for quantitative image quality evaluation of MRI parallel reconstructions: detection and perceptual difference model," *Magn. Reson. Imaging* **25**(5), 712–721 (2007).
14. C. G. Graff and E. Y. Sidky, "Compressive sensing in medical imaging," *Appl. Opt.* **54**, C23–C44 (2015).
15. Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Trans. Image Process.* **13**, 600–612 (2004).
16. M. Lustig, D. Donoho, and J. M. Pauly, "Sparse MRI: the application of compressed sensing for rapid mr imaging," *Magn. Reson. Med.* **58**, 1182–1195 (2007).
17. C. K. Abbey and H. H. Barrett, "Human-and model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *J. Opt. Soc. Am. A* **18**, 473–488 (2001).
18. A. E. Burgess, "Statistically defined backgrounds: performance of a modified nonprewhitening observer model," *J. Opt. Soc. Am. A* **11**(4), 1237–1242 (1994).
19. K. J. Myers and H. H. Barrett, "Addition of a channel mechanism to the ideal-observer model," *J. Opt. Soc. Am. A* **4**, 2447–2457 (1987).
20. Y. Chen et al., "Task-based data-acquisition optimization for sparse image reconstruction systems," *Proc. SPIE* **10136**, 264–269 (2017).

21. A. R. 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).
22. A. Pineda et al., "Optimizing constrained reconstruction in magnetic resonance imaging for signal detection," *Phys. Med. Biol.* **66**, 145014 (2021).
23. A. G. O'Neill et al., "Modeling human observer detection in undersampled magnetic resonance imaging (MRI)," *Proc. SPIE* **11599**, 85–90 (2021).
24. A. G. O'Neill, S. G. Lingala, and A. R. Pineda, "Predicting human detection performance in magnetic resonance imaging (MRI) with total variation and wavelet sparsity regularizers," *Proc. SPIE* **12035**, 256–261 (2022).
25. F. Knoll and J. Zbontar, "fastMRI: a publicly available raw k-space and dicom dataset of knee images for accelerated MR image reconstruction using machine learning," *Radiol. Artif. Intell.* **2**(1), e190007 (2020).
26. O. Dietrich et al., "Measurement of signal-to-noise ratios in MR images: influence of multichannel coils, parallel imaging, and reconstruction filters," *J. Magn. Reson. Imaging* **26**(2), 375–385 (2007).
27. K. P. Pruessmann et al., "Sense: sensitivity encoding for fast MRI," *Magn. Reson. Med.* **42**(5), 952–962 (1999).
28. M. Uecker et al., "Berkeley advanced reconstruction toolbox," *Proc. Intl. Soc. Mag. Reson. Med.* **23**, 2486 (2015).
29. B. D. Gallas, G. A. Pennello, and K. J. Myers, "Multireader multicase variance analysis for binary data," *J. Opt. Soc. Am. A* **24**(12), B70–B80 (2007).
30. Z. Yu, M. A. Rahman, and A. K. Jha, "Investigating the limited performance of a deep-learning-based SPECT denoising approach: an observer-study-based characterization," *Proc. SPIE* **12035**, 120350D (2022).
31. K. Li et al., "Assessing the impact of deep neural network-based image denoising on binary signal detection tasks," *IEEE Trans. Med. Imaging* **40**(9), 2295–2305 (2021).
32. B. D. Gallas et al., "A framework for random-effects roc analysis: biases with the bootstrap and other variance estimators," *Commun. Stat. Theor. Methods* **38**(15), 2586–2603 (2009).
33. B. J. Smith and S. L. Hillis, "Multi-reader multi-case analysis of variance software for diagnostic performance comparison of imaging modalities," *Proc. SPIE* **11316**, 113160K (2020).

**Alexandra G. O'Neill** received her BS degree in mathematics and psychology from Manhattan College in 2022. Currently, she is working as a clinical research coordinator in the Division of Neuropsychiatry and Neuromodulation at Massachusetts General Hospital. Her research interests include modeling of human observer performance and statistical analysis in psychology.

**Emely L. Valdez** received her BS degree in mathematics from Manhattan College. Currently, she is working as a mathematics teacher at DreamYard Preparatory High School. Her research interest includes modeling of human observer performance.

**Sajan Goud Lingala** received his bachelor's, master's, and PhD degrees in biomedical engineering respectively from the Osmania University (2002 to 2006), Indian Institute of Technology (2006 to 2008), University of Iowa (2008 to 2014). Currently, he is working as an assistant professor of biomedical engineering and radiology at the University of Iowa. Between 2014 and 2017, he completed a postdoctoral fellowship at the University of Southern California. His research interests include rapid MRI sequence design, dynamic MRI, and model-based reconstruction using adaptive priors (e.g., learning based).

**Angel R. Pineda** completed his BS degree in chemical engineering from Lafayette College, his PhD in applied mathematics from the University of Arizona, and his postdoctoral fellowship in the Radiology Department of Stanford University. Currently, he is working as a professor of mathematics at Manhattan College. His research quantifies the information content of medical images from a task-based perspective.