

# Evaluation of deep learning-based CT reconstruction with a signal-Laplacian model observer

Gregory Ongie<sup>a</sup>, Emil Y. Sidky<sup>b</sup>, Ingrid S. Reiser<sup>b</sup>, and Xiaochuan Pan<sup>b</sup>

<sup>a</sup>Mathematical and Statistical Sciences, Marquette University, Milwaukee, WI, USA

<sup>b</sup>Department of Radiology, University of Chicago, IL, USA

## ABSTRACT

Recent studies have proposed to optimize deep learning-based CT reconstruction methods for signal detectability performance. However, obtaining objective measures of signal detectability performance of the trained reconstruction networks is challenging due to the non-linear nature of the reconstruction. We propose a simple evaluation metric based on the model observer framework. The metric is based on the performance of a specific linear observer on signal-known-exactly/background-known-exactly task. The linear observer uses the signal Laplacian as a template, which we hypothesize is a better proxy for a human model observer than the ideal/Hotelling observer. We illustrate that the proposed metric can be used to select training hyper-parameters for a CNN-model used to reconstruct synthetic sparse-view breast CT data.

**Keywords:** Deep learning, CT reconstruction, image quality assessment, model observers.

## 1. INTRODUCTION

There has been a surge of interest in training convolutional neural networks (CNNs) to reconstruct low-dose/sparse-view CT data. Most current approaches train the CNN by minimizing a pixel-wise mean-squared error (MSE) or similar loss function over a training set of images. However, these losses are insensitive to small and/or low-contrast features that are critical for screening and diagnosis (e.g., tumor spiculations or microcalcifications in breast imaging), and these subtle features can be significantly degraded in the reconstructions.

To address this issue, recent work has proposed modified CNN training procedures inspired by the model observer framework to enhance the detectability of weak signals in the reconstructions.<sup>1-3</sup> The model observer framework, based on signal detection theory, offers an objective means to evaluate how well a reconstruction method preserves fine details in the reconstructions at a statistical level.

A major challenge with these approaches – and most other non-linear CT reconstruction techniques – is how to select various tuning parameters. For example, the approach of Ongie et al.<sup>3</sup> relies on a regularization parameter that trades-off between mean-squared error of the reconstructions and signal detectability performance.

One potential solution is to measure signal detectability performance of the reconstructions in terms of the ideal observer, or a close proxy, such as the (channelized) Hotelling observer, on a signal-known-exactly/background-known-exactly (SKE/BKE) task.<sup>3</sup> However, there are several issues with this methodology. First, finding the ideal observer for a non-linear reconstruction method is challenging. Second, ideal observer performance is known to correlate poorly with human observer performance. Indeed, if the goal is to maximize performance according to the ideal observer, the optimal strategy is to not process the data at all.

Instead, in this study, we propose evaluating the signal detectability performance using a different type of model observer. The proposed observer model uses a linear test statistic using the discrete Laplacian of the signal as the template. We hypothesize that this observer model is a better proxy for human observer performance.

We illustrate the proposed observer model on simulated data in two settings: a simple denoising setting, and reconstruction of sparse-view breast CT data. In both cases, we demonstrate empirically that there is an identifiable peak in detectability performance of the signal-Laplacian observer when varying tuning parameters, unlike the ideal observer. We find this peak correlates well with our own subjective assessment of preservation of fine details in the reconstructions.

---

Further author information: Send correspondence to G. Ongie (email: gregory.ongie@marquette.edu)

## 2. METHODS

The focus of this work is evaluation of learning-based reconstruction models for sparse-view CT reconstruction. First, we briefly describe the CNN training approach proposed in Ongie et al.<sup>3</sup> that is also used in this study. Then we describe the proposed evaluation metric based on an observer model.

### 2.1 CNN Training with Observer Regularization

Let  $f_{\theta} : \mathbb{R}^d \rightarrow \mathbb{R}^d$  denote a CNN depending on parameters  $\theta \in \mathbb{R}^P$  mapping noisy sparse-view FBP images  $\mathbf{y} \in \mathbb{R}^d$  to reconstructed images  $\mathbf{x} \in \mathbb{R}^d$ , which we call the *reconstruction network*. Let  $\{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$  be a collection of training pairs, where each  $\mathbf{y}_i$  is a noisy, sparse-view FBP image and  $\mathbf{x}_i$  is the corresponding ground truth image, and let  $\mathcal{D}$  denote the corresponding empirical distribution of these training pairs. We train the parameters  $\theta$  of the reconstruction network  $f_{\theta}$  by attempting to minimize the loss given by

$$\min_{\theta} \mathbb{E}_{(\mathbf{x}, \mathbf{y}) \sim \mathcal{D}} [\|f_{\theta}(\mathbf{y}) - \mathbf{x}\|^2] + \lambda \cdot \text{ObsReg}(\theta) \quad (1)$$

The “observer regularizer”  $\text{ObsReg}(\theta)$  term is defined with respect to a user-specified distribution of random signals to be planted within the training images. In particular, we assume that pairs  $(\mathbf{s}, \hat{\mathbf{s}})$  can be randomly generated, where  $\mathbf{s}$  is the signal in input space (e.g., its sparse-view FBP) and  $\hat{\mathbf{s}}$  is the same signal represented in output space (e.g., its gridded reconstruction). Then we define  $\text{ObsReg}(\theta)$  as

$$\text{ObsReg}(\theta) = -\mathbb{E}[(f_{\theta}(\mathbf{y} + \mathbf{s}) - f_{\theta}(\mathbf{y}))^{\top} \hat{\mathbf{s}}] \quad (2)$$

where the expectation above is taken with respect to both the noisy sparse-view FBP training images  $\mathbf{y}$  and the random signal pairs  $(\mathbf{s}, \hat{\mathbf{s}})$ . The observer regularizer measures the correlation between the difference of the reconstructions with signal present/signal absent and the true signal. Minimizing this quantity maximizes their positive correlation. Intuitively, this should enhance signal detectability in the reconstructed images.

### 2.2 Evaluation of CNN Reconstruction Methods Using Observer Models

A challenge in deploying the above CNN-based reconstruction scheme is choosing the “best” regularization parameter  $\lambda$  in equation (1). This parameter trades-off between denoising capabilities of the reconstruction network and signal detectability: large values of  $\lambda$  enhance signal detectability performance at the expense of more noise in the reconstructions.

One approach, adopted in Ongie et al.,<sup>3</sup> is to measure the ability of the reconstruction network to preserve small signals on a SKE/BKE task. In this previous study, a channelized Hotelling observer (CHO) is used as a proxy for the ideal observer. For a range of regularization parameter settings, the CHO is estimated and its AUC is estimated empirically. However, it was shown that the AUC as determined by the CHO increased monotonically with regularization parameter  $\lambda$ , plateauing for sufficient large  $\lambda$  where the CNN output reconstructions nearly identical to the input noisy FBP image. Therefore, according to this metric, the “optimal” reconstruction is the noisy FBP image. While this may be optimal from an information-theoretic point of view, we conjecture this is not optimal for human observers. The main contribution of this abstract is to investigate an alternative evaluation metric that we conjecture correlates better with human model observer performance.

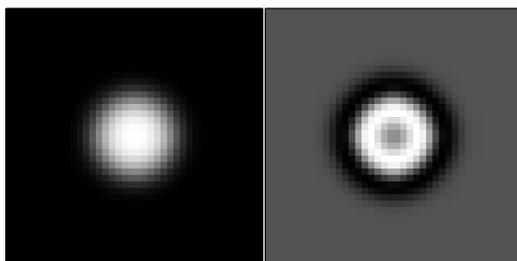
### 2.3 Proposed Observer Model

To measure performance on the SKE/BKE task we propose using a linear observer, i.e., a linear test statistic of the form  $t(\mathbf{y}) = \langle \mathbf{w}, \mathbf{y} \rangle$ , where  $\mathbf{w}$  is a fixed template image. We propose to use the discrete Laplacian of the signal as the template:  $\mathbf{w} = \Delta \mathbf{s}$ , where  $\mathbf{s}$  is the signal as used in the SKE/BKE task, and  $\Delta$  is the discrete Laplacian computed using centered finite differences.

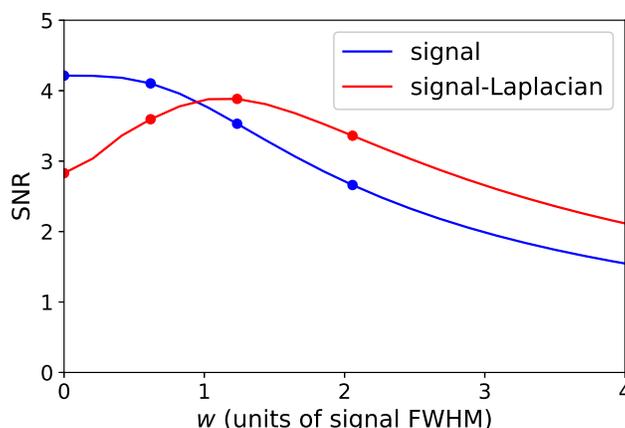
## 3. RESULTS

In order to motivate the use of the signal-Laplacian as a template for the human model observer in a SKE/BKE detection task, we consider a simple imaging system of a signal in white noise. We then apply this observer model to parameter tuning for the CNN-based image reconstruction algorithm.

**Smoothing of image containing a signal in a white noise background** We consider a  $256 \times 256$  pixel noisy image where the noise follows an uncorrelated Gaussian distribution with uniform pixel standard deviation of 2.0. Furthermore, a detection task is considered with a smoothed-disk signal centered in the middle of the image of radius 5.725 pixels and amplitude 1.0 and the image background is zero. The detection task is tantamount to classifying a shown image into either signal-present or signal-absent image hypotheses, and the confounding factor is the image noise. The observer also has the ability to apply Gaussian smoothing to the image with a full-width half-max (FWHM) parameter  $w$  as measured in units of the signal FWHM, 11.45 pixels.



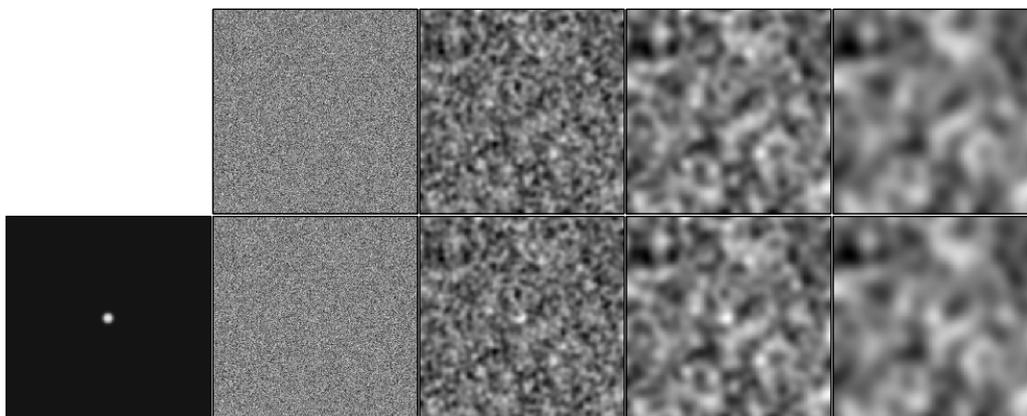
**Figure 1:** Blow-up of candidate templates for human model observer signal detection on a  $20 \times 20$  pixel grid: (Left) the signal itself and (Right) the signal-Laplacian. The considered imaging system is a signal in white noise on a  $256 \times 256$  pixel array, and the detection task is SKE/BKE.



**Figure 2:** SNR for SKE/BKE signal detection as a function of smoothing strength for the signal in white noise imaging system. The SNRs are computed with two possible human model observer templates shown in Fig. 1. For no smoothing  $w = 0$ , the signal template is the ideal observer template, and accordingly the corresponding SNR of 4.213 is the maximum possible SNR. The solid circles indicate values that correspond to the images in Fig. 3.

The ideal observer for this simple imaging system uses a test statistic that involves the dot product between the image and a template that is the signal itself, because the noise distribution is uncorrelated and uniform. Furthermore, the ideal observer would not perform smoothing at all as its SNR for detection is maximal already with  $w = 0$ . It is, however, not clear that this signal template is the one that would model a human observer. We hypothesize that a template that focuses on the edges of the signal may be more representative of a human’s strategy and we formulate this edge-focused model as the dot product of the image with the signal-Laplacian. To illustrate these two strategies, images of the signal and signal-Laplacian templates are shown in Fig. 1. We note that the signal-Laplacian has a “center-surround” structure that has been associated with human observer 2D templates for detection,<sup>4</sup> where a middle region of positive weights is surrounded by a ring of negative weights.

In Figure 2, the SNR is computed for both signal and signal-Laplacian templates and different levels of Gaussian smoothing. The two curves have quite different behavior, with signal and signal-Laplacian SNRs peaking at  $w = 0$  and  $w = 1.23$  signal-widths, respectively. In order to establish correspondence of these results with visualization, noisy image realizations for both signal-present and signal-absent hypotheses are shown in Fig.



**Figure 3:** Noisy signal-absent (top row) and signal-present (bottom row) images for the signal in white noise imaging system. The noiseless signal image is shown in the bottom row of the first column. The subsequent columns show noisy images for different levels of Gaussian smoothing,  $w = 0, 0.62, 1.23,$  and  $2.06$  signal-widths going from top to bottom. For the noisy background the same noise realization is used for all eight panels so that it is easy to observe the difference between the signal-present and signal-absent images. The gray scale for every noisy image is determined by the minimum and maximum pixel values in the corresponding signal-present image.

3 for different levels of smoothing. As this figure is only illustrative, it shows a relatively large signal that is easy to detect and the same noise realization is used on all of the noisy images; it is not intended to be representative of a true two-alternative forced-choice experiment. Starting at the top with  $w = 0$ , it is difficult to distinguish between the signal-present and signal-absent images because the noise amplitude is large compared with the signal. As  $w$  increases the noise amplitude is decreased relative to the signal because it is wider than the speckle structure due to the noise, and it becomes easier to see the signal. In the bottom row of the figure, for  $w = 2.06$  signal-widths, the smoothing significantly degrades the signal amplitude and the signal once again becomes lost in the noise. Thus, the visual trend of Figure 3 supports the SNR trend of the signal-Laplacian template from Figure 2. We note that human-observer experiments would be needed to establish this correspondence quantitatively. For this work, we go ahead and apply the observer model, specified by the dot product with the signal-Laplacian template, to determine parameter settings for the CNN-based image reconstruction algorithm.

### 3.1 Evaluation of CNN's for Sparse-View CT Reconstruction

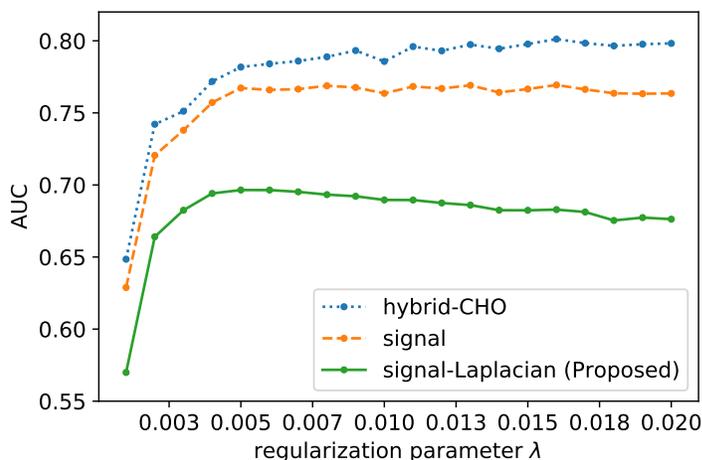
We focus on a sparse-view setting using synthetic breast CT phantoms. For training data, we generate random phantom images using a structured fibro-glandular tissue model. An initial image is generated on a  $2048 \times 2048$  pixel grid, from which we numerically simulate noisy 128-view sinogram data under a 2D circular, fan-beam scanning geometry, which is representative of the mid-plane slice of a 3D circular cone-beam scan. Noise-free ground truth images are formed by downsampling the initial image to a  $512 \times 512$  pixel grid. We also compute an initial FBP reconstruction from the simulated sparse-view sinogram data, which is passed as input to the CNN. We generate 1000 FBP and ground truth image pairs in this way to use for training. We use a U-net architecture<sup>5</sup> for the reconstruction network in all our experiments. We modify the standard U-net slightly by adding a residual “skip” connection with trainable weights.

We set up a SKE/BKE task to measure signal detectability performance on a hold-out test set of images. The test set consists of 1000 signal present realizations and 1000 signal absent realizations, all sharing the same fixed background image. To facilitate computation of the signal detectability metrics, we fix the location of the test signal to the center of the image. The signal strength is set so that the data domain AUC is 0.86.

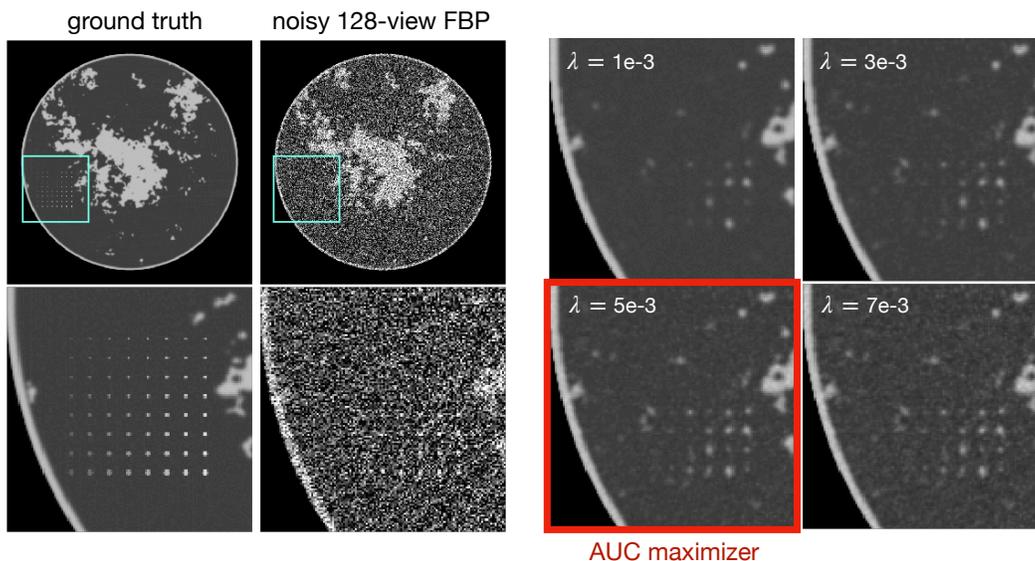
In addition to the signal observer and the proposed signal-Laplacian observer, we also compare against a channelized Hotelling observer that uses a hybrid of pixel and Laguerre-Gauss channels (hybrid-CHO), which have been found effective in estimating signal detectability performance of other nonlinear reconstruction methods.<sup>6</sup> As our figure-of-merit, we compute the area under the ROC curve (AUC) of each observer. This is estimated empirically using the two-alternative forced choice (2-AFC) calculation over the reconstructed test images.

Figure 4 shows the AUCs obtained by the model observers. We observe that AUCs of the signal observer and the hybrid-CHO observer roughly monotonically increase with increasing  $\lambda$ , plateauing at an AUC close to 0.80. The proposed signal-Laplacian observer, though giving much lower AUCs overall, reaches a peak AUC for a small value of the regularization parameter, similar to the denoising experiment above.

In Figure 5 we illustrate the correspondence between visual image quality and signal detectability metrics by reconstructing a test image containing an additional contrast-detail (CD) insert. The CD insert consists of an  $8 \times 8$  grid of point-like signals of varying widths and contrasts. Visually comparing the reconstructions obtained from different networks trained with different  $\lambda$ , the signal-Laplacian AUC maximizer ( $\lambda = 0.005$ ) gives a more faithful reconstruction of the CD insert than lower values of  $\lambda$ , while still suppressing noise.



**Figure 4:** AUC of observer models vs. observer regularization parameter  $\lambda$  used in training a CNN reconstruction network. Observe that the AUC of proposed signal-Laplacian observer reaches a maximum at  $\lambda \approx 0.005$ , while the AUC for the Hybrid-CHO and signal observer increases roughly monotonically with  $\lambda$  increasing.



**Figure 5:** CNN reconstructions of noisy 128-view breast CT phantom with contrast detail insert. A family of CNNs is trained over a range regularization parameters  $\lambda$  in (1). Highlighted in red is the output from the CNN trained with the choice of  $\lambda$  that maximizes the AUC of the signal-Laplacian observer.

## 4. CONCLUSION

We propose a model observer approach to assess signal detectability performance of non-linear CT reconstruction using CNNs. The proposed model observer is based on the signal-Laplacian, which we hypothesize is a reasonable proxy for a human model observer. We demonstrate its potential to aid in selecting hyper-parameters when training a CNN to reconstruct synthetic sparse-view breast CT data.

## REFERENCES

- [1] Wang, W., Gang, G. J., and Stayman IV, J. W., “A CT denoising neural network with image properties parameterization and control,” in [*Medical Imaging 2021: Physics of Medical Imaging*], **11595**, 115950K, International Society for Optics and Photonics (2021).
- [2] Han, M., Shim, H., and Baek, J., “Low-dose CT denoising via convolutional neural network with an observer loss function,” *Medical physics* **48**(10), 5727–5742 (2021).
- [3] Ongie, G., Sidky, E. Y., Reiser, I. S., and Pan, X., “Optimizing model observer performance in learning-based CT reconstruction,” in [*Medical Imaging 2022: Physics of Medical Imaging*], International Society for Optics and Photonics (2022).
- [4] Abbey, C. K., Lago, M. A., and Eckstein, M. P., “Comparative observer effects in 2D and 3D localization tasks,” *J. Med. Imag.* **8**, 041206 (2021).
- [5] Ronneberger, O., Fischer, P., and Brox, T., “U-net: Convolutional networks for biomedical image segmentation,” in [*International Conference on Medical image computing and computer-assisted intervention*], 234–241, Springer (2015).
- [6] Phillips, J. P., Sidky, E. Y., Ongie, G., Zhou, W., Cruz-Bastida, J., Reiser, I. S., Anastasio, M. A., and Pan, X., “A hybrid channelized hotelling observer for estimating the ideal linear observer for total-variation-based image reconstruction,” in [*Medical Imaging 2021: Image Perception, Observer Performance, and Technology Assessment*], **11599**, 115990D, International Society for Optics and Photonics (2021).