

Approximating the Hotelling observer with autoencoder-learned efficient channels for binary signal detection tasks

Jason L. Granstedt,^a Weimin Zhou,^b and Mark A. Anastasio^{a,c,*}

^aUniversity of Illinois Urbana-Champaign, Department of Computer Science, Champaign, Illinois, United States

^bShanghai Jiao Tong University, Global Institute of Future Technology, Shanghai, China

^cUniversity of Illinois Urbana-Champaign, Department of Bioengineering, Champaign, Illinois, United States

ABSTRACT. **Purpose:** The objective assessment of image quality (IQ) has been advocated for the analysis and optimization of medical imaging systems. One method of computing such IQ metrics is through a numerical observer. The Hotelling observer (HO) is the optimal linear observer, but conventional methods for obtaining the HO can become intractable due to large image sizes or insufficient data. Channelized methods are sometimes employed in such circumstances to approximate the HO. The performance of such channelized methods varies, with different methods obtaining superior performance to others depending on the imaging conditions and detection task. A channelized HO method using an AE is presented and implemented across several tasks to characterize its performance.

Approach: The process for training an AE is demonstrated to be equivalent to developing a set of channels for approximating the HO. The efficiency of the learned AE-channels is increased by modifying the conventional AE loss function to incorporate task-relevant information. Multiple binary detection tasks involving lumpy and breast phantom backgrounds across varying dataset sizes are considered to evaluate the performance of the proposed method and compare to current state-of-the-art channelized methods. Additionally, the ability of the channelized methods to generalize to images outside of the training dataset is investigated.

Results: AE-learned channels are demonstrated to have comparable performance with other state-of-the-art channel methods in the detection studies and superior performance in the generalization studies. Incorporating a cleaner estimate of the signal for the detection task is also demonstrated to significantly improve the performance of the proposed method, particularly in datasets with fewer images.

Conclusions: AEs are demonstrated to be capable of learning efficient channels for the HO. The resulting significant increase in detection performance for small dataset sizes when incorporating a signal prior holds promising implications for future assessments of imaging technologies.

© The Authors. Published by SPIE under a Creative Commons Attribution 4.0 International License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JMI.10.5.055501](https://doi.org/10.1117/1.JMI.10.5.055501)]

Keywords: objective image quality assessment; numerical observers; channelized Hotelling observer; autoencoder; neural networks

Paper 23095GR received Apr. 19, 2023; revised Aug. 31, 2023; accepted Sep. 5, 2023; published Sep. 26, 2023.

*Address all correspondence to Mark Anastasio, maa@illinois.edu

1 Introduction

Medical imaging systems are commonly optimized with consideration of a specific task.¹ Assessing the performance of such systems requires an objective metric for image quality.²⁻⁶ For signal detection tasks, the Bayesian ideal observer (IO) has been advocated for producing a figure-of-merit for assessing imaging systems because it can maximize the amount of task-specific information in the measurement data.²⁻⁷ For a binary signal detection task, the IO test statistic takes the form of a likelihood ratio. Using this likelihood ratio as a test statistic in turn maximizes the area under the receiver operating characteristics (ROC) curve.^{2-4,7} However, analytically determining the IO is generally difficult because it typically is a nonlinear function and requires complete knowledge of the statistical properties of the image data.

There has been recent progress in developing approximations for computing the IO test statistic.^{5,6} One line of research involves sampling-based methods that utilize Markov-chain Monte Carlo methods to approximate the IO, but the work in this area has largely been limited to relatively simple stochastic object models.^{3,5,8,9} Another development is the approximation of the IO with convolutional neural networks (CNNs).¹⁰⁻¹² Generative adversarial models have also been employed to approximate the IO with Markov-Chain Monte Carlo methods.^{13,14} An alternative method to approximating the IO's performance employs variational Bayesian inference.¹⁵ This line of research has shown promise for implementing task-specific optimization of sparse reconstruction methods.

A common surrogate for the often-intractable IO is the Hotelling observer (HO).¹⁶⁻¹⁹ The HO is the optimal linear discriminator for maximizing the signal-to-noise ratio (SNR) of the test statistic.^{20,21} Directly implementing the HO requires the estimation and inversion of a covariance matrix, which quickly grows as image size increases and can become intractable to compute.²² There are a few different strategies for mitigating the computational cost of inverting a large matrix.^{16,22} One method is to avoid a direct inversion by implementing an iterative approach to estimate the test statistic.² If the measurement noise covariance matrix is known and an estimate of the background covariance matrix is available, covariance matrix decomposition is another option,² with the caveat that certain situations can lead to significant bias in the performance.²³ Alternatively, the test statistic can be learned directly from the images provided that there is a sufficient amount of data.^{11,24} The most commonly employed method, however, is the implementation of channels that approximate the HO.^{5,25,26} Additionally, a recent comparison between nonlinear CNN-based methods and linear channelized methods for simple signals on simulated digital mammography (DM) images found that channelized methods can outperform CNN-based methods when training data are limited.²⁷ Conceptually, channels function by projecting the high-dimensional image data to a low-dimensional image manifold.²⁸⁻³⁰ Ideally, the manifold embedding would preserve the important features of the data.³¹ This dimensionality reduction stabilizes the process of estimating the HO when considering noisy data and limited amounts of training images.

Channelized observers are considered efficient if they closely approximate the original observer's performance.³² Prior work in computing efficient channels includes Laguerre-Gauss (LG),²⁶ singular value decomposition (SVD),³³ the filtered channel observer (FCO),³⁴ and partial least squares (PLS).³⁵ Although some methods generate generic channels that can be applied to classes of signals,²⁶ the more recent work has focused on channels that are applicable to a specific detection task.^{35,36} These channels require a dataset of images for training but are applicable to detection tasks where the signal is not known exactly or rotationally symmetric.^{35,36} In addition to learning efficient channels, there are approaches that seek to mimic the human observer's performance.³⁷⁻³⁹ An early approach learned the relationship between channel features and human observer performance with a support vector machine.³⁷ Another approach investigated optimization with respect to the HO and human observers on accelerated MRI reconstruction.³⁸ There has also been some recent work applying CNNs for directly learning a surrogate measure of human observer performance.³⁹ In the remainder of this work, however, the focus is on efficient channels.

Autoencoders (AEs) are a type of artificial neural network (ANN) that are characterized by a mirror structure, with the target output of the network similar to the input.⁴⁰⁻⁴⁴ They are designed to learn a lower-dimensional representation of the data called an embedding. The portion of the network that transforms the input to the embedding is known as the encoder, and the portion that

transforms the embedding back into the original data space is known as the decoder. A good embedding is capable of significant data compression while retaining most of the information from the original data. The data compression qualities of AEs make them desirable to use in many tasks, and they have been applied in state-of-the-art systems for classification,⁴⁵ noise reduction,⁴⁶ and regression.⁴⁷ The widespread success of the AE is due to its ability to generate low-dimensional representations of images, which increases the efficiency of further processing by attenuating noise and embedding data to its most important components. Linear AEs share some similarities with principle component analysis. In general, a linear AE with optimal weights projects the data onto a subspace spanned by the top principle directions of the data.⁴⁸

In this work, the problem of learning task-informed embeddings with an AE is explored to develop efficient channelized observer methods. AEs have previously been demonstrated to develop efficient channels for simple signal-known-exactly/background-known-statistically (SKE/BKS) detection tasks.^{36,49} This paper expands on those works by demonstrating that the learning task for an AE is equivalent to learning channels for the HO. The AE is modified to learn the optimum transformation matrix that maximizes the amount of task-specific information encoded in its latent states. Experiments are performed to validate AE-learned channels on location-known binary detection tasks, including SKS tasks with an elliptical signal on a lumpy background and detection tasks with two separate signals on a breast phantom background. The performance of the AE-learned channels in these studies is compared to state-of-the-art channelized methods, including PLS and FCO, as well as conventional methods for estimating the HO. A study is also performed to assess the ability of these channelized methods to generalize to related imaging systems. Finally, the impact of the quality of the signal estimate on channelized methods is explored.

The remainder of this work is organized as follows. In Sec. 2, the salient aspects of binary signal detection theory are presented. The HO, channelized HO (CHO), and AE are also reviewed in that section. A methodology for learning channels for a location-known binary signal detection task using an AE is developed in Sec. 3. The numerical studies and results of the proposed method for approximating the HO are included in Secs. 4 and 5, along with a comparison to other state-of-the-art methods. Finally, this paper concludes with a discussion of the work in Sec. 6.

2 Background

We consider a linear digital imaging system described as

$$\mathbf{g} = \mathcal{H}f(\mathbf{r}) + \mathbf{n}, \quad (1)$$

where $f(\mathbf{r})$ represents the object function with the spatial coordinate $\mathbf{r} \in \mathbb{R}^K$, \mathcal{H} represents a continuous-to-discrete (C-D) imaging operator that maps $\mathbb{L}_2(\mathbb{R}^K) \rightarrow \mathbb{R}^N$, $\mathbf{n} \in \mathbb{R}^N$ is the random measurement noise, and $\mathbf{g} \in \mathbb{R}^N$ denotes the measured image data vector. Hereafter, the object function $f(\mathbf{r})$ will be abbreviated as \mathbf{f} .

2.1 Overview of Binary Signal Detection Tasks

The binary signal detection task considered involves the classification of an image by an observer into one of two hypotheses: signal-present (H_1) or signal-absent (H_0). Under these hypotheses, the imaging process can be represented as

$$H_0: \mathbf{g} = \mathcal{H}\mathbf{f}_b + \mathbf{n}, \quad (2a)$$

$$H_1: \mathbf{g} = \mathcal{H}(\mathbf{f}_b + \mathbf{f}_s) + \mathbf{n}, \quad (2b)$$

where \mathbf{f}_b and \mathbf{f}_s denote a background and a signal object, respectively. Depending on the imaging task, these can be either random or fixed.

In the binary signal detection task, a real-valued test statistic $t(\mathbf{g})$ is obtained from applying an observer to a measured image. This test statistic is compared against a threshold τ to classify \mathbf{g} as satisfying either H_0 or H_1 . To determine the desired performance for the signal detection task, an ROC curve can be plotted to depict the trade-off between the false-positive fraction and the true-positive fraction by varying the threshold τ . The overall signal detection performance of the observer can be summarized by computing the area under the ROC curve (AUC).⁵⁰

2.2 Bayesian Ideal Observer and Hotelling Observer

The IO is optimal and sets the upper limit for observer performance on signal detection tasks. The IO test statistic for binary tasks is defined as a monotonic transformation of the likelihood ratio $\Lambda_{LR}(\mathbf{g})$ that takes the form^{2,3,12}

$$\Lambda_{LR}(\mathbf{g}) = \frac{p(\mathbf{g}|H_1)}{p(\mathbf{g}|H_0)}, \quad (3)$$

where $p(\mathbf{g}|H_0)$ and $p(\mathbf{g}|H_1)$ are conditional probability density functions that describe \mathbf{g} under the hypotheses of H_0 and H_1 , respectively.

An alternative to the IO for assessing signal detection performance is the HO. The HO test statistic is defined as

$$t(\mathbf{g}) = \mathbf{w}_{HO}^T \mathbf{g}, \quad (4)$$

where $\mathbf{w}_{HO} \in \mathbb{R}^M$ denotes the observer template. Let $\bar{\mathbf{g}}_j \equiv \langle \langle \mathbf{g} \rangle_{\mathbf{g}|\mathbf{f}} \rangle_{\mathbf{f}|H_j}$ denote the conditional mean averaged with respect to object randomness and noise associated with H_j . The Hotelling template \mathbf{w}_{HO} is defined as²

$$\mathbf{w}_{HO} = \left[\frac{1}{2} (\mathbf{K}_0 + \mathbf{K}_1) \right]^{-1} \Delta \bar{\mathbf{g}}, \quad (5)$$

where

$$\mathbf{K}_j = \langle \langle [\mathbf{g} - \bar{\mathbf{g}}_j][\mathbf{g} - \bar{\mathbf{g}}_j]^T \rangle_{\mathbf{g}|\mathbf{f}} \rangle_{\mathbf{f}|H_j}, \quad (6)$$

$$\Delta \bar{\mathbf{g}} = \bar{\mathbf{g}}_1 - \bar{\mathbf{g}}_0, \quad (7)$$

where \mathbf{K}_j and $\bar{\mathbf{g}}_j$ represent the covariance matrix and mean of the measured data \mathbf{g} under hypothesis H_j , respectively. The mean signal $\Delta \bar{\mathbf{g}}$ is the difference of the means under the two hypotheses. A regularized Moore–Penrose (MP) pseudoinverse can be applied instead of the inverse in Eq. (5) in cases where the covariance matrix is ill-conditioned.²

The SNR associated with the test statistic t is another commonly employed FOM for assessing signal detection performance and is given by²¹

$$\text{SNR}_t = \frac{\langle t \rangle_1 - \langle t \rangle_0}{\sqrt{\frac{1}{2}\sigma_0^2 + \frac{1}{2}\sigma_1^2}}, \quad (8)$$

where $\langle t \rangle_j$ and $\sigma_j^2 = \langle (t - \langle t \rangle_j)^2 \rangle_j$ are the mean and variance of $t(\mathbf{g})$ under hypothesis H_j . The HO maximizes the SNR of its test statistic $t(\mathbf{g})$, but unlike the IO it does not necessarily maximize the AUC.²¹

2.3 Channels

In some cases, the direct estimation of the HO can become intractable. The covariance matrix in Eq. (5) may not be invertible due to the size of the images or insufficient quantity of data. As images become larger, the amount of data required to obtain a robust estimate noticeably increases. Medical image datasets frequently do not contain sufficient images to robustly estimate \mathbf{K}_j due to large image sizes or the limited access to large scale annotated datasets.

To mitigate this problem, a channelized version of the image \mathbf{g} can be introduced as²¹

$$\mathbf{v} = \mathbf{T}\mathbf{g} = \mathbf{T}(\mathcal{H}\mathbf{f} + \mathbf{n}), \quad (9)$$

where \mathbf{v} is a $M \times 1$ channel-reduced image and \mathbf{T} is a $M \times N$ matrix. The dimensionality reduction is determined by the number of channels $M < N$. Applying the HO to the channel-reduced data yields the CHO,²¹ with the test statistic taking the form

$$T(\mathbf{v}) = \mathbf{w}_v^T \mathbf{v} = (\mathbf{K}_v^{-1} \Delta \bar{\mathbf{v}})^T \mathbf{v}, \quad (10)$$

where $\mathbf{K}_v = \frac{1}{2} [\mathbf{K}_{v,1} + \mathbf{K}_{v,0}]$ and $\Delta \bar{\mathbf{v}} = \bar{\mathbf{v}}_1 - \bar{\mathbf{v}}_0$, where $\mathbf{K}_{v,j} = \langle \langle [\mathbf{v} - \bar{\mathbf{v}}_j][\mathbf{v} - \bar{\mathbf{v}}_j]^T \rangle_{\mathbf{v}|\mathbf{f}} \rangle_{\mathbf{f}|H_j}$ and $\bar{\mathbf{v}}_j = \langle \langle \mathbf{v} \rangle_{\mathbf{v}|\mathbf{f}} \rangle_{\mathbf{f}|H_j}$ for $j = (0,1)$.

Efficient channels should maximize the retained task-relevant information to provide an efficient approximation of the HO. There are several methods that exist for selecting efficient channels. One of the first was LG channels.³² These channels are a combination of a Gaussian function with a Laguerre polynomial and were proposed due to their structural similarity with the Hotelling template for certain detection tasks. These channels are suitable for a smooth rotationally symmetric signal on a lumpy background but may have suboptimal performance for arbitrary signals and more complex backgrounds.³⁵

An alternative to LG channels are SVD channels.³³ These channels are singular vectors that form a basis for image vectors in the range of the imaging operator. The most efficient set of channels constructed from this method involved decomposing the noiseless signal image by use of the singular vectors and choosing the top m of them to form the channel set. However, this method is computationally expensive and system-specific.

Two current state-of-the-art methods for generating efficient channels capable of considering arbitrary signals and backgrounds without any specific knowledge of the imaging system are PLS³⁵ and the FCO.³⁴ The PLS method applies a data reduction technique that iteratively constructs a number of latent vectors that maximize the covariance between the data and a binary label representing the presence of a signal. The PLS method represents an attractive method to use in limited-data cases and/or large image sizes and works well with noisy and heavily correlated data. However, the technique can suffer from a degradation of performance when the amount of available image data is small.³⁵

FCO channels were initially developed as anthropomorphic channels to approximate human signal detection performance for irregularly shaped signals.³⁴ However, FCO channels have also been explored as efficient channels for the HO.^{34,51,52} The FCO convolves a selected set of baseline channels with the signal before computing the observer template. For the experiments in this paper, LG channels were selected as the baseline set of channels due to both LG's past success³² and similar decisions with the FCO method in more recent work.^{51,52} This realization of the FCO method will be referred to below as a convolutional LG observer.

2.4 Neural Networks for Approximating the IO

A feed-forward ANN is a system of computational units associated with tunable parameters called weights.^{53,54} A feed-forward ANN is capable of approximating any continuous function if it has a sufficiently complex architecture.^{55,56} ANNs have been employed to form numerical observers, with the focus on directly estimating the test statistic.^{10–12} Kupinski et al.¹² utilized conventional fully connected neural networks to approximate the IO on low-dimensional extracted image features. Zhou and Anastasio extended this work to higher-dimensional data and allowed for native processing of image data by replacing the FCNN with a CNN.^{10,11} However, these approaches focus on learning the test statistic directly and may require a complex network and a large amount of training data to accurately approximate the IO.

2.5 Autoencoders

A specialized type of ANN is the AE.^{40–44} The AE is characterized by a mirror structure, with the input of the network similar to the target output. An AE has three distinct components: an encoder, an embedding, and a decoder. The encoder transforms the input to the embedding, which generally has a significantly reduced dimensionality compared to the input. The decoder transforms the embedding into the target output. In a canonical AE, the decoder is specified to reconstruct an approximation of the input to the encoder. AEs are frequently employed for their data compression properties in state-of-the-art systems for classification,⁴⁵ regression,⁴⁷ noise reduction,⁴⁶ anomaly detection,⁵⁷ and image recovery⁵⁸ tasks. Additional performance improvements can be made by injecting additional information into the AE training process. Studies have shown that exploiting *a priori* information via modifications to the loss function can introduce task-specific information in the training of AEs.^{59,60} In contrast to previous work with ANNs, an AE is usually trained in an unsupervised way.⁴⁵

In general, the layers in an AE specify many sets of matrix multiplications with added bias terms and nonlinear transformations. By restricting the operations to only matrix multiplications, linear AEs can be obtained. In these cases, the encoder and decoder can each be described by a transformation matrix that transforms to or from the data embedding. Such a simplified network

is considered in this work since this configuration's encoder has a natural parallel with the channel matrix in the CHO. The input to the network is an image, and the target output is either the input image or a related version of the input image, depending on the task.

Another aspect of AEs that has recently been considered is the concept of tied weights.⁶¹ Tied weights further enforce the mirror-like structure of the AE by forcing the encoder and decoder matrices to be symmetric. Tied-weight AEs have been shown to perform similar to untied-weight AEs but require less data to train because of the reduction in parameters.

An optimization problem is solved to determine the weights of the AE by minimizing a reconstruction loss. Let \mathbf{W}_1 and \mathbf{W}_2 be $N \times M$ weight matrices that parameterize the encoder and decoder of the linear AE, respectively. The optimization problem to obtain the optimal weights of a linear AE is

$$\hat{\mathbf{W}}_1, \hat{\mathbf{W}}_2 = \arg \min_{\mathbf{W}_1, \mathbf{W}_2} \|\mathbf{W}_2 \mathbf{W}_1^T \mathbf{g} - \mathbf{g}^*\|^2, \quad (11)$$

where the target reconstruction is represented by \mathbf{g}^* . This target reconstruction can be the same as or different from the input data but is usually closely related. For example, in denoising problems, the target output is a clean version of the input image.

The solution of the optimization problem is computed by minimizing the loss function using a variation of the backpropagation algorithm.⁶² The conventional loss function for an AE is the mean squared error between the input and the output of the network. Given P vectorized background images \mathbf{g}_i , the conventional loss function corresponding to a zero-bias linear AE is⁴⁰

$$L_{\text{con}}(\mathbf{W}_1, \mathbf{W}_2) = \frac{1}{P} \sum_{i=1}^P \|\mathbf{W}_2 \mathbf{W}_1^T \mathbf{g}_i - \mathbf{g}_i^*\|_2^2. \quad (12)$$

The data embedding of the trained AE is represented as

$$\hat{\mathbf{y}} = \hat{\mathbf{W}}_1^T \mathbf{g}, \quad (13)$$

and the corresponding reconstruction is $\hat{\mathbf{g}} = \hat{\mathbf{W}}_2 \hat{\mathbf{y}}$.

3 Method: Autoencoder-Learned Channels

A method for learning efficient channels for the CHO with an AE is described below. A connection between AE weights and the CHO framework is established to illustrate the connection between the learned data embeddings and more traditional channels.

3.1 Autoencoder Channels and Linear Autoencoders

The learned weights of an AE have an additional interpretation when considered in the framework of a signal detection task. The weights define a mapping from the high-dimensional image space to a low-dimensional embedding space. This is conceptually equivalent to the CHO channel matrix \mathbf{T} . Thus the learned AE weights can be employed as channels for the CHO by setting $\mathbf{T} = \mathbf{W}_1^T$ in Eq. (9). Intuitively, these AE-learned channels capture the data most important for reconstructing the image. Combining a conventional AE with directly learning an observer on the data embedding has also been proposed.⁴⁹ However, here AE channels with a conventional method for obtaining the CHO will be considered.

The embedding in Eq. (13) causes the AE to encode the entirety of the input image. This makes the conventional AE suboptimal for learning channels because a significant portion of the data embedding is dedicated to reconstructing certain components of the image that may not be highly relevant to the detection task. To circumvent this, as described below, information about the signal can be incorporated into the AE training process to preserve task-specific information.

3.2 Task-Specific Autoencoders

A modification to the loss function to improve the learned data embedding and resulting signal detection performance for AE-channels is presented here. Ideally, the entirety of the AE embedding would be dedicated to the task-specific information. This would minimize the proportion of the embedding that is dedicated to extraneous information and lead to a more efficient set of

channels. By changing the AE's target reconstruction to the mean signal image for a binary signal detection task, the background and noise are suppressed during the reconstruction process. This results in an embedding in which the signal can be accurately represented. This substitution of the target reconstruction minimizes the MSE between the reconstructed image and the estimated signal image and takes the form of

$$L_{\text{task}}(\mathbf{W}_1, \mathbf{W}_2) = \frac{1}{P} \sum_{i=1}^P \|\mathbf{W}_2 \mathbf{W}_1^T \mathbf{g}_i - I(\mathbf{g}_i) \Delta \bar{\mathbf{g}}\|_2^2, \quad (14)$$

where $\Delta \bar{\mathbf{g}}$ is defined in Eq. (7) and $I(\cdot)$ is the indicator function that returns 1 if the signal is present and 0 otherwise. Since the loss function employs label information, this approach is a supervised learning method. Considering the background as noise to be suppressed permits the entire capacity of the embedding to focus on the task-specific information. Although a linear AE projects data into a subspace spanned by the data's top principle directions,⁴⁸ the proposed modification removes this association. Using the signal template as the target image assists the training process in identifying an embedding that preserves task-specific information. As shown below, this modification to the loss function is capable of generating efficient channels for the CHO. A diagram of the AE with both the conventional and task-based approach for the signal detection task is provided in Fig. 1, with a sample reconstruction from AEs trained using both loss functions shown in Fig. 2. Both the task-specific and conventional loss functions can be minimized by use of a gradient-descent method, with specific implementation details provided in Sec. 4.6.

4 Numerical Studies

Numerical simulation studies were conducted to evaluate the performance of the proposed method for learning efficient channels for the CHO. All simulations addressed BKS signal detection tasks. Four distinct binary signal detection tasks were considered. Using a lumpy stochastic object model, a location-known task and a fixed-centroid signal-known-statistically task were considered. These tasks enabled the HO to be estimated both using covariance matrix decomposition² and direct computation according to Eq. (5). These observers will be referred to

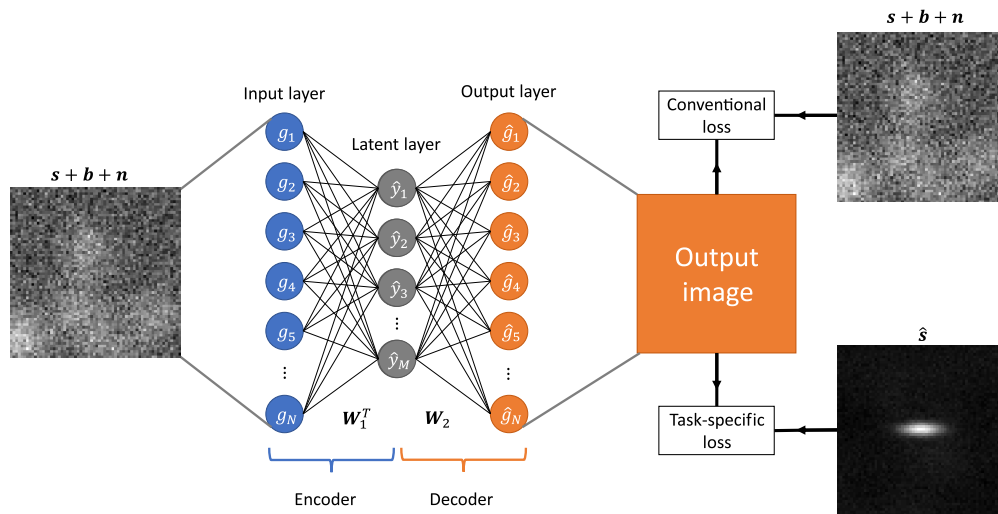


Fig. 1 Diagram of the proposed method. A noisy image is the input to an encoder, which is mapped to an M -dimensional latent space to encode the information. The encoder transformation matrix is given by \mathbf{W}_1^T . The embedded representation is then multiplied by the decoder transformation matrix \mathbf{W}_2 to return to image space and generate the output image. Two different loss functions are considered for training this model. The conventional loss function computes the MSE between the output image and the input image. This approach attempts to reconstruct the entire input image. The second considered loss function is the task-specific loss, which calculates the difference between the output image and estimated signal image $\hat{\mathbf{s}}$. This loss maximizes the signal-specific information of the input image.

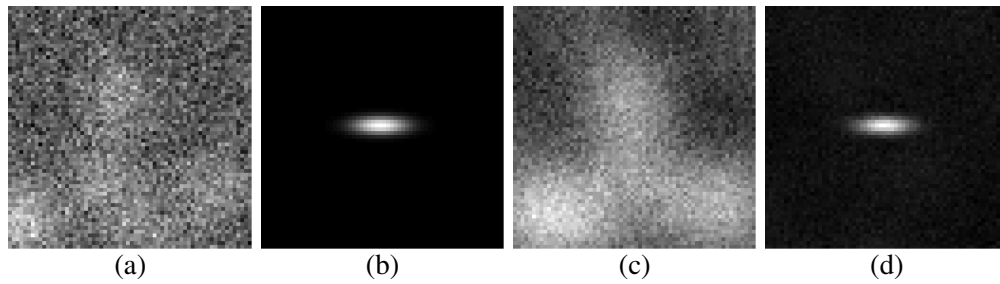


Fig. 2 Reconstructed images corresponding to each of the loss functions. The grayscale in each case is adjusted to maximize visibility. Panel (a) is the input image, which contains the faint Gaussian elliptical signal in (b). The conventional AE with 20 channels reconstructs the image in (c) while the task-specific tied-weight AE with 10 channels reconstructs the image in (d). Note that the reconstructed image in (c) is noticeably less noisy than the input image (a), which it is attempting to reconstruct. This is due to the limited number of latent states in the model embedding the largest structures in the input images. Noise cannot be effectively encoded for an image, so it is attenuated.

as HO-CMD and HO-Direct, respectively. On a breast phantom background, two location-known signal detection tasks using signals of different shapes and sizes were considered. These tasks allowed for the evaluation of channelized methods on a more realistic medical imaging task. ROC curves were fit by use of a binormal model^{50,63,64} with the fitted AUC values reported. The experimental results are reported in distinct sections based on the image background model, with the details for each signal detection task and the training of neural networks are given in the appropriate sections.

4.1 Signal Detection Tasks That Utilize a Lumpy Background Model

Two different signal detection tasks were performed with consideration of a lumpy stochastic object model⁶⁵ with an idealized parallel-hole collimator system.³ Further details about each of the components are provided below.

4.1.1 Lumpy background

A stochastic lumpy object model was used as the background⁶⁵

$$\mathbf{f}_b(\mathbf{r}) = \sum_{n=1}^L l(\mathbf{r} - \mathbf{r}_n | a, s), \quad (15)$$

where $L \sim \text{Pois}(\bar{L} = 5)$ is the number of lumps that is sampled from Poisson distribution with the mean set to 5 and $l(\mathbf{r} - \mathbf{r}_n | a, s)$ is the lump function modeled by a symmetric 2D Gaussian function with amplitude a and width s given by

$$l(\mathbf{r} - \mathbf{r}_n | a, s) = a \exp\left(-\frac{(\mathbf{r} - \mathbf{r}_n)^T(\mathbf{r} - \mathbf{r}_n)}{2s^2}\right), \quad (16)$$

where \mathbf{r}_n is the uniformly sampled position of the n 'th lump. The magnitude and width of the lumps were set to the frequently employed values of $a = 1$ and $s = 7$.^{10,35,49,65} An example of a signal-present image in the dataset with a circular signal is located in Fig. 3.

4.1.2 Imaging system

The stylized imaging system in these studies was a linear C-D mapping describing an idealized parallel-hole collimator system with a point response function given by^{3,66}

$$\mathbf{h}_m(\mathbf{r}) = \frac{h}{2\pi w^2} \exp\left(-\frac{(\mathbf{r} - \mathbf{r}_m)^T(\mathbf{r} - \mathbf{r}_m)}{2w^2}\right), \quad (17)$$

with the height $h = 40$, the width $w = 0.5$, and the location of the m 'th pixel in \mathbf{r} denoted by m .

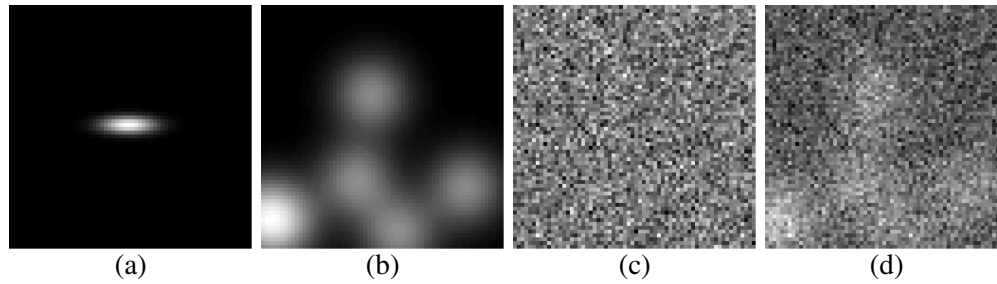


Fig. 3 Sample generation for signal-present images used in the lumpy background experiments. The grayscale in each case was adjusted to maximize visibility. The signal image (a) was added to the lumpy background (b) and the Gaussian noise (c) to produce the composite dataset image (d). The signal image is an elliptical Gaussian signal with width $\sigma_x = 5$, $\sigma_y = 1.5$.

4.1.3 Signals

The signal function \mathbf{f}_s was a 2D Gaussian function:

$$\mathbf{f}_s(\mathbf{r}) = A \exp(-(\mathbf{R}_\theta(\mathbf{r} - \mathbf{r}_c))^T \mathbf{D}^{-1}(\mathbf{R}_\theta(\mathbf{r} - \mathbf{r}_c))), \quad (18)$$

where $A = 0.2$ is the amplitude and \mathbf{r}_c is the coordinate of the signal location. Here \mathbf{R}_θ is the Euclidean rotation matrix that rotates the Gaussian by an angle of θ and is given by

$$\mathbf{R}_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}, \quad (19)$$

and \mathbf{D} is a scaling matrix that controls the width of the Gaussian along each axis and is given by

$$\mathbf{D} = \begin{bmatrix} 2\sigma_x^2 & 0 \\ 0 & 2\sigma_y^2 \end{bmatrix}. \quad (20)$$

For both experiments involving the lumpy background, the elliptical Gaussian signal was set to have the parameters $\sigma_x = 5$ and $\sigma_y = 1.5$. The image size was selected to be 64×64 with the signal centered at $\mathbf{r}_c = [32, 32]^T$. The value of θ varied depending on the type of task.

4.1.4 Detection tasks

The first signal detection task considered a single orientation of the signal by setting $\theta = 0$. The signal template was computed according to Eq. (7), which resulted in a noisy estimate of the signal. The second signal detection task sampled θ uniformly from the set $[0 \text{ deg}, 45 \text{ deg}, 90 \text{ deg}, 135 \text{ deg}]$. This allowed for four distinct orientations of the elliptical Gaussian. The mean signal was also computed with Eq. (7), which resulted in a noisy estimate of the signal averaged across the four possible realizations.

4.1.5 Dataset generation

A training set of 60,000 unique background images with noise were generated for the lumpy object model. The background images were generated separately from the signal image in Eq. (18) using the appropriate background model. Each background image was summed with a unique noise vector drawn from an independent and identically distributed Gaussian distribution $\mathbf{n}_m \sim \mathcal{N}(0, \delta^2)$ with a mean of 0 and standard deviation $\delta = 20$. These images were then paired, with half designated for signal present and half for signal absent. Each signal present image was summed with the signal image to generate the final training dataset of 30,000 paired images. Another set of 5000 paired images was generated for determining the channel covariance matrix after the channels had been learned and a further set of 5000 paired images were held out as a testing dataset.

4.2 Location-Known Tasks That Utilize a Breast Phantom Dataset

Two additional signal detection tasks were performed with consideration of a breast phantom stochastic object model employing the VICTRE dataset.^{51,52} This dataset contains simulated DM

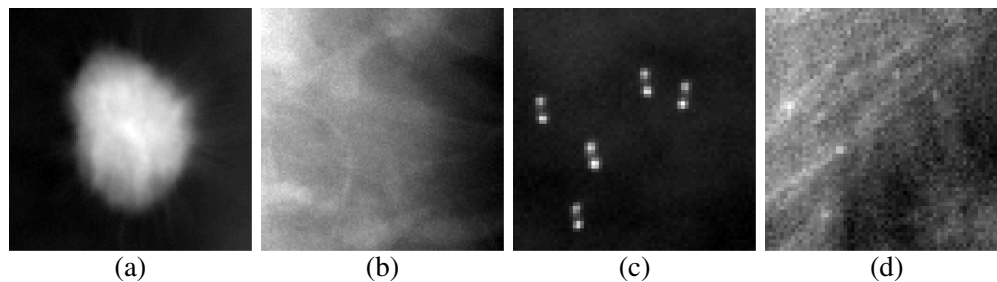


Fig. 4 Sample estimated signals and images from the VICTRE breast phantom dataset. The gray-scale was adjusted in each case to maximize visibility. Panels (a) and (c) contain the mean signal image of the spiculated mass and microcalcification cluster, respectively. Panels (b) and (d) are sample images of those corresponding signals embedded into a fatty breast phantom. Note that the signal in (c) appears to be doubled. This artifact is a result of the extracted scattered breast ROIs not aligning with the other breast types in the provided public dataset.

images and was employed previously in a location-known mathematical observer study to evaluate imaging systems.^{51,52} The images were divided into four categories of breast types: extremely dense, heterogeneously dense, scattered fibroglandular, and fatty. The signals in the dataset were microcalcification clusters and spiculated masses. Regions of interest (ROIs) from the simulated DM image were extracted to generate the images used in the signal detection task. For each signal, there were associated signal-absent and signal-present images. The spiculated mass ROIs were 109×109 pixels and the microcalcification ROIs were 65×65 pixels. The considered microcalcification signal was a consistent arrangement of five microcalcifications, which were considered jointly as one signal for the detection task. The signal remained constant throughout all the signal-present images, but a clean signal image was assumed to be unavailable. An estimation of the signal was obtained from the difference of the mean signal-present and signal-absent images according to Eq. (7), making this a location-known task.^{51,52}

For each type of signal, 12,500 total images were selected from the dataset to form training, validation, and testing sets of 5000, 625, and 625 paired images, respectively. The breast types selected maintained the proportions employed in the VICTRE study.^{51,52} The signals were estimated by taking the mean of the signal-present images and subtracting the mean of the signal-absent images for the combined training and validation dataset. Sample images and estimated signals are included in Fig. 4.

4.3 Location Known Tasks Considering a Provided Signal

The tasks in Secs. 4.1 and 4.2 consider a mean signal that is computed separately for each dataset size. This represents the case where the training data are used to estimate the mean signal. However, one can also consider the performance of methods when the mean signal is provided. This represents cases where a prior about the signal is known and can be leveraged to improve detection performance. Such experiments isolate the contribution of the mean signal estimate from additional images provided by larger datasets. This experiment formulation has also been employed previously to evaluate the dependence of channelized methods on the quantity of background images.³⁵ The experiments in Secs. 4.1 and 4.2 were repeated, setting the mean signal to be the estimate from the training and validation data at the maximum dataset size for cases.

4.4 Tasks Evaluating the Generalizability of Channelized Methods

Since the proposed method of determining channels is dependent on the mean signal image, there is the reasonable concern that AE-learned channels may fail to generalize outside of the learned dataset. Thus two studies were considered to evaluate the ability of the learned channelized methods to generalize.

4.4.1 Domain shift study

The first case considers the effects of domain shift on the channels. Domain shift is the difference in distributions between the datasets an algorithm is trained and evaluated upon.⁶⁷ A method's

robustness to domain shift corresponds to generalization performance when a set of channels developed for one imaging system is employed in another related system with different properties. Such a circumstance may occur when using channels to tune an imaging system's parameters, as it is undesirable to recompute channels whenever part of the system is changed. Characterization of such performance is important when considering the application of channels developed with consideration of one imaging system to data produced by another imaging system. Medical imaging datasets are frequently difficult to obtain, and it may be beneficial to leverage previously collected datasets to develop channels for use with new imaging systems.

An experiment was developed to investigate the impact of domain-shift on the learned channels. To evaluate observer performance in the presence of domain shift, a source dataset and a target dataset are necessary. The source dataset is defined as the training dataset for a channelized method. Likewise, the target dataset is the dataset upon which the channelized method is evaluated. The difficulty of the problem depends on the difference in the distributions between the source and target dataset.

To evaluate generalization performance across multiple instances of domain shift, three datasets were constructed and all possible combinations of source and target datasets were considered. The imaging system, location-known elliptical signal, and stochastic object model for the background were reused from Sec. 4.1.1. These parameters correspond with a location-known detection task for a stylized PET imaging system. The three datasets were constructed by setting the width of the imaging system in Eq. (17) to 1.0, 2.0, and 4.0. Thus the domain shift is the difference in image distributions from stylized PET imaging systems with different widths.

The noise model was also altered from the previous studies. A mixed Poisson/Gaussian noise model was considered, as it is known to be a reasonable approximation of PET noise.⁶⁸ Each pixel of the image was sampled as a random Poisson variable with the mean set to the intensity of the voxel and combined with additive i.i.d. Gaussian noise. This noise model takes the form of

$$\mathbf{n}_m = \mathbf{p}_m + \sigma_m, \quad (21)$$

where $\mathbf{p}_m \sim \text{Poiss}(\mathbf{g}_m)$ and $\sigma_m \sim \mathcal{N}(0, 1)$. After adding the noise, the images were scaled to $[0, 1]$.

The generated datasets had training, validation, and testing sizes of 30,000, 5000, and 5000 paired signal present/absent images, respectively. The same set of backgrounds were considered for each dataset.

4.4.2 Amalgamated dataset study

Due to the difficulty of obtaining medical imaging data and the benefits of additional data for improving performance, it is sometimes beneficial to consider related images acquired from multiple imaging conditions. For this combined dataset to be beneficial for application across multiple imaging conditions, the imaging conditions must be similar enough such that the benefits of the additional data outweigh the loss of imaging-condition specific channels. The performance of the obtained channels upon the individual imaging conditions that contributed to the combined dataset is relevant for evaluating the benefits of such a technique.

To evaluate such a case, pairs of images were evenly sampled from the three constructed datasets in Sec. 4.4.1 to generate a fourth dataset. Backgrounds were sampled without replacement and selected from train/validation/testing sets independently to avoid contamination. This dataset served as an additional source and was denoted as the amalgamated dataset. The performance of channels trained on the amalgamated dataset was evaluated on each of the three component datasets to simulate deploying the learned channels to each of the contributing imaging conditions.

4.5 AE Topology

The considered network topology was a tied-weight AE with no nonlinear or bias terms. This structure parallels the CHO formulation in Eq. (9), as the AE is learning the transformation matrix \mathbf{T} . Tied weights were chosen because they couple the encoder and the decoder by enforcing $\mathbf{W}_1^T = \mathbf{W}_2$, making the encoder a transpose of the decoder. This formulation prevents loss of

information that may solely exist in the decoder since only the encoder is employed as the transformation matrix. Additionally, tied weight AEs have fewer parameters to train and thus are expected to perform better in limited-data experiments.³⁶

4.6 Experimental Parameters

4.6.1 Training details

AE-channels were determined by minimizing the modified AE loss function in Eq. (14). The models were trained in Tensorflow⁶⁹ using the Adam algorithm.⁶² The AE weights were initialized using a truncated normal initializer with a standard deviation of 5×10^{-6} . The models were trained for 500 epochs. In cases where the considered dataset contained more than 500 images, pre-training the models on a subset of 500 images for 500 epochs to burn in the network sometimes improved performance. A mini-batch size of 250 was employed, with an equal number of signal-present and signal-absent images in each mini-batch. The learning rate was determined empirically and set to 5×10^{-3} for the VICTRE phantom background study and 1×10^{-5} for the lumpy background study. All networks were trained on a single NVIDIA TITAN X GPU.

Several channelized reference methods were implemented to be compared against the AE-learned channels, including convolutional LG,³⁴ PLS,³⁵ and the nonprewhitening matched filter (NPWMF).²¹ The HO-Direct²¹ was also computed on each subset using Eq. (5). The MP pseudoinverse was employed instead of the traditional inverse in these cases. Regularization was applied to the MP pseudoinverse in the form of truncating scaled singular values below 1×10^{-6} , with the cutoff point determined empirically. A grid search on the entire training dataset for each background was performed for each dataset size to select the parameters for all methods, with the maximum number of channels capped at 20. This grid search also implicitly provided multiple random initializations for the AE.

4.6.2 Evaluation

Subsets of the training data were considered to evaluate the performance of models as the amount of available data varied. The VICTRE experiments detailed in Sec. 4.2 contained subsets of size $K = 250, 500, 1000, 2000,$ and 5000 image pairs. The larger lumpy background experiments detailed in Sec. 4.1 also considered sets of $10,000, 15,000, 20,000, 25,000,$ and $30,000$ image pairs. The generalization and amalgamation studies considered a single case of $30,000$ images.

The standard train-validate-test scheme⁷⁰ was employed to evaluate performance. The AE and competing methods were given the training data and signal estimate to operate on, with the performance evaluated on the validation data to select the best set of parameters. For all methods, the mean signal image used during training was computed from the difference of the signal present and signal absent training images. Once the parameters were determined for each method, the CHO was numerically determined according to Eq. (10). The final models were then evaluated on the testing set to obtain the AUC values.

The HO-CMD was also computed for the experiments on lumpy backgrounds to analyze the efficiency of the channels for each method. This method has been demonstrated to approximate the HO when noiseless images are available and the noise model is known,² forming an upper bound for the experiments and permitting the evaluation of the efficiency of the channelized methods. The empirical background covariance matrix was calculated using the combined training and validation datasets for a total of $70,000$ noiseless background images. This method was unavailable for estimating the HO of the VICTRE experiments as noiseless images were not available.

5 Results

The results for the limited-image tests for the lumpy model and VICTRE breast phantom model are provided in Figs. 5 and 6. A visualization of the generated observer templates for each of the considered methods at the maximum number of considered images is included in Fig. 7, with a diagonal cross-section of some of the observers contained in Fig. 8. Overall, the proposed task-informed AE method was competitive with the state-of-the-art channelized methods for both

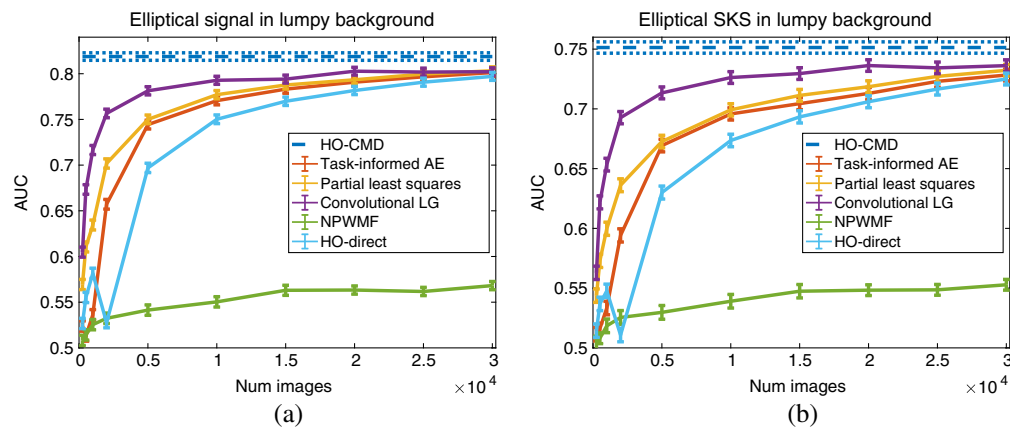


Fig. 5 Performance of the CHO on varying training dataset sizes for the lumpy background model. Panel (a) contains the results for the location-known elliptical signal while (b) contains the SKS elliptical signal results. The error bars correspond to the standard deviation of the fit AUC values. The HO-CMD is provided as an estimate of the upper bound of the HO, given an infinite amount of images, and is included to benchmark the efficiency of the channels for all methods.

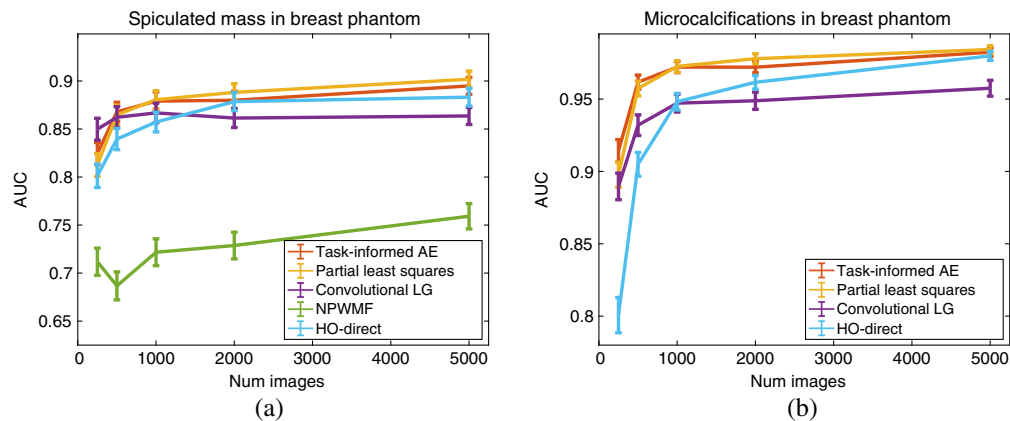


Fig. 6 Performance of the CHO on varying training dataset sizes for the VICTRE breast phantom model. Panel (a) contains the results for the spiculated mass signal while (b) contains the microcalcification cluster results. The error bars correspond to the standard deviation of the fit AUC values. The analytic HO estimate is not included as the clean images are not available to generate an estimate. The NPWMF is also omitted in (b) since it has an AUC below 0.53 for all considered dataset sizes, significantly less than the other methods.

the lumpy background and VICTRE phantom background cases. For the lumpy background cases, the best performing model was convolutional LG. This is likely due to the smooth, symmetric signal supporting LG channels. The task-informed AE and PLS methods produced AUC values that were within statistical error of each other at most of the considered dataset sizes. The HO-Direct had inferior performance to all channelized methods while also requiring significantly more computation to evaluate. Thus some channelized methods outperformed the standard method of computing the HO. The HO-CMD serves as an upper bound. As the amount of training data increases, it is expected that the channelized methods will approach the HO-CMD provided that sufficient channels are available since the CHO is a special case of the HO. However, this was not investigated as it requires significant amounts of data and is not a practical use case.

A few additional architectures were considered. The first modification to the architecture was replacing the decoder of the AE with a fully connected neural network and minimizing a sigmoid cross-entropy loss employing the image label rather than a mean image. This modification had inferior performance when compared to the proposed method. One potential reason for this is less information available for training the method—one label as opposed to a full signal image to compute the MSE. Previous works have demonstrated that it is possible to closely

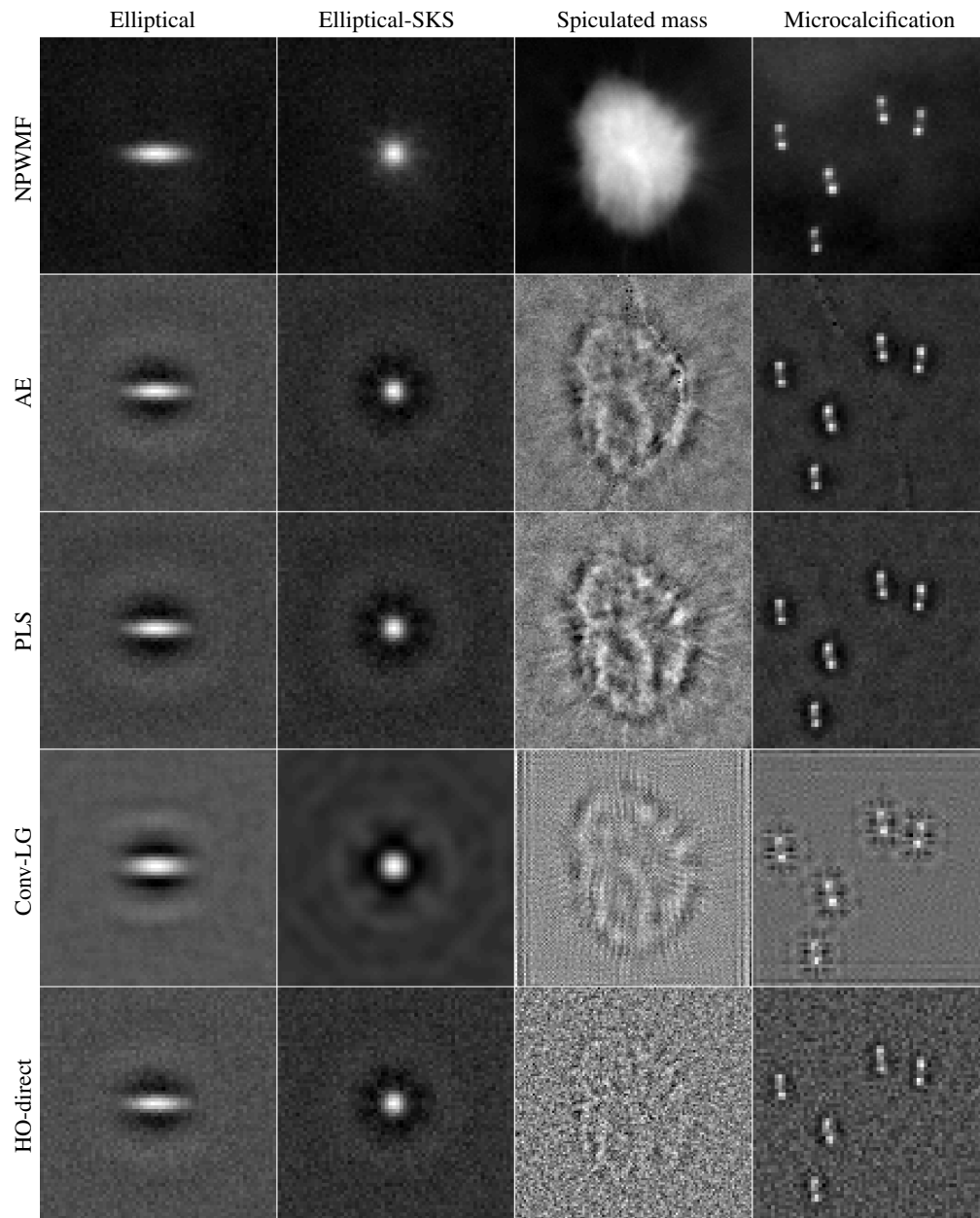


Fig. 7 Visualization of the observer templates obtained from each of the channelized methods for the considered backgrounds and signals. Each column contains results for a given signal/background combination and each row contains the observer templates for a specific channelized method.

approximate the HO when employing a label instead of an image, but a more sophisticated loss function must be considered.^{11,49,71} This architecture is also incompatible with the experiment proposed in Sec. 4.3. A second modification was adding a ReLU function after the encoder to introduce a nonlinear function into the training process. This modification required the model weights to be untied for the training process to converge. This architecture also had inferior performance to the proposed method. A potential reason for this is the untied weights. The AE learns to leverage the nonlinear function to improve the signal image reconstruction and lower the loss function, but when the weights of the AE are used as channels in the CHO they no longer benefit from the nonlinear function. The symmetry between the AE training method and the CHO discussed in Sec. 3 is broken. Finally, the conventional AE was also tested but failed to exceed 0.55 AUC in all four experiments.

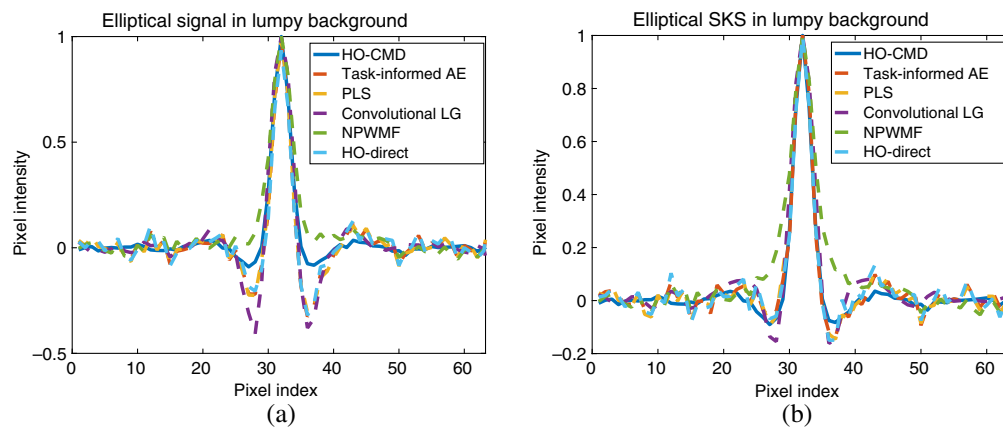


Fig. 8 Lineplots of the considered observers for the lumpy background model on the 30,000 image training dataset. The lineplot is a diagonal cross section of the observer template from the upper left to the lower right. Panel (a) contains the results for the location-known signal while (b) contains the SKS elliptical results.

In the VICTRE background case, the PLS and task-informed AE methods maintained their trend of the produced channels resulting in AUCs within statistical error of each other. However, in this case, both methods performed noticeably better than convolutional LG. The evolution of the channels for the spiculated mass as the number of images are increased is in Fig. 9. There was a noticeable seam in the observer templates for the task-informed AE channels on the phantom data, which is a remnant of the random initialization of the training method. This seam does not significantly affect observer performance, which may be why it was not removed in the training process. An additional contributing factor to the emergence of this seam may be the more complicated, nonsymmetric signal. Compared to the lumpy background experiments, the more complex VICTRE signal likely contributes to the substandard performance of the convolutional LG method. In addition, the increased structure complexity of the signal and the relatively higher noise are likely contributing factors in the earlier plateau of observer performance in the VICTRE case compared to the lumpy background experiments.

The learned channels for the 30,000 location-known lumpy image case and the 5000 VICTRE spiculated mass case are included in Figs. 10 and 11, respectively. Many of the channels are similar to one another in the features they extract and can be removed without significant loss of performance. These extraneous channels likely exist due to the AE training process. Random initializations generate different starting locations for each channel, which is iteratively optimized by the AE training process. During this process, the channels are updated to better jointly reconstruct the signal image. Thus even if the final model makes inefficient use of its full channel budget, the channels are influenced by their interactions during the training process. One of the limitations of this approach is its sensitivity to the random initialization, which can result in models of dramatically varying quality even with the same structure.

In all four of the considered detection tasks, the performance of the AE-learned channels most closely followed the PLS channels. The PLS filters for the elliptical and spiculated mass signals are included in Figs. 12 and 13. In contrast to the AE-learned channels, there is a distinct pattern for each of the considered filters. Thus while the performance of the AE-learned channels scales approximately the same as PLS with additional images, the channels the two methods generate are quite distinct.

Since the AE training process simultaneously updates all channels, there is no natural ordering for the importance of the channels. To evaluate the contributions of each channel provided to the overall AUC, an ablation study was performed for the AE-learned channels for the elliptical signal with 30,000 training images and the spiculated mass signal with 5000 training images. The channel that contributes the least was iteratively pruned to form a rank ordering of the importance of the channels. The results of the ablation study are in Fig. 14. Although the first few channels contribute significantly to the performance of the observer, the results quickly plateau. The plateau occurs more slowly with the more complex spiculated mass signal, indicating that signal complexity impacts the minimum number of beneficial channels. Beyond this point, the random

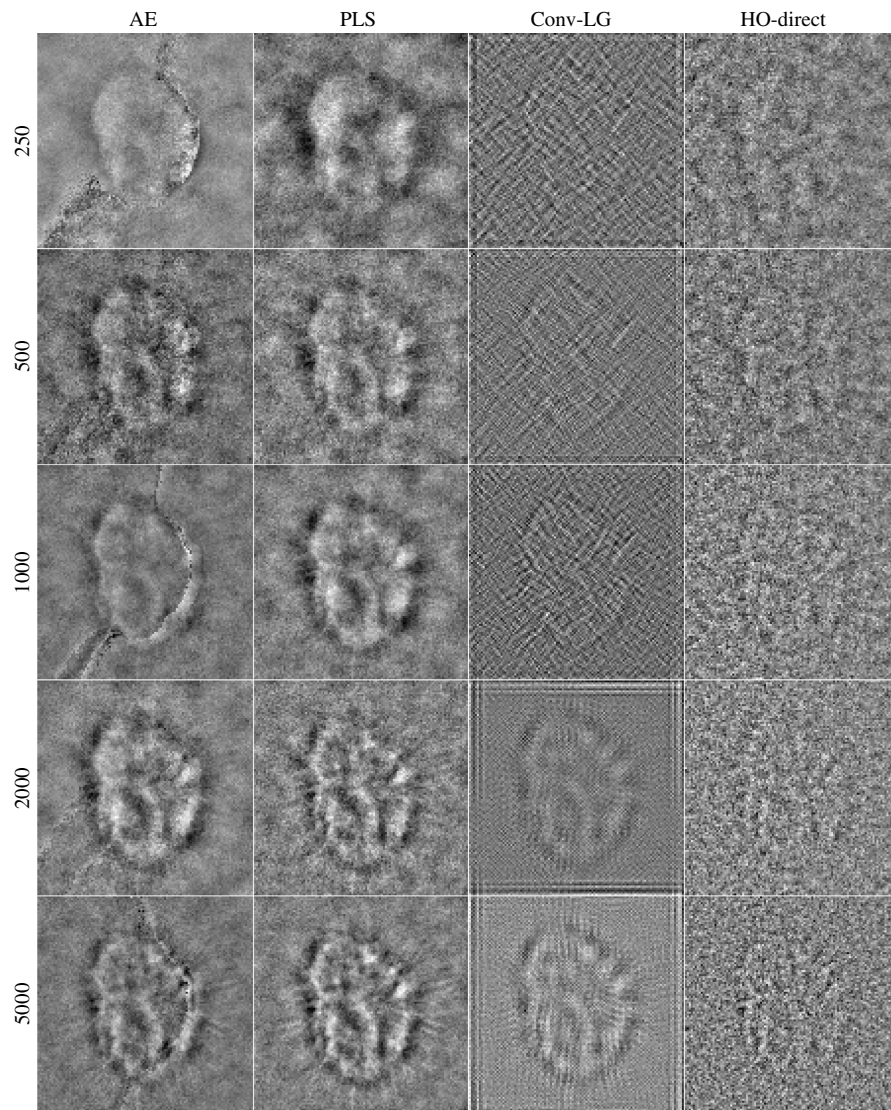


Fig. 9 Observer templates associated with the VICTRE spiculated mass experiment. Each column contains one of the considered channelized methods and the rows containing the size of the associated training dataset.

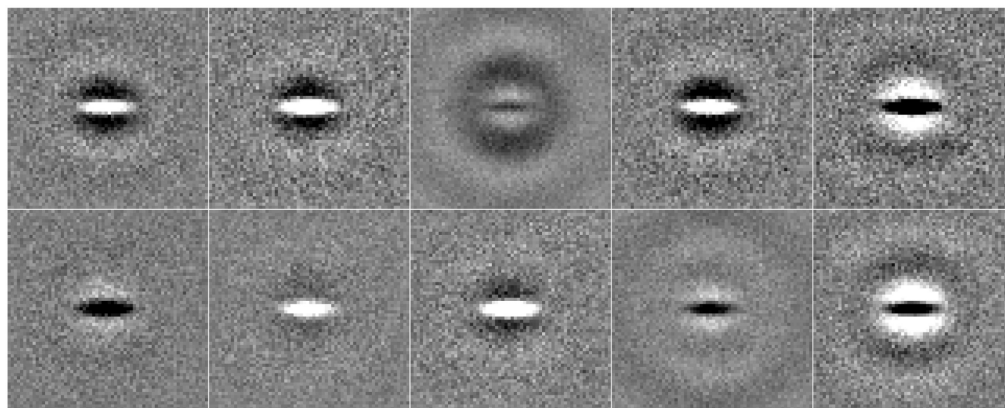


Fig. 10 Location-known elliptical channels associated with the AE-learned HO template for the 30,000 paired image dataset on the lumpy background. The grayscale is constant and fixed. The channels are ordered left to right, top to bottom. The ordering of the channels was determined through an ablation study.

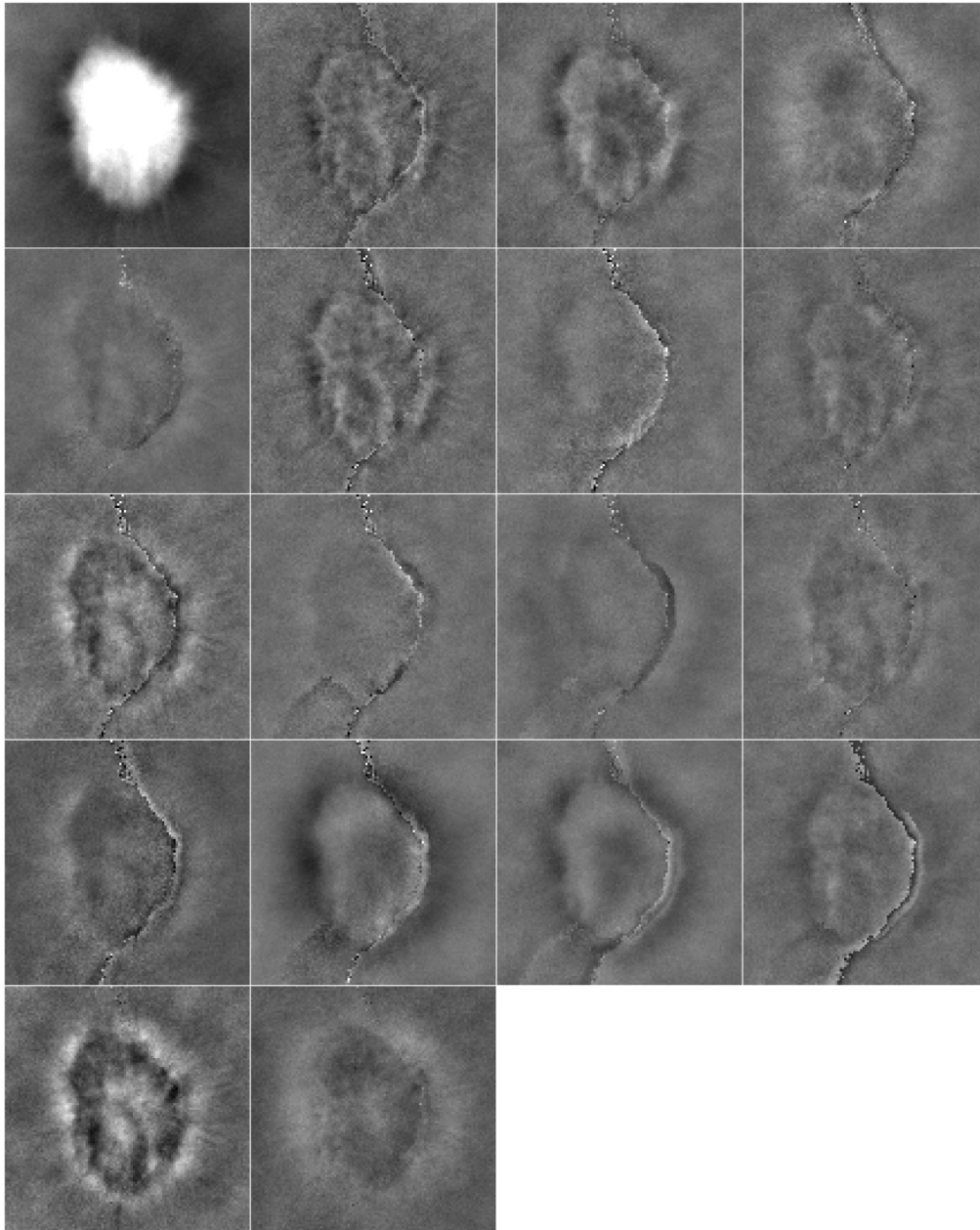


Fig. 11 Eighteen channels associated with the AE-learned HO template for the VICTRE spiculated mass detection task, displayed on the range $[-0.01, 0.01]$. The channels are ordered left to right, top to bottom. The ordering of the channels was determined through an ablation study. Note the presence of a seam running through several of the channels. This seam does not significantly impact observer performance and is likely a result of the stochastic learning process and more complicated structure of the signal.

initializations provided by sweeping over a range of the channels likely matter more than the additional channel budget.

The results for the fixed-signal lumpy background experiments are in Fig. 15. For the lumpy background cases, the task-informed AE channels performed significantly better than the PLS channels for all but the largest dataset sizes. In those cases, performance was comparable. Convolutional LG channel performance was relatively static since the models were tuned at the maximum dataset size, and it is not a learning method but were the best performing channels for the majority of the lumpy dataset sizes considered. However, both the PLS and task-informed AE channels outperformed convolutional LG when sufficient images were available. Compared to

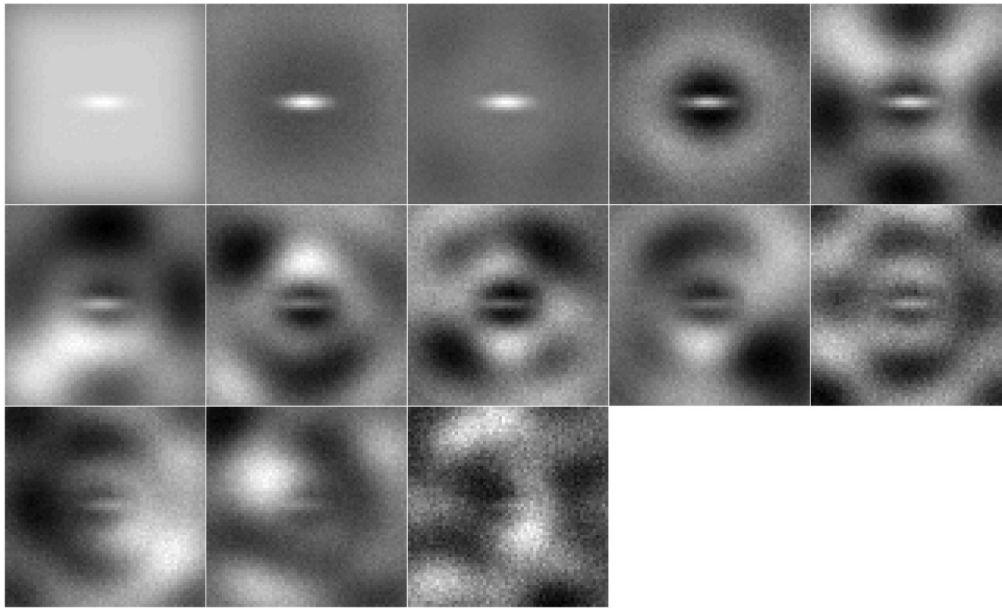


Fig. 12 Location-known elliptical channels associated with PLS for the 30,000 paired image dataset on the lumpy background. The grayscale is constant and fixed. The channels are ordered left to right, top to bottom. The channels are presented in the order produced by the iterative PLS method.

the standard experiments in Fig. 5, the performance of the task-informed AE and convolutional LG methods significantly improve and the PLS method retains approximately the same curve. This performance increase is especially notable for the smaller dataset sizes.

In the fixed-signal VICTRE background case, contained in Fig. 16, the task-informed AE-learned channels outperform every other tested method for the smaller training subsets. Given a sufficient amount of data, the AE and PLS channels approach the same AUC and are approximately equivalent. This occurred more quickly for the larger spiculated mass signal than the smaller microcalcification clusters. Compared to the fixed-signal lumpy background experiments, the performance of the observers on the VICTRE data plateaued more quickly. This is likely due to the relatively higher noise in the lumpy background experiments and the more complicated signal in the VICTRE case. The significant boost in the performance of the AE method when a cleaner signal image is available demonstrates the potential of the method to leverage prior information when it is available. Compared to the standard case in Fig. 6, the AE and convolutional LG methods again demonstrate improvement with the better signal estimate, whereas the PLS method's performance remains consistent. Overall, the ablation study results indicate that the proposed method is more sensitive to the quality of the mean signal than the amount of training images. This characteristic is shared with the FCO method but not PLS.

Although the ROI varied across the considered tasks, it did not appear to be a dominant factor in determining when the plateau occurred in observer performance. In particular, the spiculated mass detection task with the larger ROI had its AUC plateau more quickly than the simple elliptical signal with a smaller ROI. Despite the number of weights in the AE scaling with the size of the ROI, the statistical properties of the detection task are likely the determining factor for the amount of training data required. Thus the number of images required to reach the plateau is likely connected to the difficulty of the signal detection task. For the comparison of the spiculated mass and elliptical signal tasks, the difference in difficulty is most likely due to the magnitude of noise.

The results for the domain shift and amalgamation studies are included in Table 1. Visualizations of the observer templates obtained for each of the considered width values are included in Fig. 17. For the 2.0 and 4.0 source system widths, the channels obtained from the task-informed AE generalized well to target imaging systems with widths both above and below the source system. The improvement was especially noticeable when applying the channels

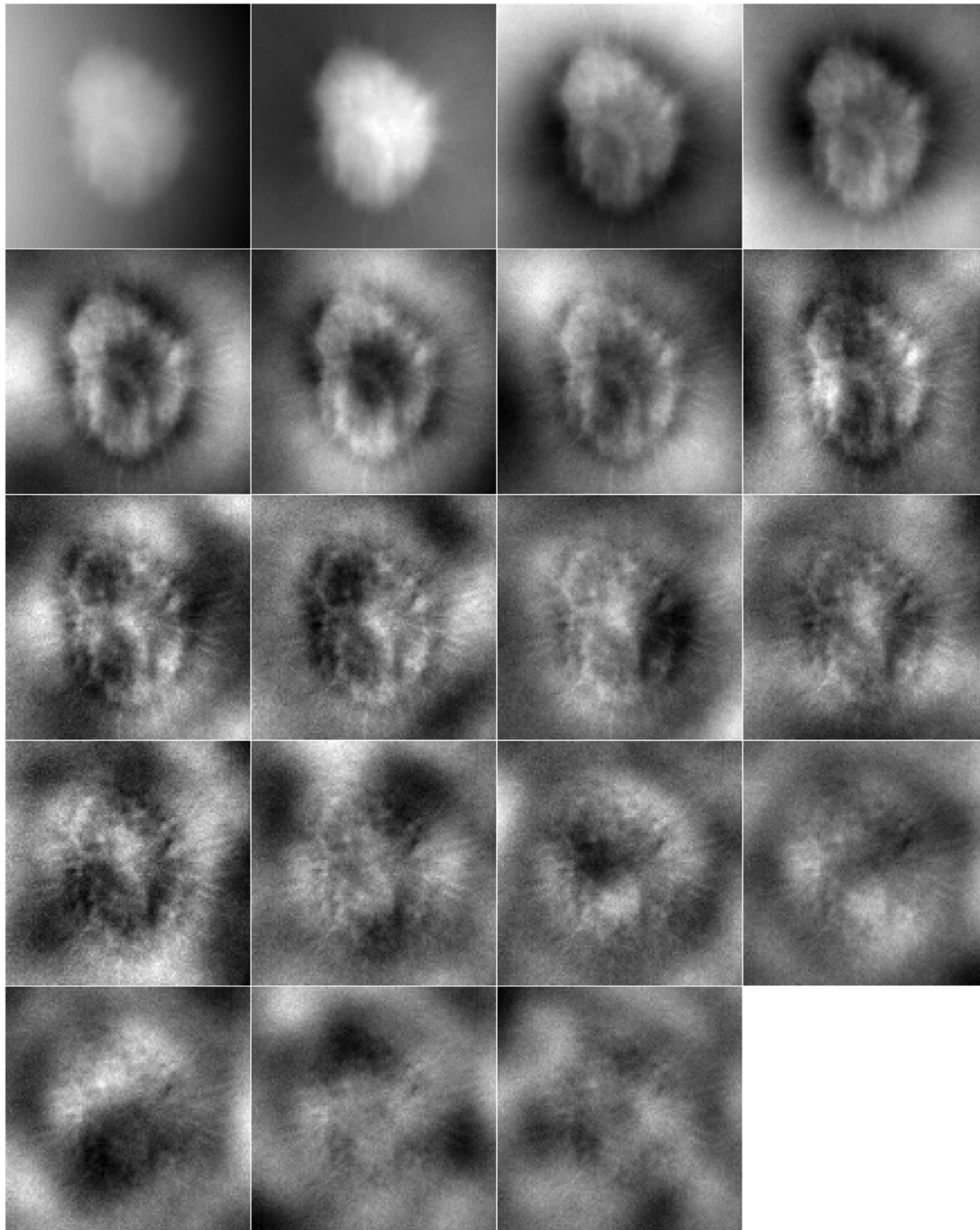


Fig. 13 Nineteen channels associated with the PLS HO template for the VICTRE spiculated mass detection task, displayed on the range $[-0.01, 0.01]$. The channels are ordered left to right, top to bottom. The channels are presented in the order produced by the iterative PLS method.

obtained from the 4.0 imaging system width data to the 1.0 imaging system width target images. In the 1.0 source imaging system width case, the channels from the task-informed AE were competitive when applied to data from the target imaging systems of 2.0 and 4.0 widths and outperformed both Conv-LG and HO-Direct channels, but PLS channels produced superior results. This indicates that the learned channels possess some measure of generalizability and resilience to errors in the employed signal image, despite the dependence of the proposed method on the estimate of the signal. PLS-generated channels demonstrated unusual behavior in that they generalized well to imaging systems with more blur but had noticeably inferior performance to the other considered methods when evaluated on imaging systems with less blurring. This is noticeable in the case with the channels obtained from the 1.0 source imaging system width generalizing the best of the considered methods to the 4.0 target system. However, PLS also performed the worst out of the observers when considering the channels obtained from the

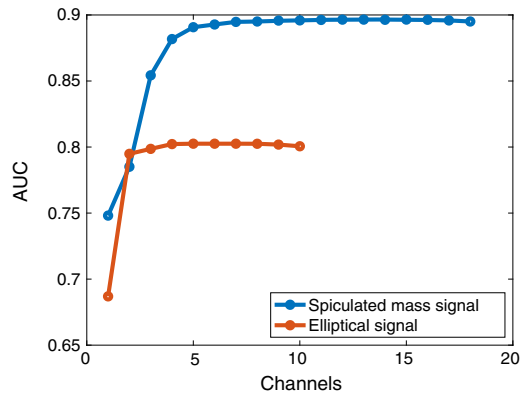


Fig. 14 Results of the ablation study performed for the elliptical and spiculated mass experiments. The channel contributing the least was iteratively pruned to produce a rank ordering of the channels. A plateau occurs after a certain number of channels, depending on the signal complexity.

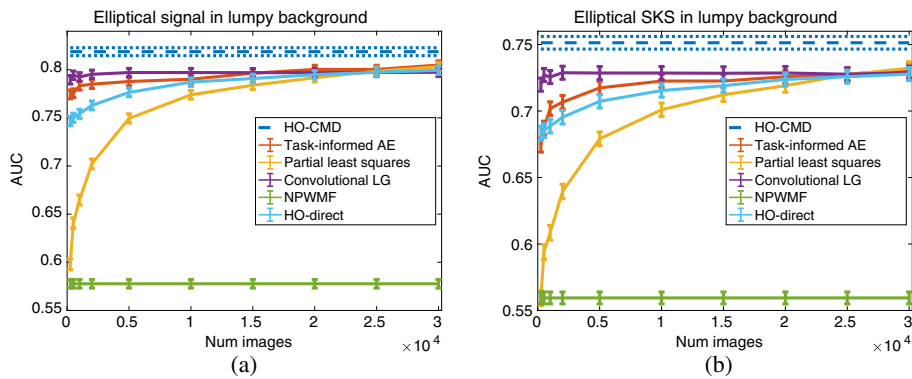


Fig. 15 Performance of the CHO on varying training dataset sizes for the lumpy background model with a fixed mean signal estimated from the training and validation datasets. Panel (a) contains the results for the location-known elliptical signal while (b) contains the SKS elliptical signal results. The error bars correspond to the standard deviation of the fit AUC values. The HO-CMD is provided as an estimate of the upper bound of the HO, is given an infinite amount of images, and is included to benchmark the efficiency of the channels for all methods.

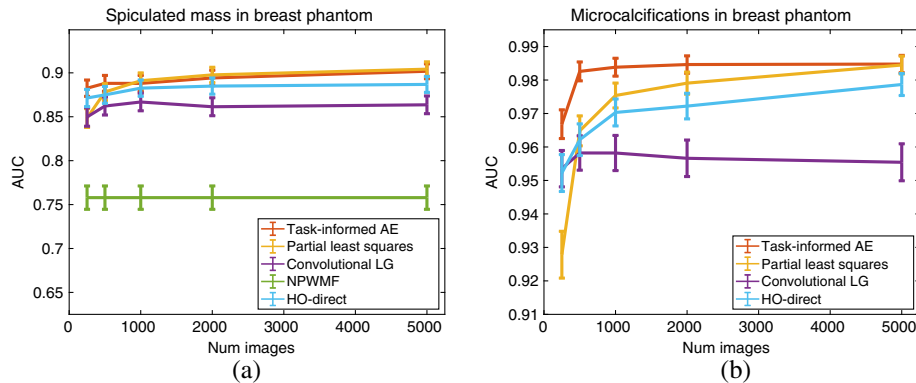


Fig. 16 Performance of the CHO on varying training dataset sizes for the VICTRE breast phantom model with a fixed mean signal estimated from the training and validation datasets. Panel (a) contains the results for the spiculated mass signal while (b) contains the microcalcification cluster results. The error bars correspond to the standard deviation of the fit AUC values. The analytic HO estimate is not included as the clean images are not available to generate an estimate. The NPWMF is also omitted in (b) since it has an AUC below 0.53 for all considered dataset sizes, significantly less than the other methods.

Table 1 Results from the domain shift and amalgamation studies. Observers for each method were first obtained from each source system. Each of these observers were then applied to each target imaging system to evaluate their performance in the presence of domain shift. The source amalg case refers to the even sampling of the source system datasets to generate an amalgamated dataset, which was evaluated in the same way for each target imaging system. The standard error for all of the values is ± 0.005 . The highest values that are statistically significant for each of the target imaging systems within a source image system width are in bold.

Source system width	Observer method	Target system width		
		1.0	2.0	4.0
1.0	Task-informed AE	0.99	0.91	0.68
	PLS	0.99	0.92	0.71
	Conv LG	0.99	0.90	0.64
	HO-Direct	0.99	0.90	0.64
2.0	Task-informed AE	0.99	0.92	0.71
	PLS	0.99	0.92	0.71
	Conv LG	0.98	0.91	0.69
	HO-direct	0.99	0.92	0.70
4.0	Task-informed AE	0.93	0.88	0.73
	PLS	0.85	0.80	0.72
	Conv LG	0.89	0.85	0.73
	HO-direct	0.92	0.87	0.73
Amalg	Task-informed AE	0.98	0.92	0.73
	PLS	0.96	0.90	0.74
	Conv LG	0.98	0.91	0.71
	HO-direct	0.97	0.92	0.74

4.0 source imaging system data to the 1.0 target imaging system, with the performance of PLS lagging the performance of the task-informed AE channels by 0.08 AUC.

During the course of the experiments, it was observed that the convolutional LG channels were especially sensitive to the quality of the estimated signal. When provided with the signal used to generate the data in both the location-known and SKS lumpy experiments, the method outperformed all other competitors. When fewer images were used to generate the mean signal image and thus there is more noise in the signal image, such as in the VICTRE phantom dataset, the performance degraded significantly. This is illustrated in the dichotomy between standard and fixed signal results in Figs. 5 and 15. In the fixed signal case, the performance of the method barely changes as the number of paired images sweep from 500 to 30,000. Thus the performance improvement in the standard case is almost entirely due to the increase in the quality of the mean signal image. This is also visible in the observer templates constructed using convolutional LG channels, as they differ significantly from the templates constructed via other channelized methods. Although the task-informed AE-learned channels attempt to reconstruct the given signal image directly, and thus would seem to be impacted more by noise, the method was more robust to error in the estimated signal than the convolutional LG approach. Denoising AEs is known to reduce noise in an image since the AE embedding's limited dimensionality lacks the capacity to represent the high-frequency components of the noise. The reduced sensitivity of the proposed method to minor variations in the data may be a result of this property.

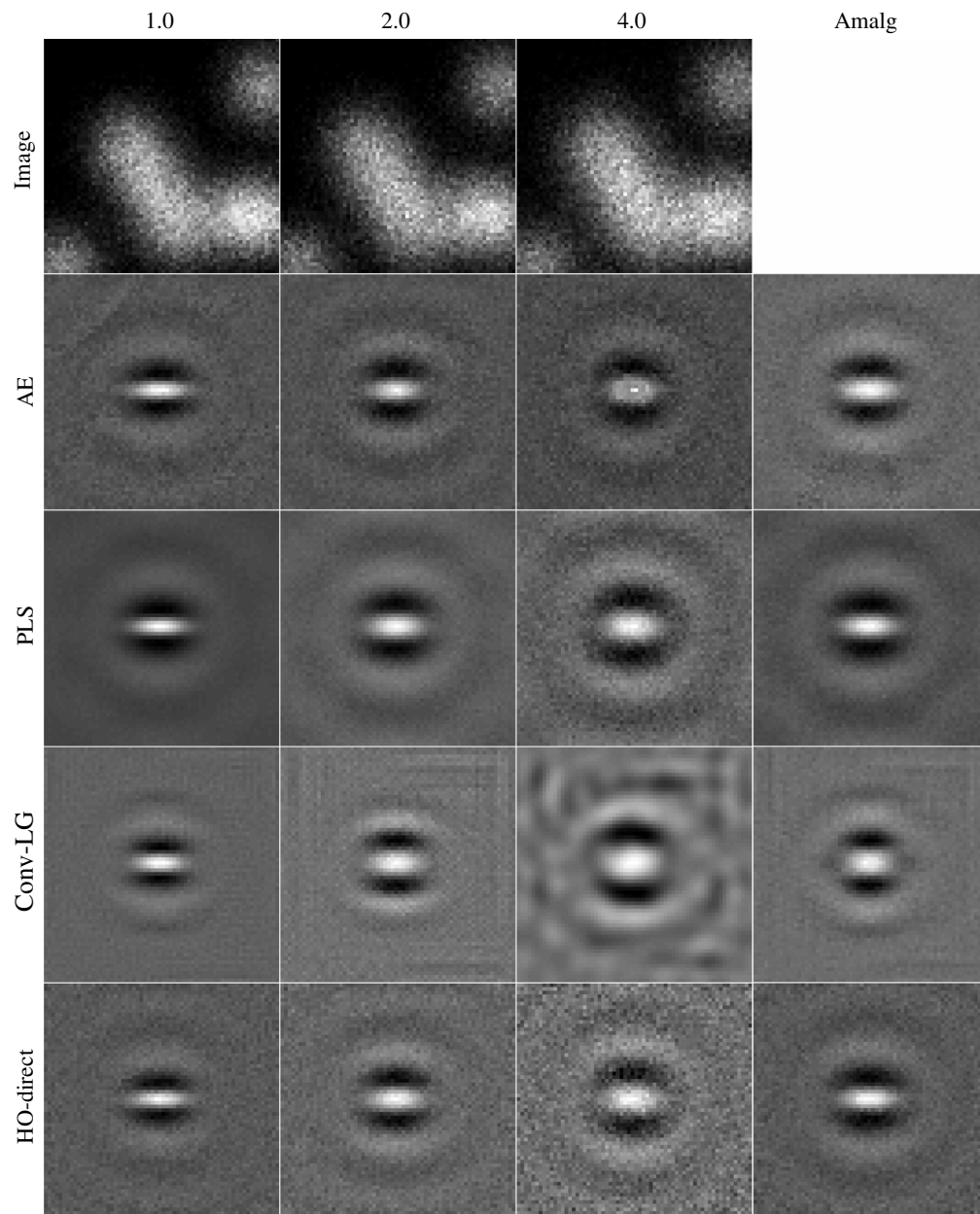


Fig. 17 Sample images and templates from the generalization and amalgamation studies. The columns contain the results from varying the width of the imaging system, increasing from left to right. The first row contains a sample signal absent image from the dataset. The remaining rows contain the obtained observer templates for the considered methods.

6 Discussion and Conclusion

This study demonstrated that AEs are capable of learning efficient CHO channels for both location known and fixed-centroid SKS signal detection tasks. Data embeddings and observer channels were demonstrated to be fundamentally related, with the task of optimizing a data embedding to preserve signal-specific information equivalent to determining an efficient channel selection for the CHO. Furthermore, the presented method of computing channels is capable of meeting or exceeding the performance of state-of-the-art methods on the investigated tasks.

Channels were learned for the CHO by minimizing the reconstruction loss of an AE. Modification of the AE loss function to focus only on task-specific information involving the signal was found to have a significant benefit over using the conventional AE loss. Empirical sweeps over the network topology revealed that the AE could efficiently approximate the HO for

a wide range of cases utilizing comparable numbers of channels to other approaches. The proposed method was equivalent to state-of-the-art approaches for the lumpy and VICTRE backgrounds. When a higher-quality signal estimate was provided, the AE-channels demonstrated significantly higher performance and were superior to the compared channelized methods on the VICTRE breast phantom dataset containing complicated signals. This increase was especially notable on smaller dataset sizes. Additionally, ability of the channels to generalize to imaging systems with varying parameters was evaluated. On average, the proposed method outperformed the other considered channelized approaches in cross-imaging system performance despite the injection of the signal image during the training process. In terms of appearance and performance, the AE channels were closest to PLS channels and the performance of the two methods was typically within error. However, AE-learned channels were noticeably better at generalizing. The AE-learned channels were sensitive to the random initialization of the weights and frequently learned redundant channels. The training scheme can likely be further improved with a more robust approach to weight initialization.

Opportunities for future work include evaluating the AE-learned channels with a channelized IO and extending the formulation to both more sophisticated SKS cases and 3D input images. The channels should work directly for any standard Markov chain Monte Carlo method for estimating the IO. Although the current form of the loss function for learning AE-channels requires knowing the signal centroid, it could be generalized by considering convolutional AEs.⁷²

Another aspect that can be further developed is the resilience of the method to deterministic errors in the employed signal image. For instance, the supplied mean signal might be offset or larger/smaller than the original image. Although the generalizability study explored this in a limited way, a more structured study would be beneficial for developing more resilient methods. A deterministic shift of the mean image may benefit from untying the encoder and decoder, for instance.

The superior performance of AE-learned channels on smaller datasets and medically realistic phantoms also expands the applicability of the method to real-world cases, and the method should be tested on experimental data to identify remaining challenges in tuning the AE. The proposed method may also be applied toward learning anthropomorphic channels by replacing the ground-truth labels with human-generated labels during the training process. Acquiring human labels is a complex task that has several different elements in the study design, depending on the specific detection task and desired use case.^{73–76}

Disclosures

The authors declare no potential conflicts of interest.

Data and Code Availability

Code and data for this paper are publicly available at <https://github.com/jasonlg/AETSI>.

Acknowledgments

This work was supported in part by the US National Institutes of Health (NIH), Awards EB031772 (subproject 6366), EB031585, and EB034249. Preliminary results of this work were presented at SPIE Medical Imaging 2019 and published as an SPIE Proceedings paper.³⁶

References

1. R. F. Wagner and D. G. Brown, "Unified SNR analysis of medical imaging systems," *Phys. Med. Biol.* **30**(6), 489 (1985).
2. H. H. Barrett and K. J. Myers, *Foundations of Image Science*, John Wiley & Sons (2013).
3. M. A. Kupinski et al., "Ideal-observer computation in medical imaging with use of Markov-Chain Monte Carlo techniques," *J. Opt. Soc. Am. A* **20**(3), 430–438 (2003).
4. S. Park et al., "Channelized-ideal observer using Laguerre–Gauss channels in detection tasks involving non-Gaussian distributed lumpy backgrounds and a Gaussian signal," *J. Opt. Soc. Am. A* **24**(12), B136–B150 (2007).
5. S. Park and E. Clarkson, "Efficient estimation of ideal-observer performance in classification tasks involving high-dimensional complex backgrounds," *J. Opt. Soc. Am. A* **26**(11), B59–B71 (2009).

6. F. Shen and E. Clarkson, "Using Fisher information to approximate ideal-observer performance on detection tasks for lumpy-background images," *J. Opt. Soc. Am. A* **23**(10), 2406–2414 (2006).
7. W. Vennart, "ICRU report 54: medical imaging—the assessment of image quality," *Radiography* **3**, 243–244 (1996).
8. X. He, B. S. Caffo, and E. C. Frey, "Toward realistic and practical ideal observer (IO) estimation for the optimization of medical imaging systems," *IEEE Trans. Med. Imaging* **27**(10), 1535–1543 (2008).
9. C. K. Abbey and J. M. Boone, "An ideal observer for a model of X-ray imaging in breast parenchymal tissue," *Lect. Notes Comput. Sci.* **5116**, 393–400 (2008).
10. W. Zhou and M. A. Anastasio, "Learning the ideal observer for SKE detection tasks by use of convolutional neural networks," *Proc. SPIE* **10577**, 1057719 (2018).
11. W. Zhou, H. Li, and M. Anastasio, "Approximating the ideal observer and Hotelling observer for binary signal detection tasks by use of supervised learning methods," *IEEE Trans. Med. Imaging* **38**, 2456–2468 (2019).
12. M. A. Kupinski et al., "Ideal observer approximation using Bayesian classification neural networks," *IEEE Trans. Med. Imaging* **20**(9), 886–899 (2001).
13. W. Zhou and M. A. Anastasio, "Markov-Chain Monte Carlo approximation of the ideal observer using generative adversarial networks," *Proc. SPIE* **11316**, 113160D (2020).
14. W. Zhou, U. Villa, and M. A. Anastasio, "Ideal observer computation by use of Markov-chain Monte Carlo with generative adversarial networks," *IEEE Trans. Med. Imaging* (2023).
15. Y. Chen et al., "Reconstruction-aware imaging system ranking by use of a sparsity-driven numerical observer enabled by variational Bayesian inference," *IEEE Trans. Med. Imaging* **38**, 1251–1262 (2019).
16. H. H. Barrett et al., "Task-based measures of image quality and their relation to radiation dose and patient risk," *Phys. Med. Biol.* **60**(2), R1 (2015).
17. I. Reiser and R. Nishikawa, "Task-based assessment of breast tomosynthesis: effect of acquisition parameters and quantum noise," *Med. Phys.* **37**(4), 1591–1600 (2010).
18. A. A. Sanchez, E. Y. Sidky, and X. Pan, "Task-based optimization of dedicated breast CT via Hotelling observer metrics," *Med. Phys.* **41**(10), 101917 (2014).
19. S. J. Glick, S. Vedantham, and A. Karellas, "Investigation of optimal kVp settings for CT mammography using a flat-panel imager," *Proc. SPIE* **4682**, 392–403 (2002).
20. H. H. Barrett et al., "Linear discriminants and image quality," *Image Vision Comput.* **10**(6), 451–460 (1992).
21. H. H. Barrett et al., "Model observers for assessment of image quality," *Proc. Natl. Acad. Sci. U. S. A.* **90**(21), 9758–9765 (1993).
22. H. H. Barrett et al., "Megalopinakophobia: its symptoms and cures," *Proc. SPIE* **4320**, 299–308 (2001).
23. M. A. Kupinski, E. Clarkson, and J. Y. Hesterman, "Bias in Hotelling observer performance computed from finite data," *Proc. SPIE* **6515**, 65150S (2007).
24. W. Zhou, H. Li, and M. A. Anastasio, "Learning the Hotelling observer for SKE detection tasks by use of supervised learning methods," *Proc. SPIE* **10952**, 1095208 (2019).
25. H. H. Barrett et al., "Stabilized estimates of Hotelling-observer detection performance in patient-structured noise," *Proc. SPIE* **3340**, 27–44 (1998).
26. B. D. Gallas and H. H. Barrett, "Validating the use of channels to estimate the ideal linear observer," *J. Opt. Soc. Am. A* **20**, 1725–1738 (2003).
27. G. Kim et al., "A convolutional neural network-based model observer for breast ct images," *Med Phys* **47**(4), 1619–1632 (2020).
28. J. B. Tenenbaum, "Mapping a manifold of perceptual observations," in *Adv. in Neural Inf. Process. Syst. 10*, M. I. Jordan, M. J. Kearns, and S. A. Solla, Eds., MIT Press, pp. 682–688 (1998).
29. D. Beymer and T. Poggio, "Image representations for visual learning," *Science* **272**(5270), 1905–1909 (1996).
30. M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.* **3**(1), 71–86 (1991).
31. K. Q. Weinberger and L. K. Saul, "Unsupervised learning of image manifolds by semidefinite programming," *Int. J. Comput. Vision* **70**, 77–90 (2006).
32. 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).
33. S. Park, J. M. Witten, and K. J. Myers, "Singular vectors of a linear imaging system as efficient channels for the Bayesian ideal observer," *IEEE Trans. Med. Imaging* **28**, 657–668 (2009).
34. I. Diaz et al., "Derivation of an observer model adapted to irregular signals based on convolution channels," *IEEE Trans. Med. Imaging* **34**(7), 1428–1435 (2015).
35. J. M. Witten, S. Park, and K. J. Myers, "Partial least squares: a method to estimate efficient channels for the ideal observers," *IEEE Trans. Med. Imaging* **29**, 1050–1058 (2010).
36. J. L. Granstedt, W. Zhou, and M. A. Anastasio, "Autoencoder embedding of task-specific information," *Proc. SPIE* **10952**, 1095207 (2019).

37. J. G. Brankov et al., "Learning a channelized observer for image quality assessment," *IEEE Trans. Med. Imaging* **28**, 991–999 (2009).
38. 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**, 109520A (2019).
39. B. Kim, M. Han, and J. Baek, "Convolutional neural network-based anthropomorphic model observer for breast cone-beam CT images," *Proc. SPIE* **11316**, 113160Y (2020).
40. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*, D. E. Rumelhart and J. L. McClelland C. PDP Research Group, Eds., Vol. **1**, pp. 318–362, MIT Press, Cambridge, MA (1986).
41. G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.* **18**(7), 1527–1554 (2006).
42. G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science* **313**, 504–507 (2006).
43. Y. Bengio and Y. Lecun, *Scaling Learning Algorithms Towards AI*, MIT Press (2007).
44. D. Erhan et al., "Why does unsupervised pre-training help deep learning?" *J. Mach. Learn. Res.* **11**, 625–660 (2010).
45. P. Vincent et al., "Extracting and composing robust features with denoising autoencoders," in *Proc. Twenty-Fifth Int. Conf. Mach. Learn. (ICML'08)*, W. W. Cohen, A. McCallum, and S. T. Roweis, Eds., ACM, pp. 1096–1103 (2008).
46. M. Nishio et al., "Convolutional auto-encoder for image denoising of ultra-low-dose ct," *Heliyon* **3**(8), e00393 (2017).
47. Z. Zhang, Y. Song, and H. Qi, "Age progression/regression by conditional adversarial autoencoder," in *IEEE Conf. Comput. Vision and Pattern Recognit. (CVPR)*, pp. 4352–4360 (2017).
48. D. Kunin et al., "Loss landscapes of regularized linear autoencoders," in *Proc. 36th Int. Conf. Mach. Learn.*, K. Chaudhuri and R. Salakhutdinov, Eds., Vol. **97**, pp. 3560–3569 (2019).
49. J. L. Granstedt, W. Zhou, and M. A. Anastasio, "Learning efficient channels with a dual loss autoencoder," *Proc. SPIE* **11316**, 113160C (2020).
50. C. E. Metz, "ROC methodology in radiologic imaging," *Investig. Radiol.* **21**(9), 720–733 (1986).
51. A. Badano et al., "Evaluation of digital breast tomosynthesis as replacement of full-field digital mammography using an in silico imaging trial," *JAMA Network Open* **1**, e185474 (2018).
52. R. Zeng et al., "Computational reader design and statistical performance evaluation of an in-silico imaging clinical trial comparing digital breast tomosynthesis with full-field digital mammography," *J. Med. Imaging* **7**(4), 042802 (2020).
53. J. Schmidhuber, "Deep learning in neural networks: an overview," *Neural Networks* **61**, 85–117 (2015).
54. Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature* **521**(7553), 436 (2015).
55. K. Hornik, M. Stinchcombe, and H. White, "Multilayer feedforward networks are universal approximators," *Neural Networks* **2**(5), 359–366 (1989).
56. K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks* **4**(2), 251–257 (1991).
57. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: a survey," *ACM Comput. Surv.* **41**, 1–58 (2009).
58. A. Mousavi, A. B. Patel, and R. G. Baraniuk, "A deep learning approach to structured signal recovery," in *53rd Annu. Allerton Conf. Commun., Control, and Comput. (Allerton)*, pp. 1336–1343 (2015).
59. J. Snoek, R. Adams, and H. Larochelle, "On nonparametric guidance for learning autoencoder representations," in *Proc. Fifteenth Int. Conf. Artif. Intell. and Stat.*, N. D. Lawrence and M. Girolami, Eds., Vol. **22**, pp. 1073–1080, PMLR, La Palma, Canary Islands (2012).
60. J. Snoek, R. P. Adams, and H. Larochelle, "Nonparametric guidance of autoencoder representations using label information," *J. Mach. Learn. Res.* **13**, 2567–2588 (2012).
61. P. Vincent et al., "Stacked denoising autoencoders: learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.* **11**, 3371–3408 (2010).
62. D. P. Kingma and J. Ba, "Adam: a method for stochastic optimization," in *3rd Int. Conf. Learn. Represent.*, San Diego, California, Y. Bengio and Y. LeCun, Eds., (2015).
63. C. E. Metz and X. Pan, "'Proper' binormal ROC curves: theory and maximum-likelihood estimation," *J. Math. Psychol.* **43**(1), 1–33 (1999).
64. X. Pan and C. E. Metz, "The 'proper' binormal model: parametric receiver operating characteristic curve estimation with degenerate data," *Acad. Radiol.* **4**(5), 380–389 (1997).
65. J. Rolland and H. H. Barrett, "Effect of random background inhomogeneity on observer detection performance," *J. Opt. Soc. Am. A* **9**(5), 649–658 (1992).
66. M. A. Kupinski et al., "Experimental determination of object statistics from noisy images," *J. Opt. Soc. Am. A* **20**(3), 421–429 (2003).

67. B. Sun, J. Feng, and K. Saenko, "Return of frustratingly easy domain adaptation," in *Proc. Thirtieth AAAI Conf. Artif. Intell., AAAI'16*, AAAI Press, pp. 2058–2065 (2016).
68. M. Makitalo and A. Foi, "Optimal inversion of the generalized Anscombe transformation for Poisson–Gaussian noise," *IEEE Trans. Image Process.* **22**(1), 91–103 (2013).
69. M. Abadi et al., "Tensorflow: a system for large-scale machine learning," in *OSDI*, Vol. 16, pp. 265–283 (2016).
70. I. Goodfellow et al., *Deep Learning*, Vol. 1, MIT Press, Cambridge (2016).
71. J. L. Granstedt et al., "Learned Hotelling observers for use with multi-modal data," *Proc. SPIE* **12035**, 1203512 (2022).
72. M. Ranzato et al., "Unsupervised learning of invariant feature hierarchies with applications to object recognition," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1–8 (2007).
73. D. Georgian-Smith et al., "Can digital breast tomosynthesis replace full-field digital mammography? A multireader, multicase study of wide-angle tomosynthesis," *Am. J. Roentgenol.* **212**(6), 1393–1399 (2019).
74. C. Castella et al., "Mass detection in breast tomosynthesis and digital mammography: a model observer study," *Proc. SPIE* **7263**, 72630O (2009).
75. L. Yu et al., "Prediction of human observer performance in a 2-alternative forced choice low-contrast detection task using channelized Hotelling observer: impact of radiation dose and reconstruction algorithms," *Med. Phys.* **40**(4), 041908 (2013).
76. F. Jäkel and F. A. Wichmann, "Spatial four-alternative forced-choice method is the preferred psychophysical method for Naïve observers," *J. Vision* **6**, 13–13 (2006).

Jason L. Granstedt received his BS and MS degrees in electrical engineering from Virginia Polytechnic Institute and State University in 2015 and 2017, respectively. He is currently a PhD candidate in the Department of Computer Science at UIUC. His research interests include task-based analysis of images and application of machine learning techniques to medical imaging.

Weimin Zhou is an assistant professor at the Global Institute of Future Technology at Shanghai Jiao Tong University (SJTU). He received his PhD in electrical engineering from Washington University in St. Louis. His research focuses on biomedical imaging, machine learning, image science, and visual perception. He is the recipient of the SPIE Community Champion and serves as a program committee member of SPIE Medical Imaging.

Mark A. Anastasio is the Donald Biggar Willett Professor in Engineering and the head of the Department of Bioengineering at UIUC. He is a fellow of SPIE, the American Institute for Medical and Biological Engineering, and the International Academy of Medical and Biological Engineering. His research addresses computational image science, inverse problems in imaging, and machine learning for imaging applications. He has contributed to emerging biomedical imaging technologies, including photoacoustic computed tomography and ultrasound computed tomography.