Journal of Medical Imaging

MedicalImaging.SPIEDigitalLibrary.org

Design of a practical model-observer-based image quality assessment method for x-ray computed tomography imaging systems

Hsin-Wu Tseng
Jiahua Fan
Matthew A. Kupinski

SPIE.

Hsin-Wu Tseng, Jiahua Fan, Matthew A. Kupinski, "Design of a practical model-observer-based image quality assessment method for x-ray computed tomography imaging systems," *J. Med. Imag.* **3**(3), 035503 (2016), doi: 10.1117/1.JMI.3.3.035503.

Journal of Medical Imaging 3(3), 035503 (Jul–Sep 2016)

Design of a practical model-observer-based image quality assessment method for x-ray computed tomography imaging systems

Hsin-Wu Tseng,[a,b,*] Jiahua Fan,[b] and Matthew A. Kupinski[a]
[a]The University of Arizona, College of Optical Sciences, Tucson, Arizona 85721, United States
[b]CT Engineering, GE Healthcare, Waukesha, Wisconsin 53188, United States

Abstract. The use of a channelization mechanism on model observers not only makes mimicking human visual behavior possible, but also reduces the amount of image data needed to estimate the model observer parameters. The channelized Hotelling observer (CHO) and channelized scanning linear observer (CSLO) have recently been used to assess CT image quality for detection tasks and combined detection/estimation tasks, respectively. Although the use of channels substantially reduces the amount of data required to compute image quality, the number of scans required for CT imaging is still not practical for routine use. It is our desire to further reduce the number of scans required to make CHO or CSLO an image quality tool for routine and frequent system validations and evaluations. This work explores different data-reduction schemes and designs an approach that requires only a few CT scans. Three different kinds of approaches are included in this study: a conventional CHO/CSLO technique with a large sample size, a conventional CHO/CSLO technique with fewer samples, and an approach that we will show requires fewer samples to mimic conventional performance with a large sample size. The mean value and standard deviation of areas under ROC/EROC curve were estimated using the well-validated shuffle approach. The results indicate that an 80% data reduction can be achieved without loss of accuracy. This substantial data reduction is a step toward a practical tool for routine-task-based QA/QC CT system assessment. [DOI: 10.1117/1.JMI.3.3.035503]

Keywords: model observer; CHO; CSLO; data reduction; leave-one-out likelihood.

Paper 16013PRRR received Jan. 22, 2016; accepted for publication Jul. 1, 2016; published online Jul. 28, 2016.

1 Introduction

For detection tasks and combined detection/estimation tasks, in which the observer must detect a signal and estimate parameters of the signal (e.g., contrast and size), the channelized Hotelling observer (CHO)[1–11] and the channelized scanning-linear observer (CSLO)[12] have successfully been utilized to measure image quality. To implement these two observers, first-order and second-order statistics are needed, which are typically estimated using sample images from a training dataset. For an M-element regions of interest (ROI), without channelization, the inverse of the sample covariance matrix exists only when the number of images in the training dataset is larger than $M+1$. In a CT imaging system, however, it is difficult to have a training dataset larger than $M+1$ because M is usually very large—even for a small ROI. The process of channelization in which the image data are processed by a small number of channels to produce channel outputs provides a solution to this problem. The number of channels is typically much less than M. For instance, the number of dense-difference of Gauss (DDOG)[13] channels selected in this study is 10, so the dimension of each channelized sample becomes 10, whereas the ROI size might contain $M=100\times100=10^4$ pixels. Thus, the minimum required training size reduces to 11 instead of 10^4+1 in this specific example. The channelization makes the CHO and CSLO possible for limited datasets in CT systems. Although the channelized model observers become possible tools in CT imaging systems, the number of required images to estimate both the mean vector and the covariance matrix accurately is still quite large. Thus, conventional channelized model observers cannot be used as a quality assurance tool for routine system validation. In this study, an approach is proposed that reduces the required number of CT images and enables routine system evaluations using model observers for task-based assessments of image quality.

The goal of this study is to explore different methods to reduce data demand for computing CHO/CSLO performance and propose a practical model-observer-based approach for potential routine QA/QC of CT imaging systems. We will use high-dose images to estimate the first-order statistics, i.e., the mean signal image, and use the leave-one-out covariance (LOOC)[14] method to estimate the second-order statistics, i.e., the covariance matrices.

2 Materials and Methods

2.1 *Real Data Preparation*

In this work, low contrast (LC) objects embedded in four different phantoms were imaged on a GE Discovery CT750 HD CT system. Two of the phantoms were QRM-LC-FD1 and QRM-LC-FD4 phantoms (QRM Quality Assurance in Radiology and Medicine GmbH, Moehrendorf, Germany) and the other two

*Address all correspondence to: Hsin-Wu Tseng, E-mail: tseng@optics.arizona.edu

2329-4302/2016/$25.00 © 2016 SPIE

were MITA CCT 183 and MITA CCT 189 phantoms (The Phantom Laboratory, Salem, New York). The LC objects embedded in each phantom varied in size, contrast, or both. In this study, we used them differently for different study tasks. The MITA CCT 183 phantom was used for pure detection tasks and the validation of our proposed ideas, the QRM-LC-FD1 was used for combined detection and size estimation tasks, the QRM-LC-FD4 was used for combined detection and contrast estimation tasks, and the MTIA CCT 189 phantom was used for a combination of detection, size and contrast estimation tasks. The properties of LC objects in these four phantoms are listed in Tables 1–4. Only axial scans were used in this work. The slice thickness was 0.625 mm and the collimator aperture used was 20 mm. The x-ray voltage was kept at 120 kVp and current range was tuned to ensure that the task being performed was neither too difficult nor too easy.

Various radiation dose protocols were used and for each dose level, 50 repeat scans were acquired. The image reconstruction algorithm used in this study was filtered backprojection (FBP). From each of these scans, 10 individual realizations of signal and noise ROI were extracted from each scan for each different low-contrast signal. Thus, there were 500 individual image pairs for each LC object at every dose level. The images were reconstructed at a field of view of 180 mm with a matrix size of 512×512 image pixels. ROIs with a size of 100×100 pixels containing the LC object under study in the center (signal-present case) and the noise (signal-absent case) were extracted from the image slices, which is about 35 mm × 35 mm in size.

Table 1 MITA CCT183 phantom.

Object	Diameter (mm)	Contrast relative to background (HU)
1	3	14
2	5	7
3	7	5
4	10	3

Table 2 QRM-LC-FD1 phantom.

Object	Diameter (mm)	Contrast relative to background (HU)
1	3	−2
2	5	−2
3	7	−2
4	10	−2

Table 3 QRM-LC-FD4 phantom.

Object	Diameter (mm)	Contrast relative to background (HU)
1	5	−5
2	5	−10
3	5	−25
4	5	−50

Table 4 MITA CCT189 phantom.

Object	Diameter (mm)	Contrast relative to background (HU)
1	3	14
2	5	7
3	7	5
4	10	3
5	15	14
6	15	7
7	15	5
8	15	3

2.2 *Simulation Data Preparation, Analytical Solution, and Baseline*

Since the number of acquired images is always limited in experiments, simulation data were generated to find the baseline performance and help understand the variability in our methods. Because the two key components, first- and second-order statistics, needed in our observer models are the same for pure detection tasks and combined detection and estimation tasks, without loss of generality, our simulation was focused on the case of pure detection task. Simulated image pairs for signal-absent and signal-present classes are given by

$$\text{signal} - \text{absent image}\,\mathbf{g}_1 = \mathbf{n} + \mathbf{b} = \mathbf{HW} + \mathbf{b}, \tag{1}$$

$$\text{signal} - \text{present image}\,\mathbf{g}_2 = \mathbf{n} + \mathbf{b} + \mathbf{s} = \mathbf{HW} + \mathbf{b} + \mathbf{s}, \tag{2}$$

where subscript index 1 and 2 denote signal-absent and signal-present, respectively. The signal $\mathbf{s}$ is circular LC objects with blurred edges. The term $\mathbf{b}$ is the phantom background and was estimated using real signal-absent images. The noise structure $\mathbf{n}$ was simulated using correlated Gaussian noise. The convolution of a blurring filter with white Gaussian noise can be represented with matrix operator $\mathbf{H}$ and a vector $\mathbf{W}$, respectively. The matrix $\mathbf{H}$ can be generated from the autocorrelation function $\mathbf{R_G}$ that was estimated using real images. $\mathbf{R_G}$ is given by

$$\mathbf{R_G} = \frac{1}{N}\sum_{n=1}^{N} \mathbf{g}_{1,n}^{\text{real}} \mathbf{g}_{1,n}^{\text{real}} = \langle \mathbf{g}_1 \mathbf{g}_1^t \rangle$$
$$= \langle (\mathbf{HW} + \mathbf{b})(\mathbf{HW} + \mathbf{b})^t \rangle = \sigma^2 \mathbf{H}\mathbf{H}^t + \mathbf{b}\mathbf{b}^t, \tag{3}$$

where superscript real signifies real image data, t denotes the transpose operation, and σ is the standard deviation estimated from real image data. $\mathbf{R_G}$, σ, and $\mathbf{b}$ are all known and dose level dependent. The blurring kernel $\mathbf{H}$ is calculated by taking the square root of $\frac{1}{\sigma^2}(\mathbf{R_G} - \mathbf{b}\mathbf{b}^t)$.

The figure-of-merit (FOM) of this study is the area under the ROC curve (AUC). The analytical solution of AUC can be calculated through the noise covariance matrix.[15] The noise covariance $\mathbf{K_n}$ can be determined via

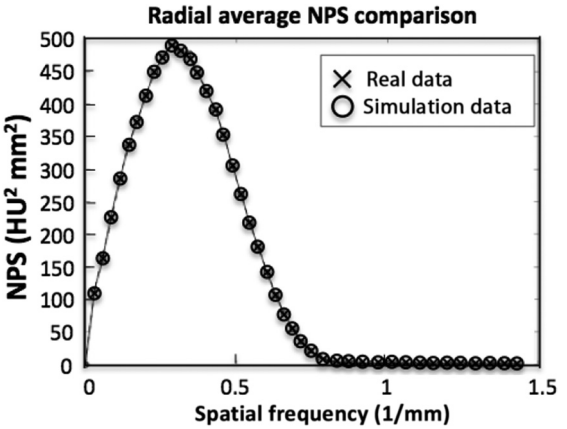

Fig. 1 Radial average NPS profile at 120 kVp 80 mAs: simulation data (black-circle solid line) and real data (black-dashed cross line).

$$\mathbf{K_n} = \langle \mathbf{n}\mathbf{n}^t \rangle = \langle (\mathbf{HW})(\mathbf{HW})^t \rangle = \langle \mathbf{HWW}^t\mathbf{H}^t \rangle = \sigma^2 \mathbf{HIH}^t = \sigma^2 \mathbf{HH}^t = \mathbf{R_G} - \mathbf{bb^t}. \quad (4)$$

The AUC can be approximated through the SNR of the test statistics using

$$\mathrm{SNR} = \sqrt{\mathbf{s}^t \mathbf{K_n^{-1}} \mathbf{s}} = \sqrt{\mathbf{s}^t (\mathbf{R_G} - \mathbf{bb^t})^{-1} \mathbf{s}}, \quad (5)$$

$$\mathrm{AUC} = \frac{1}{2}\left[1 + \mathrm{Erf}\left(\frac{\mathrm{SNR}}{2}\right)\right], \quad (6)$$

where SNR is the signal-to-noise ratio, and Erf is the error function. Since $\mathbf{R_G}$ is a function of dose level so the SNR and AUC are also functions of dose level.

To validate the simulated images, the radial average noise power spectrum (NPS)[16] profile of the simulation and the real images were compared. The comparison (Fig. 1) indicated that the simulated noise background is very similar to the real noise background. An example of a simulated image and a real image is shown in Fig. 2. The analytically derived AUC values (solid line) were compared with the AUC values calculated using the CHO method from 20,000 pairs of simulation images (dashed line in Fig. 3), and the comparisons demonstrated that the above derivation of the analytical solution is reasonable.

To better understand any biases present while using a limited number of real data, we compared the AUC values generated using finite numbers of simulated images with the AUC values generated by the analytical solution and 20,000 pairs of simulated images. We found that the AUC values of CHO using 240 training images and 260 testing images (solid line with triangle)

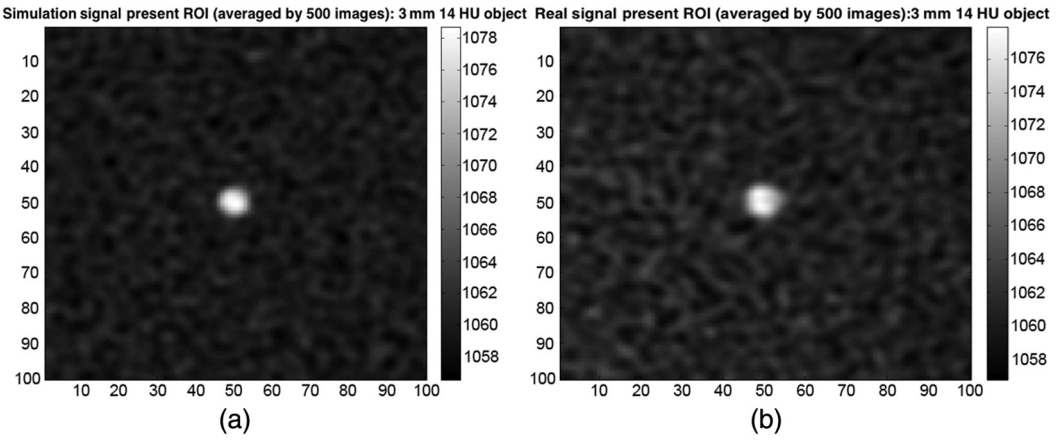

Fig. 2 Signal-present image averaged by 500 images at 120 kVp 80 mAs: (a) simulation image and (b) real image.

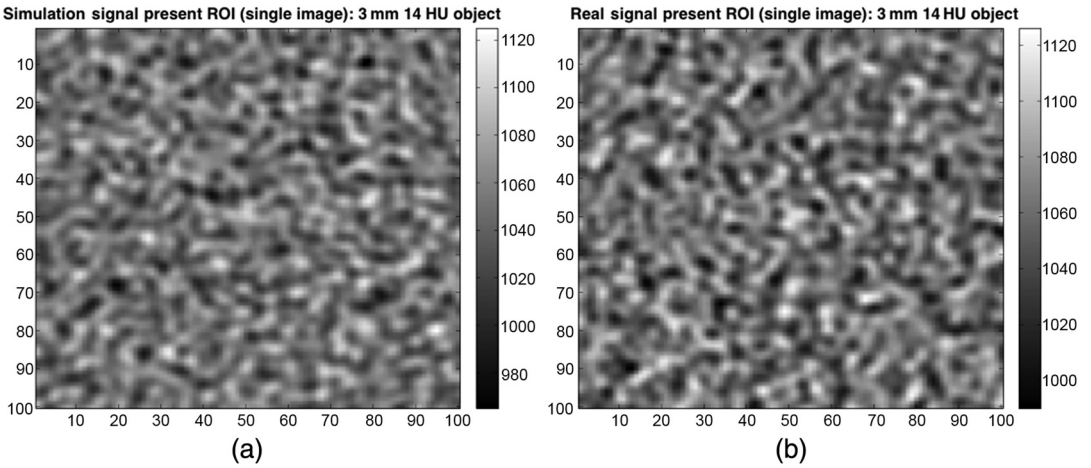

Fig. 3 Signal-absent image at 120 kVp 80 mAs: (a) simulation image and (b) real image.

were very close to both the analytical solution (solid line) and the 20,000 image pairs CHO (dashed line) for all four objects with accuracy ranging from 98.25% to 99.87% (Fig. 4). This high accuracy in the simulation study indicates that 500 real image pairs will result in negligible bias and can be used as the baseline for the rest of our experimental study.

2.3 *Traditional Approach*

2.3.1 *Channelized Hotelling observer*

The CHO approach requires a relatively large amount of training data to estimate the means and covariance matrices to generate accurate estimates of task performance. The image template calculated by the CHO is given by

$$\mathbf{W}_{\mathrm{CHO}}^{t} = \bar{\mathbf{s}}^{t}\mathbf{S}^{-1}, \tag{7}$$

where $\bar{\mathbf{s}}$ is the sample estimated mean signal and $\mathbf{S}$ is the common covariance matrix defined by averaging the sample covariance matrices of the two classes. They are computed by the following equations:

$$\mathbf{x}_i = \mathbf{T}^{\mathbf{t}}\mathbf{g}_i, \tag{8}$$

$$\mathbf{m}_i = \frac{1}{N_i}\sum_{k=1}^{N_i}\mathbf{x}_{i,k}, \tag{9}$$

$$\bar{\mathbf{s}} = (\mathbf{m}_2 - \mathbf{m}_1), \tag{10}$$

$$\mathbf{\Sigma}_i = \frac{1}{N_i - 1}\sum_{k=1}^{N_i}(\mathbf{x}_{i,k} - \mathbf{m}_i)(\mathbf{x}_{i,k} - \mathbf{m}_i)^t, \tag{11}$$

$$\mathbf{S} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{\Sigma}_i, \tag{12}$$

where subscript index $i = 2{,}1$ refers to signal-present and signal-absent classes, respectively, and $\mathbf{m}_i$ is the mean of channelized data set $\mathbf{x}_i$ for i'th class. In this paper, $\mathbf{T}$ is the 10 DDOG channels proposed by Abbey and Barrett,[17] $\mathbf{g}_i$ is the image vector before channelization, $\mathbf{\Sigma}_i$ is the sample covariance matrix for the ith class, and L is the number of classes. Since there are only two classes in the pure detection task, $L = 2$ for the detection task.

2.3.2 *Channelized scanning-linear observer*

Similar to the CHO, the first- and second-order statistics need to be calculated before generating the FOM for the CSLO. The concept of scanning-linear estimation[13] involves the calculation of the mode of the posterior density or to approximate maximum *a posteriori* (MAP) estimation. This calculation requires a scan over parameter space to compare solutions and find the maximum. The general formula of MAP estimation is

$$\hat{\boldsymbol{\theta}}_{\mathrm{MAP}} = \mathrm{argmax}_{\theta}[\mathrm{pr}(\boldsymbol{\theta}|\mathbf{g})] \quad = \mathrm{argmax}_{\theta}\left[\frac{\mathrm{pr}(\mathbf{g}|\boldsymbol{\theta})\mathrm{pr}(\boldsymbol{\theta})}{\mathrm{pr}(\mathbf{g})}\right], \tag{13}$$

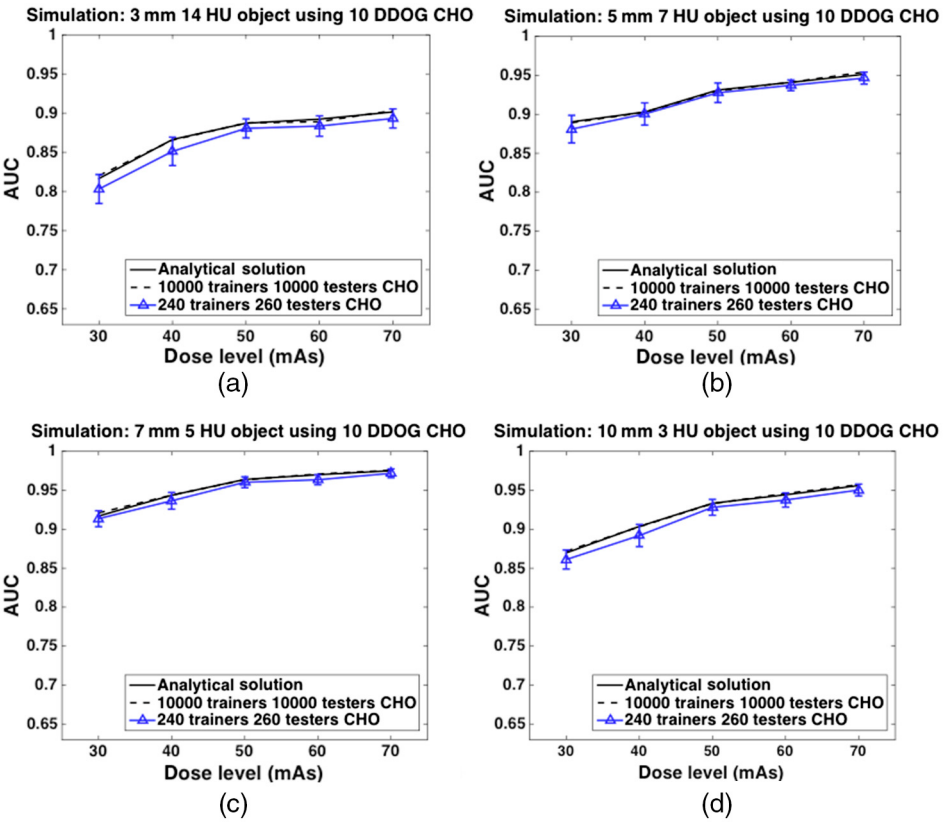

Fig. 4 Comparisons of AUC values of analytical solutions versus CHO using 20,000 images versus CHO using 500 images. (a) Object 1: 3 mm 14 HU, (b) object 2: 5 mm 7 HU, (c) object 3: 7 mm 5 HU, and (d) object 4: 10 mm 3 HU.

where the $\boldsymbol{\theta}$ and $\hat{\boldsymbol{\theta}}$ are the parameters we are interested in and their estimated parameters, respectively, $\mathrm{pr}(\mathbf{g}|\boldsymbol{\theta})$ is the likelihood of data conditioned on parameters, $\mathrm{pr}(\boldsymbol{\theta})$ and $\mathrm{pr}(\mathbf{g})$ are the likelihood of parameter $\boldsymbol{\theta}$ and dataset $\mathbf{g}$. To easily optimize this function, we will consider the Gaussian likelihood as was done in Ref. 13. Note this approximation does not imply that joint pdf $\mathrm{pr}(\mathbf{g},\boldsymbol{\theta})$ is also Gaussian. Based on this approximation, the conditional likelihood $\mathrm{pr}(\mathbf{g}|\boldsymbol{\theta})$ can be described by

$$\mathrm{pr}(\mathbf{g}|\boldsymbol{\theta}) \cong \frac{1}{\sqrt{2\pi^M \det(\mathbf{K}_{\mathbf{g}|\boldsymbol{\theta}})}} \exp\left\{\frac{-1}{2}[\mathbf{g}-\bar{\mathbf{g}}(\boldsymbol{\theta})]^t \mathbf{K}_{\mathbf{g}|\boldsymbol{\theta}}^{-1}[\mathbf{g}-\bar{\mathbf{g}}(\boldsymbol{\theta})]\right\}, \tag{14}$$

where $\bar{\mathbf{g}}(\boldsymbol{\theta})$ is the mean image averaged over the parameters $\boldsymbol{\theta}$, $\mathbf{K}_{\mathbf{g}|\boldsymbol{\theta}}$ is the sample covariance matrix conditioned on the parameters, and $\det(\cdot)$ is the determinant of the matrix. Instead of evaluating the covariance matrix and its inverse for every parameter, our second approximation is to use the mean of $\mathbf{K}_{\mathbf{g}|\boldsymbol{\theta}}$ averaged over all $\boldsymbol{\theta}$. To avoid the exponential term of Eq. (14), the common strategy is to operate the natural logarithm on both sides and ignore the term independent of the parameter $\boldsymbol{\theta}$, which leads the second line of Eq. (13) to become

$$\ln[\mathrm{pr}(\mathbf{g}|\boldsymbol{\theta})] \cong \frac{-1}{2}[\mathbf{g}-\bar{\mathbf{g}}(\boldsymbol{\theta})]^t \bar{\mathbf{K}}_{\mathbf{g}}^{-1}[\mathbf{g}-\bar{\mathbf{g}}(\boldsymbol{\theta})] + \ln[\mathrm{pr}(\boldsymbol{\theta})]. \tag{15}$$

Thus, the scanning-linear estimator that maximizes the posterior density under these approximations is equivalent to

$$\hat{\boldsymbol{\theta}}_{\mathrm{SL}}(\mathbf{g}) = \mathrm{argmax}_\theta\{\bar{\mathbf{g}}(\boldsymbol{\theta})^t \bar{\mathbf{K}}_{\mathbf{g}}^{-1}\mathbf{g} - \frac{1}{2}\bar{\mathbf{g}}(\boldsymbol{\theta})^t \bar{\mathbf{K}}_{\mathbf{g}}^{-1}\bar{\mathbf{g}}(\boldsymbol{\theta}) + \ln[\mathrm{pr}(\boldsymbol{\theta})]\}. \tag{16}$$

The first term is a linear operation applied on the testing image data by $\bar{\mathbf{g}}(\boldsymbol{\theta})^t \bar{\mathbf{K}}_{\mathbf{g}}^{-1}$ from the training data set. The second term is a shifted term due to the different parameters $\boldsymbol{\theta}$. The third term can be ignored in our study because $\mathrm{pr}(\boldsymbol{\theta})$ is a constant. So the final equation becomes

$$\hat{\boldsymbol{\theta}}_{\mathrm{SL}}(\mathbf{g}) = \mathrm{argmax}_\theta[\bar{\mathbf{g}}(\boldsymbol{\theta})^t \bar{\mathbf{K}}_{\mathbf{g}}^{-1}\mathbf{g} - \frac{1}{2}\bar{\mathbf{g}}(\boldsymbol{\theta})^t \bar{\mathbf{K}}_{\mathbf{g}}^{-1}\bar{\mathbf{g}}(\boldsymbol{\theta})]. \tag{17}$$

This observer operates on the data linearly, even though, in general, the linear template is a nonlinear function of $\boldsymbol{\theta}$. In the estimation process, the observer seeks the value of $\boldsymbol{\theta}$ that will maximize this linear operation.

For the channelized scanning-linear observer (CSLO), all image data needs to go through channelization before using Eq. (17). The input to the CSLO is the channelized testing images of different unknown physical properties, $\mathbf{x}^{\mathrm{testing}}(\boldsymbol{\theta})$. The output of the CSLO in this study is the estimated parameters denoted by the vector $\hat{\boldsymbol{\theta}}_{\mathrm{CSL}}(\mathbf{x})$. The relation between the input and output, according to Eq. (17), then can be described as

$$\hat{\boldsymbol{\theta}}_{\mathrm{CSL}}(\mathbf{x}) = \mathrm{argmax}_{\boldsymbol{\theta}}[\mathbf{s}(\boldsymbol{\theta})^t \mathbf{S}^{-1}\mathbf{x}^{\mathrm{testing}}(\boldsymbol{\theta}) - \frac{1}{2}\mathbf{s}(\boldsymbol{\theta})^t \mathbf{S}^{-1}\mathbf{s}(\boldsymbol{\theta})], \tag{18}$$

where $\mathbf{s}(\boldsymbol{\theta})$ is $\mathbf{T} * \bar{\mathbf{g}}(\boldsymbol{\theta})$, $\mathbf{S}$ is $\mathbf{T}\bar{\mathbf{K}}_{\mathbf{g}}^{-1}\mathbf{T}^{\mathbf{t}}$ and $\mathbf{x}^{\mathrm{testing}}(\boldsymbol{\theta})$ is $\mathbf{T} * \mathbf{g}^{\mathrm{testing}}(\boldsymbol{\theta})$.

2.4 *Proposed Approach*

In this paper, we will develop methods for computing observer performance that rely on less real data and can thus be computed more readily. Specifically, we will study whether we can use 40 images from one high-dose acquisition to estimate the signal and 20 images at the appropriate dose to estimate the covariance matrix $\mathbf{S}$. In the following sections, the methods we used to estimate signal and covariance matrix will be illustrated.

2.4.1 *Estimated signal*

Unlike the traditional method, the proposed estimated signal in detection is given by

$$\bar{\mathbf{s}}_{\text{High-dose}} = (\mathbf{m}_{2\,\text{High-dose}} - \mathbf{m}_{1\,\text{High-dose}}), \tag{19}$$

where $\mathbf{m}_{2\,\text{High-dose}}$ and $\mathbf{m}_{1\,\text{High-dose}}$ are the mean of the channelized image data sets of signal-present and signal-absent cases from high-dose images, respectively. The calculation of $\mathbf{m}_{i\,\text{High-dose}}$ is the same as $\mathbf{m}_i$ in Eq. (9) but requires less data. For combined tasks, the estimated signal with fewer samples in CSLO is given by

$$\mathbf{s}_{\text{High-dose}}(\boldsymbol{\theta}) = \mathbf{m}(\boldsymbol{\theta}) = \frac{1}{N}\sum_{k=1}^{N} \mathbf{x}_{k\,\text{High-dose}}(\boldsymbol{\theta}), \tag{20}$$

where $\mathbf{x}_{k\,\text{High-dose}}$ is the channelized image data of signal-present image from high-dose images of class i.

2.4.2 *Estimated covariance matrix*

The covariance matrix was estimated by the LOOC method proposed by Hoffbeck and Landgrebe.[14] In pattern recognition, an important problem is the effect of limited training samples on classification performance because training data is not always abundant and is expensive. When the ratio of the number of training samples to the dimensionality of data becomes small, deterioration of classification happens because the parameter estimation becomes variable. The LOOC method developed by Hoffbeck and Landgrebe estimates the covariance matrix robustly when the training data are limited. The covariance matrix was estimated by the combination of a sample covariance matrix, a common covariance matrix, and the diagonal matrices associated with them. The equation of i'th covariance matrix proposed by Hoffbeck and Landgrebe is given by

$$\mathbf{C}_i(\alpha_i) = \alpha_{i1}\mathrm{diag}(\boldsymbol{\Sigma}_i) + \alpha_{i2}\boldsymbol{\Sigma}_i + \alpha_{i3}\mathbf{S} + \alpha_{i4}\mathrm{diag}(\mathbf{S}), \tag{21}$$

where elements of mixing parameters $\alpha_i = [\,\alpha_{i1} \quad \alpha_{i2} \quad \alpha_{i3} \quad \alpha_{i4}\,]$ are required to sum to unity

$$\sum_{j=1}^{4} \alpha_{ij} = 1. \tag{22}$$

The values of mixing parameters α_i is selected to maximize the average log Gaussian likelihood of leave-one-out samples, which is called leave-one-out likelihood (LOOL). It should be noted here that the values of mixing parameters α_i are different at different radiation dose levels. In general, the more samples of α_i, the better the improvement, i.e., less bias compared to baseline. Both 50 and 100 samples for α_i were investigated, and

similar performance could be found in both cases when using our proposed approach. Due to the sufficiently high accuracy provided by 100 samples, 100 random samples of α_i were used independently for each dose level in this study. The LOOL is computed as follows:

$$\mathrm{LOOL}_i(\alpha_i) = \frac{1}{N_i}\sum_{k=1}^{N_i} \ln\{f[\mathbf{x}_{i,k}|\mathbf{m}_{i/k}, \mathbf{C}_{i/k}(\alpha_i)]\}, \tag{23}$$

where subscript index i/k means that the quantity is computed without using k'th samples from class i. The function f is the Gaussian likelihood

$$f[\mathbf{x}_{i,k}|\mathbf{m}_{i/k}, \mathbf{C}_{i/k}(\alpha_i)] = \frac{1}{\sqrt{(2\pi)^P|\mathbf{C}_{i/k}|}} \times \exp\left[\frac{-1}{2}(\mathbf{x}_{i,k}-\mathbf{m}_{i/k})^t\mathbf{C}_{i/k}^{-1}(\mathbf{x}_{i,k}-\mathbf{m}_{i/k})\right], \tag{24}$$

where P is the number of pixels. In this study, P is equal to the number of channels, 10, because all the image vectors were channelized by 10 DDOG channels.[17] The associated leave-one-out quantities are then computed as follows:

$$\mathbf{m}_{i/k} = \frac{1}{N_i-1}\sum_{\substack{j=1\\ j\neq k}}^{N_i} \mathbf{x}_{i,j}, \tag{25}$$

$$\mathbf{\Sigma}_{i/k} = \frac{1}{N_i-2}\sum_{\substack{j=1\\ j\neq k}}^{N_i} (\mathbf{x}_{i,j}-\mathbf{m}_{i/k})(\mathbf{x}_{i,j}-\mathbf{m}_{i/k})^t, \tag{26}$$

$$\mathbf{S}_{i/k} = \left(\frac{1}{L}\sum_{\substack{j=1\\ j\neq i}}^{L}\mathbf{\Sigma}_j\right) + \frac{1}{L}\mathbf{\Sigma}_{i/k}, \tag{27}$$

$$\mathbf{C}_{i/k}(\alpha_i) = \alpha_{i1}\mathrm{diag}(\mathbf{\Sigma}_{i/k}) + \alpha_{i2}\mathbf{\Sigma}_{i/k} + \alpha_{i3}\mathbf{S}_{i/k} + \alpha_{i4}\mathrm{diag}(\mathbf{S}_{i/k}). \tag{28}$$

Once the LOOL $\mathbf{C}_i$ of each class is determined, the final covariance matrix used in our approach can be calculated by the average of all covariance matrices from all classes. The final covariance matrix can be written as

$$\mathbf{C} = \frac{1}{L}\sum_{i=1}^{L}\mathbf{C}_i. \tag{29}$$

We will use $\mathbf{C}$ as the estimated covariance matrix using the LOOL method for the rest of this paper.

2.5 *Figure-of-Merit and the Relevant Mean and Variance*

It is now widely accepted that image quality should be evaluated using task-based criteria such as the performance of an observer (model or human) performing a medically relevant task. However, because human observer studies are typically very costly and time consuming, the use of a mathematical model observer becomes a very attractive alternative. In this study, we focus on the application of data reduction on CHO and CSLO in CT imaging systems. The input to our observer models is channelized image $\mathbf{x}^{\mathrm{testing}}$. Note here that $\mathbf{x}^{\mathrm{testing}}$ could be either signal-present or signal-absent and if there is a signal present, the size and contrast of the signal may vary depending upon which phantom is used. The output of our observer models is the test statistics. For the detection task, the test statistics can be described as

$$\mathbf{t}_l = \mathbf{W}\mathbf{x}_l^{\mathrm{testing}}, \tag{30}$$

where $l = 1$ for signal-absent and $l = 2$ for signal-present case, respectively, and $\mathbf{W} = \mathcal{S}^{\mathbf{T}}\bar{\mathbf{K}}^{-1}$. For the combined task of detection and estimation, the test statistics can be expressed as

$$\mathbf{t}_l[\boldsymbol{\theta}_\mathbf{l} = \hat{\boldsymbol{\theta}}_{\mathrm{CSLO(HL-CSLO)},\mathbf{l}}], \tag{31}$$

where $\boldsymbol{\theta}$ is the physical property of signal. $\hat{\boldsymbol{\theta}}_{\mathrm{CSLO(HL-CSLO)},\mathbf{l}}$ is given by

$$\hat{\boldsymbol{\theta}}_{\mathrm{CSLO(HL-CSLO)},\mathbf{l}}(\mathbf{x}_l^{\mathrm{testing}}) = \mathrm{argmax}_{\boldsymbol{\theta}}[\boldsymbol{\mathcal{S}}(\boldsymbol{\theta})^T\bar{\mathbf{K}}^{-1}\mathbf{x}_l^{\mathrm{testing}} - \frac{1}{2}\boldsymbol{\mathcal{S}}(\boldsymbol{\theta})^{\mathbf{T}}\bar{\mathbf{K}}^{-1}\mathcal{S}(\boldsymbol{\theta})], \tag{32}$$

where the channelized signal $\boldsymbol{\mathcal{S}}$ and the channelized covariance $\mathbf{K}$ can be estimated by the traditional method and proposed method. They can be described as

$$\mathcal{S} = \begin{cases} \mathbf{s}_{\mathrm{full}} \text{ using 240 training samples} \\ \mathbf{s}_{\mathrm{less}} \text{ using 40 training samples} \\ \mathbf{s}_{\mathrm{High\text{-}dose,less}} \text{ using 40 high-dose training samples} \end{cases}, \tag{33}$$

$$\mathbf{K} = \begin{cases} \mathbf{S}_{\mathrm{full}} \text{ using 240 training samples} \\ \mathbf{S}_{\mathrm{less}} \text{ using 20 training samples} \\ \mathbf{C}_{\mathrm{less}} \text{ using 20 training samples} \end{cases}, \tag{34}$$

$(\mathbf{s}_{\mathrm{full}}, \mathbf{S}_{\mathrm{full}})$ case is the baseline as mentioned before. The $(\mathbf{s}_{\mathrm{less}}, \mathbf{S}_{\mathrm{less}})$ case is the traditional approach with less data and $(\mathbf{s}_{\mathrm{High\text{-}dose,less}}, \mathbf{C}_{\mathrm{less}})$ case is the proposed approach using less data. It should be noted that there are two other obvious methods for reducing the data requirements: using only $\mathbf{s}_{\mathrm{High\text{-}dose,less}}$ and conventional covariance estimates or using $\mathbf{C}_{\mathrm{less}}$ with a conventional estimate of the signal. However, we found that using either of these methods alone is suboptimal in terms of the performance compared with our proposed method and will not be presented here for brevity. To investigate the improvement of our proposed method with less data, we focused on the comparisons among these three cases, i.e., baseline, traditional approach, and the proposed approach using the high-dose signal and the LOOL-estimate of covariance.

ROC curve and EROC curve[18] can be generated after generating the test distributions $\mathbf{t}_l$. The FOM of CHO and CSLO is the area under ROC curve (AUC) and the area under EROC (EAUC) curve, respectively. Since the selection of mixing parameters α_i could depend upon the training data, the mean values and variances of the FOM for all tasks were estimated by the shuffle method.[19]

2.6 *Validation Using Simulation Data*

To verify our proposed method, three approaches were studied using both simulation and real-data study: (1) baseline-traditional approach using 240 training samples and 260 testing samples; (2) traditional approach using 20 training samples and 80 testing samples; and (3) proposed approach using 20 training samples and 80 testing samples. The baseline results represent, to our best approximation, the results we would obtain with nearly infinite data. Thus, the closer to the baseline, the higher the accuracy is. The accuracy is defined as the ratio of the mean of the AUC value estimated by the testing method to the mean value of the AUC estimated by baseline. Again, first- and second-order statistics, needed to be estimated, are the same for the pure detection task and the combined detection and estimation task. Without loss of generality, the validation will be focused on the case of the pure detection task using simulation data. The MITA CCT183 phantom was used for this verification study. The accuracy is in the range of 88% to 98% for the traditional method using fewer data in simulation. For our proposed method using fewer data, the accuracy is in the range of 97% to 100%. The comparisons have been shown in Fig. 5. The simulation suggested that our proposed method provides more accurate results than the traditional approach while the amount of data used in training and testing was reduced to 20% of the original amount of data.

3 Results

In our study, four tasks were investigated: (1) detection tasks, (2) detection and size estimation tasks, (3) detection and contrast estimation tasks, and (4) detection, size, and contrast estimation tasks. For all tasks in this study, the location of signal was known exactly and was always in the center of ROIs. For each task, three methods were considered: (i) the baseline, traditional method with full data, (ii) traditional method with less data, and (iii) proposed method (HL-CHO/HL-CSLO) with less data. Full data refers to using 240 images for training and 260 images for testing. Less data means that there are only 20 images for training and 80 images for testing. All the images were real CT images. The FOM for the pure detection task and the combination of detection and estimation tasks is AUC and EAUC value, respectively. The mean and variance of AUC or EAUC at each dose level are estimated by the shuffle method using 20 different shuffles of the data. The accuracy is defined by the ratio of mean value of the less data to the mean value of baseline. The x-ray current range was chosen such that the FOM value was neither close to 1 nor close to the guessing value. FOM value equal to 1 means that the observer can easily detect if the signal/tumor exists and distinguish the physical properties of signal/tumor. The guessing value will be described later in this paper.

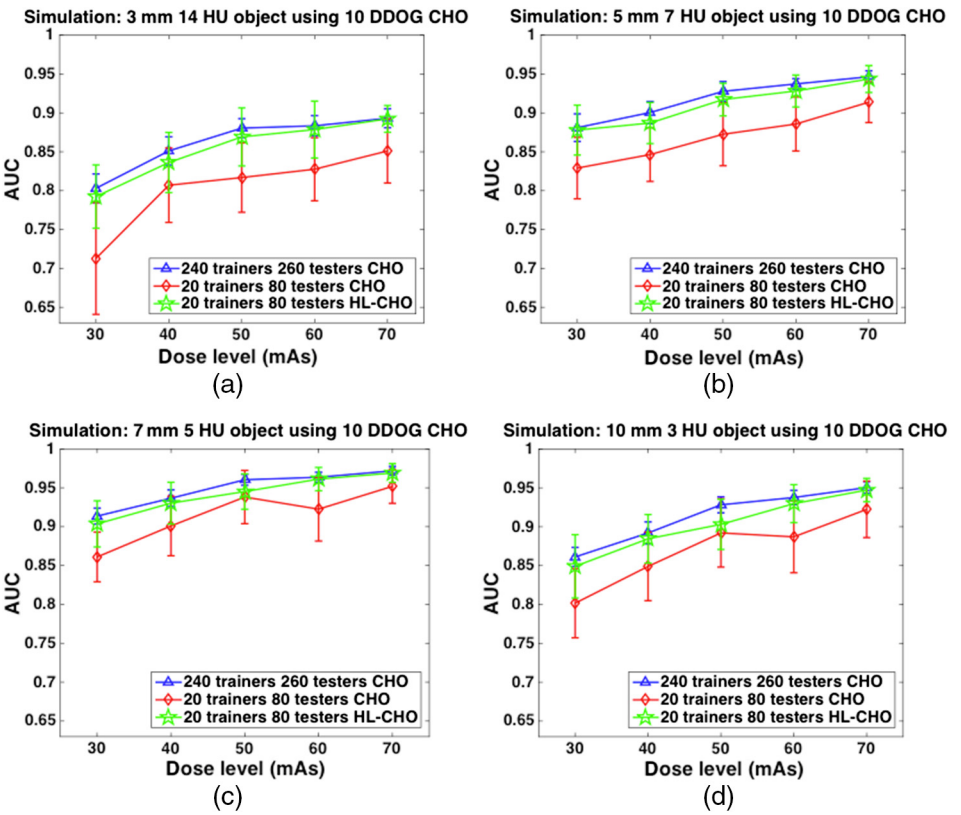

Fig. 5 Simulation study. Comparisons of AUC values for 500 image pairs CHO (baseline), 100 image pairs CHO (traditional approach), and 100 image pairs HL-CHO (proposed approach) from simulation image data sets. (a) Object 1: 3 mm 14 HU, (b) object 2: 5 mm 7 HU, (c) object 3: 7 mm 5 HU, and (d) object 4: 10 mm 3 HU.

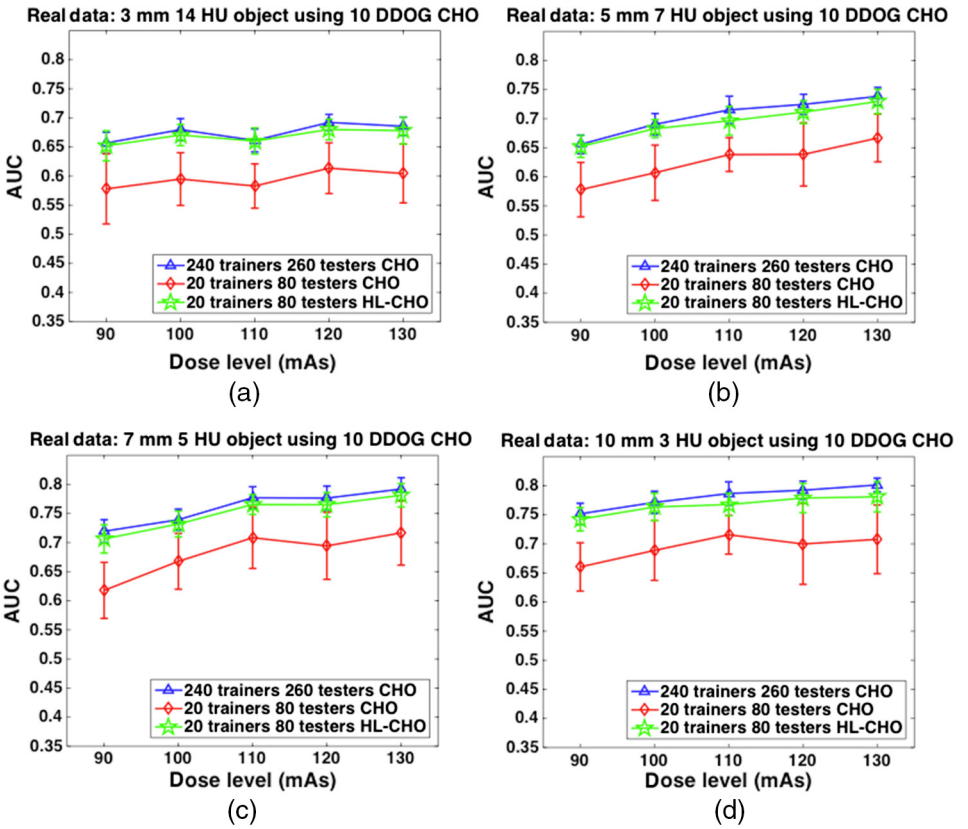

Fig. 6 Detection task. Comparisons of AUC values for 500 image pairs CHO (baseline), 100 image pairs CHO (traditional approach), and 100 image pairs HL-CHO (proposed approach) from real image data sets. (a) Object 1: 3 mm 14 HU, (b) object 2: 5 mm 7 HU, (c) object 3: 7 mm 5 HU, and (d) object 4: 10 mm 3 HU.

For the pure detection task, the MITA CCT 183 phantom was used. With ROC analysis, the guessing observer has an AUC of 0.5. The accuracy is 85% to 92% and 97% to 100% for CHO and HL-CHO with less data, respectively. The plots of comparisons are shown in Fig. 6. The results of the pure detection task indicated that our proposed approach can efficiently reduce the required data without losing accuracy.

For the detection and size estimation task, the QRM-LC-FD1 phantom was used for this task. The blind guessing value is $\frac{1}{2}\frac{1}{\text{the number of estimated physical properties}}$. Again the $\frac{1}{2}$ comes from detection by guessing. The number of estimate physical properties, in this case, is 4. So the blind guessing value is 0.125. In this task, the observer not only had to distinguish if the signal exists, but also estimate the size of signal. The results are shown in Fig. 7. In this task, accuracy is 84% to 88% and 97% to 99% for CSLO and HL-CSLO method, respectively. Based on the results of this study, the proposed approach needs only 20% of the full dataset to reach about the same performance of the traditional approach with the full dataset.

For the detection and contrast estimation tasks, the phantom QRM-LC-FD4 was used for this task. Similar to the task of combined detection and size estimation, there were also four objects of the same size with four different contrasts so the blind guessing value is also 0.125. The observer, in this task, had to tell if there is a signal in the testing image or not and determined the contrast of the signal. Results show that the accuracy is 92% to 96% for CSLO and 97% to 100% for HL-CSLO. The plot of EAUC versus radiation dose level is shown in Fig. 8. For the CSLO approach, the accuracy in the combined detection and contrast estimation tasks is better than the accuracy in the combined detection and size estimation tasks due to the contrasts of objects in QRM-LC-FD4 phantom being higher than the contrasts of objects in QRM-LC-FD1 phantom. This increases the accuracy in the lower dose range. Similar to the previous task, the proposed approach still maintains higher accuracy compared with the traditional approach.

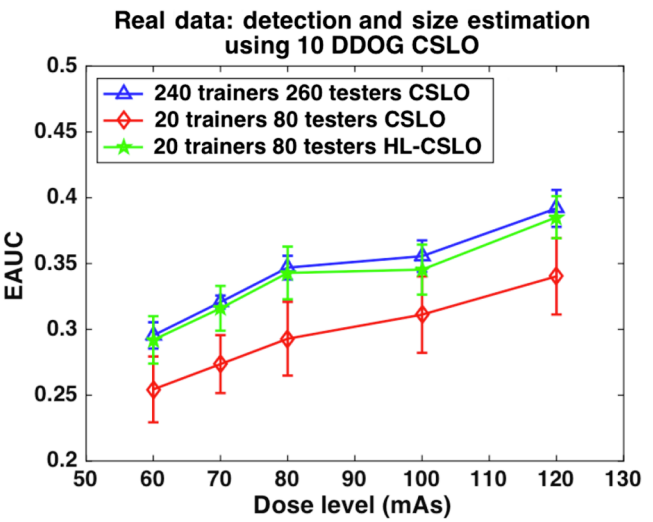

Fig. 7 Detection and size estimation tasks. Comparisons of EAUC values for 500 image pairs CSLO, 100 image pairs CSLO, and 100 image pairs HL-CSLO from real image data sets.

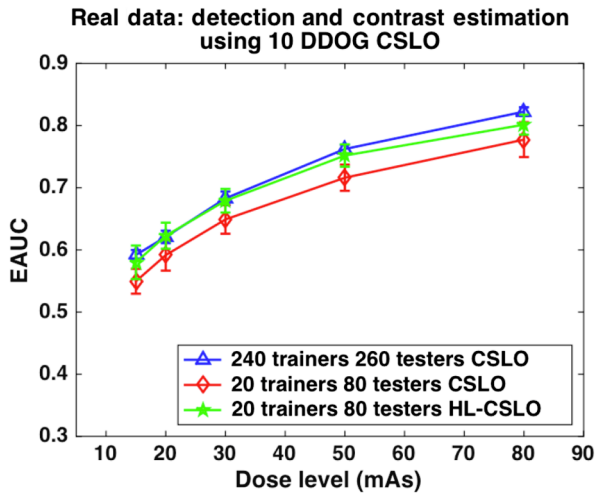

Fig. 8 Detection and contrast estimation tasks. Comparisons of EAUC values for 500 image pairs CSLO (baseline), 100 image pairs CSLO (traditional approach), and 100 image pairs HL-CSLO (proposed approach) from real image data sets.

In the study of combination of detection, size, and contrast estimation, the observer had not only to decide whether the signal is in the image or not but also make decisions on the size and the contrast of the detected signal. The MITA CCT 189 phantom was used. The blind guessing value in this complicated case is 0.0625 according to the guessing value equation discussed above and the number of estimate physical properties listed in Table 4. The accuracy of CSLO method is 88% to 91% and the accuracy of HL-CSLO is 94% to 99%. The results are shown in Fig. 9. For the CLSO approach with less data, the accuracy in this task is slightly higher than the combination task of the detection and size estimation. This does not imply that the CSLO approach with less data in the combination tasks of detection, size, and contrast estimation has better performance because the background of phantom (MITA CCT 189) used in this task was different. The proposed approach, HL-CSLO, again keeps high accuracy even in the low-radiation dose level range.

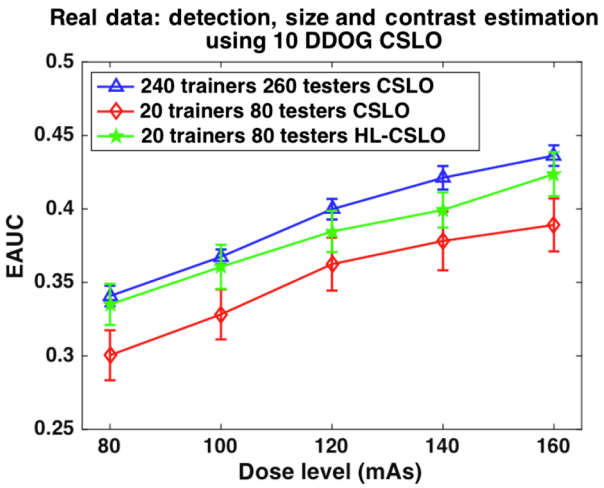

Fig. 9 Detection, size, and contrast estimation tasks. Comparisons of EAUC values for 500 image pairs CSLO (baseline), 100 image pairs CSLO (traditional approach), and 100 image pairs HL-CSLO (proposed approach) from real image data sets.

Table 5 Accuracy summary of 80% data reduction.

	Traditional method	Proposed method
Detection task	85.88% to 91.16%	97.38% to 99.86%
Detection and size estimation tasks	84.43% to 87.51%	97.13% to 98.85%
Detection and contrast estimation tasks	92.84% to 95.28%	97.46% to 99.69%
Detection, size and contrast estimation tasks	88.17% to 90.65%	94.80% to 98.33%

The accuracies for all the tasks performed by the traditional method and proposed method are listed in Table 5. All the studies were based on 80% data reduction. Overall, the traditional approaches, CHO and CSLO, both suffer while the data reduced especially in the case of lower radiation dose level. The proposed approaches, HL-CHO and HL-CSLO, have stable performance with high accuracy for all the radiation dose levels in all tasks.

4 Conclusions and Discussions

In our studies in which both simulated and real data were tried separately, we found that the proposed methods can match the baseline with high accuracy while using much less data. This is viewed as a positive result since we would ideally like our proposed approach to have the same performance as the baseline. Thus, instead of using 500 image pairs (240 training and 260 testing), the approach requires only 100 image pairs (20 training and 80 testing) for every dose level for all the tasks. This implies that instead of 50 scans for every radiation dose level, we need only 10 scans—a substantial reduction. The 40 images generated at high dose levels can be repeatedly used at different dose levels. The amount of image data required can be reduced up to 80% with our proposed method. Thus, the usage of model observers to assess image quality is a step closer to becoming a practical tool for routine QA/QC of CT imaging systems.

The high-dose images efficiently estimate the signal information with much less data because of the higher signal-to-noise ratio in the image data. Note this property is valid only for images from a linear reconstruction algorithm such as FBP. For iterative reconstruction images, the signal estimated by high dose will be an approximation because the processing is nonlinear and dependent upon the dose level. The sample covariance matrix is an unbiased estimator of the true covariance matrix when given large samples. In the limit of large samples, the sample covariance matrix approaches the true covariance. But if the number of samples is small, then a poor estimate of the covariance matrix occurs. In classification tasks, this results in over-fitting when the same data used to estimate the covariance are used to test the classifier. Over-fitting generally means that biases are introduced into the classification problem. Under-fitting implies that a classifier is too simple to match the real complexities in the data. Shrinkage, which is also called regularization, is a statistical scheme that combines a nominally overfit model with one that is manifestly underfit to produce a model with little bias.[20] In this study, we focus on the approach of LOOL. The channelized covariance matrices estimated by the process of maximizing the LOOL makes the channelized

covariance matrices more accurate. The covariance matrix estimated by LOOL mixes of the sample channelized covariance matrices and common channelized covariance matrices. This hybrid of matrices makes the inverse of matrix more stable with a sample size closer to the dimension of the channelized vector. The combination of channelized estimated signals from high-dose image and channelized-LOOL covariance matrices makes the AUC or EAUC values very close to the baseline.

Our results consistently indicated that the data can be reduced up to 80% if our proposed method is applied on the CHO and CSLO. Thus, for both pure detection and combination of detection and estimation tasks, the proposed methods, HL-CHO and HL-CSLO, can be used to quantify the image quality performance such as the dose saving capability and the low contrast detectability (LCD) performance of a CT system. It should be noted that the LOOL method was demonstrated on FBP images but does not rely on any of the assumptions that FBP uses. The simulation showed that our baseline was close to the true baseline and did rely on FBP-like assumption. This step was performed only for demonstration purposes and is not an integral part of the LOOL method. The limitation in this study is that the signal shape has to be circular symmetric due to the selection of the D-DOG channels. Thus, under controlled imaging conditions, HL-CHO and HL-CSLO can also be used to compare the image quality performance of different algorithms, components on one system. They can also be potentially used as a tool for the acceptance test for a newly developed iterative reconstruction method.

Acknowledgments

This project was supported by GE Healthcare, Waukesha, Wisconsin. The authors appreciate the technical discussions and support from members of the Advanced Sciences and Engineering team in GE Healthcare.

References

1. J. Yao and H. H. Barrett, "Predicting human performance by a channelized Hotelling observer model," *Proc. SPIE* **1786**, 161–168 (1992).
2. H. H. Barrett et al., "Model observers for assessment of image quality," *Proc. Natl. Acad. Sci. U. S. A.* **90**, 9758–9765 (1993).
3. C. K. Abbey, H. H. Barrett, and D. W. Wilson, "Observer signal-to-noise ratios for the ML-EM algorithm," *Proc. SPIE* **2712**, 47–58 (1996).
4. H. C. Gifford et al., "Channelized hotelling and human observer correlation for lesion detection in hepatic SPECT imaging," *J. Nucl. Med.* **41**(3), 514–521 (2000).
5. S. Young et al., "Estimating breast tomosynthesis performance in detection tasks with variable-background phantoms," *Proc. SPIE* **7258**, 725800 (2009).
6. L. Platisa et al., "Using channelized Hotelling observers to quantify temporal effects of medical liquid crystal display on detection performance," *Proc. SPIE* **7627**, 76270U (2010).
7. J. G. Brankov, "Optimization of the internal noise models for channelized Hotelling observer," in *2011 IEEE Int. Symp. on Biomedical Imaging: From Nano to Macro*, pp. 1788–1791 (2011).
8. S. Kulkarni et al., "A channelized Hotelling observer study of lesion detection in SPECT MAP reconstruction using anatomical priors," *Phys. Med. Biol.* **52**(12), 3601–3617 (2007).
9. L. Platisa et al., "Channelized Hotelling observers for the assessment of volumetric imaging data sets," *J. Opt. Soc. Am. A* **28**(6), 1145–1163 (2011).
10. 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).
11. H. Tseng et al., "Assessing image quality and dose reduction of a new x-ray computed tomography iterative reconstruction algorithm using model observers," *Med. Phys.* **41**(7), 071910 (2014).
12. H. Tseng, J. Fan, and M. A. Kupinski, "Combination of detection and estimation tasks using channelized scanning linear observer for CT imaging systems," *Proc. SPIE* **9416**, 94160H (2015).
13. M. K. Whitaker, E. Clarkson, and H. H. Barrett, "Estimating random signal parameters from noisy images with nuisance parameters: linear and scanning-linear methods," *Opt. Express* **16**(11), 8150–8173 (2008).
14. J. P. Hoffbeck and D. A. Landgrebe, "Covariance matrix estimation and classification with limited training data," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**(7), 763–767 (1996).
15. H. H. Barrett and K. Myers, *Foundations of Image Science*, Wiley, Hoboken, New Jersey (2003).
16. J. T. Dobbin et al., "Intercomparison of methods for image quality characterization. II. Noise power spectrum," *Med. Phys.* **33**(5), 1466–1475 (2006).
17. C. K. Abbey and H. H. Barrett, "Human-and-model-observer performance in ramp-spectrum noise: effects of regularization and object variability," *J. Opt. Soc. Am. A* **18**(3), 473–488 (2001).
18. E. Clarkson, "Estimation receiver operating characteristic curve and ideal observers for combined detection/estimation tasks," *J. Opt. Soc. Am. Opt. Image Sci. Vis.* **24**(12), B91–B98 (2007).
19. K. Fukunaga and R. R. Hayes, "Effects of sample size in classifier design," *IEEE Trans. PAMI* **11**, 873–885 (1989).
20. J. Theiler, "The incredible shrinking covariance estimator," *Proc. SPIE* **8391**, 83910P (2012).

Hsin-Wu Tseng received his BS degree in physics from National Tsing Hua University, Taiwan, his MS degree in physics from University of Florida, Gainesville, Florida, and his PhD in optical sciences and engineering from the University of Arizona, Tucson, Arizona. His research focuses on task-based image quality accessments, CT physics, and algorithm developments.

Jiahua Fan received his BS and MS degrees in electrical engineering from Tianjin University, China, and his PhD in electrical and computer engineering from University of Arizona, Tucson, Arizona. After working on medical-display and medical image processing as an assistant research scientist at University of Arizona, he joined GE Healthcare, Waukesha, Wisconsin. Currently, he is a principal scientist and works on image quality, physics and reconstruction topics of CT systems. He has over 70 scientific publications and 12 patents issued.

Matthew A. Kupinski is a professor at the University of Arizona in the College of Optical Sciences. He received his PhD in 2000 from the University of Chicago and joined the faculty at the University of Arizona in 2002. He is the recipient of the 2007 Mark Tetalman Award and has worked in diverse areas of imaging including x-ray, gamma-ray, diffuse optical, magnetic resonance, and neutron imaging.