Published in final edited form as:
IEEE Trans Radiat Plasma Med Sci. 2024 April ; 8(4): 439–450. doi:10.1109/trpms.2024.3379215.

DEMIST: A Deep-Learning-Based Detection-Task-Specific Denoising Approach for Myocardial Perfusion SPECT

Md Ashequr Rahman [Student Member, IEEE],
Department of Biomedical Engineering, Washington University, St. Louis, MO 63130 USA.

Zitong Yu [Student Member, IEEE],
Department of Biomedical Engineering, Washington University, St. Louis, MO 63130 USA.

Richard Laforest,
Mallinckrodt Institute of Radiology, Washington University, St. Louis, MO 63130 USA.

Craig K. Abbey,
Department of Psychological and Brain Sciences, University of California at Santa Barbara, Santa Barbara, CA 93106 USA.

Barry A. Siegel,
Mallinckrodt Institute of Radiology, Washington University, St. Louis, MO 63130 USA.

Abhinav K. Jha [Senior Member, IEEE]
Department of Biomedical Engineering and the Mallinckrodt Institute of Radiology, Washington University, St. Louis, MO 63130 USA

Abstract

There is an important need for methods to process myocardial perfusion imaging (MPI) single-photon emission computed tomography (SPECT) images acquired at lower-radiation dose and/or acquisition time such that the processed images improve observer performance on the clinical task of detecting perfusion defects compared to low-dose images. To address this need, we build upon concepts from model-observer theory and our understanding of the human visual system to propose a detection task-specific deep-learning-based approach for denoising MPI SPECT images (DEMIST). The approach, while performing denoising, is designed to preserve features that influence observer performance on detection tasks. We objectively evaluated DEMIST on the task of detecting perfusion defects using a retrospective study with anonymized clinical data in patients who underwent MPI studies across two scanners (N = 338). The evaluation was performed at low-dose levels of 6.25%, 12.5%, and 25% and using an anthropomorphic channelized Hotelling observer. Performance was quantified using area under the receiver operating characteristics curve (AUC). Images denoised with DEMIST yielded significantly higher AUC compared to

Corresponding author: Abhinav K. Jha. a.jha@wustl.edu.

All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that could have an influence or are relevant to the work reported in this article.

This article has supplementary material provided by the authors and color versions of one or more figures available at https://doi.org/10.1109/TRPMS.2024.3379215.

corresponding low-dose images and images denoised with a commonly used task-agnostic deep learning-based denoising method. Similar results were observed with stratified analysis based on patient sex and defect type. Additionally, DEMIST improved visual fidelity of the low-dose images as quantified using root mean squared error and structural similarity index metric. A mathematical analysis revealed that DEMIST preserved features that assist in detection tasks while improving the noise properties, resulting in improved observer performance. The results provide strong evidence for further clinical evaluation of DEMIST to denoise low-count images in MPI SPECT.

Keywords

Detection; image denoising; myocardial perfusion imaging (MPI); objective task-based evaluation; single-photon emission computed tomography (SPECT)

I. Introduction

Single-photon emission computed tomography (SPECT) myocardial perfusion imaging (MPI) has an established and well-validated role in evaluating patients with known or suspected coronary artery disease [1]. For diagnosis of this disease, the clinical task performed on MPI-SPECT images is the detection of focally reduced tracer uptake (perfusion defects) reflecting reduced blood flow in the myocardial wall. Typically, in clinical MPI-SPECT protocols, patients are administered a radiopharmaceutical tracer, such as Tc-99m sestamibi or Tc-99m tetrofosmin, under stress and rest conditions. For a protocol involving a Tc-99m radiopharmaceutical with rest and stress imaging performed on a single day, the administered activity can be as high as 48 mCi [2]. Thus, developing protocols to reduce this administered dose are well poised for a strong clinical impact [3], [4]. Additionally, current MPI-SPECT acquisition protocols can take up to around 12–15 min, during which time, the patient is required to be stationary. This is a challenge, especially for older patients, which are a large fraction of the patient population [5]. Thus, methods to reduce acquisition time can make MPI-SPECT more comfortable for patients, less susceptible to patient motion, and can also lead to increased clinical throughput and reduced cost of imaging. However, reducing this dose and/or acquisition time results in a lower number of detected counts in the projection data, which, when reconstructed, yields images with deteriorated image quality in terms of the ability to reliably detect perfusion defects. Thus, there is an important need to develop methods to process low-count MPI-SPECT images for improved performance on detection tasks.

In recent years, deep learning (DL)-based methods have shown promise in processing MPI-SPECT images [6], [7], [8], [9], [10], [11], [12], [13], [14], [15], particularly in image denoising for predicting normal-dose images from low-dose images [12], [13], [14], [15]. Typically, these denoising approaches are trained by minimizing a loss function based on image fidelity, such as pixel-wise mean squared error (MSE), between the actual normal-dose image and low-dose image predicted by the deep network. These methods have usually been evaluated with fidelity-based metrics, such as root MSE (RMSE) and structural similarity index metric (SSIM), where the results have indicated that the methods provide improved performance compared to low-dose images. However, it is well recognized that for clinical translation, DL-based denoising methods need to be evaluated on performance in clinically relevant tasks [16], [17], [18], [19]. At an early stage of translation, model observers provide a mechanism to perform such evaluation [20]. However, of the various DL-based denoising methods proposed for MPI SPECT, those that have been evaluated on the clinical task of detecting perfusion defects have not shown improved performance [19], [21]. Recent studies in other imaging modalities have also yielded similar findings [22], [23]. While it is well recognized that any image-processing method cannot improve the performance of ideal observers due to data-processing inequality [18], for suboptimal observers, such as human observers, improving detection performance may be possible. Further, the detection task on MPI-SPECT images is clinically performed by human observers. Thus, in this manuscript, we investigate the development of denoising methods that explicitly demonstrate improved performance on the task of detecting perfusion defects in MPI-SPECT images with an anthropomorphic model observer that has been shown to emulate human observer performance on this task [24], [25].

To investigate the limited performance of DL-based denoising methods on detection tasks in MPI SPECT, Yu et al. [21] conducted a mathematical analysis with a commonly used DL-based denoising method that used pixel-wise MSE as the loss function. They analyzed the detection performance of a numerical observer that has been observed to emulate human-observer performance in MPI SPECT. Their analysis revealed that the method was improving the noise characteristics of the images, which, in isolation, would have improved observer performance. However, the analysis also revealed that the method was discarding features used to perform the detection task, which eventually translated to no improvement in observer performance. These observations indicate that a denoising method that can preserve detection-task-specific features may improve observer performance on detection tasks. Recently, in the context of X-ray CT, a few DL-based denoising methods have been proposed with the aim of preserving features that assist in the detection task [26], [27], [28]. These methods typically incorporate a hybrid loss consisting of image fidelity and task-specific terms, where the latter term has been incorporated in the form of signal-to-noise ratio (SNR) [26], binary cross-entropy loss associated with a DL-based observer [27], and perceptual loss obtained from features extracted by a pretrained Visual Geometry Group network [28]. Results from these studies support the idea that preserving task-specific features may assist with improving performance on detection task. However, the methods proposed have limitations to the applicability to the SPECT denoising problem, such as assuming 2-D images, defect-known-exactly setups, use of ground-truth phantom as the target/label, and limited interpretability of the task-specific loss term. Additionally, the methods have been evaluated using stylized studies. For clinical applicability, evaluation of such methods with clinical data and on clinically relevant tasks is needed.

Motivated by these observations from prior studies, we propose a DL-based task-specific denoising method for 3-D MPI SPECT. The method builds upon concepts from the literature on model observers and our understanding of the human visual system to preserve detection-task-specific features while performing denoising. We objectively evaluate the proposed method on the task of detecting perfusion defects using a retrospective study with anonymized clinical MPI-SPECT data. Additionally, we evaluate the effect of population characteristics, including patient sex and perfusion defect types, on the detection-task performance. Preliminary results of this work have been presented previously [29].

II. Proposed Task-Specific Denoising Method

A. Theory

1) Problem Formulation: We propose the method in the context of reducing radiation dose in MPI SPECT, although the methodology can be applied in the context of reducing acquisition time. This is because reducing either dose or acquisition time eventually leads to a reduction in detected counts and the underlying objective of the proposed method is to denoise the low-count images.

Consider a SPECT system imaging a tracer distribution (object) within the human body, described by a vector $f(\boldsymbol{r})$, where $\boldsymbol{r} \in \mathbb{R}^3$ denotes the 3-D coordinates, and yielding projection data, denoted by the M-dimensional vector $\boldsymbol{g}$. Consider that the object and projection data lie in the Hilbert space $\mathbb{L}_2(\mathbb{R}^3)$ and the M-dimensional Euclidean space $\mathbb{E}^{\boldsymbol{M}}$, respectively. Here, $\mathbb{L}_2(\mathbb{R}^3)$. enotes the space of all square-integrable functions on $\mathbb{R}^3$. Thus, the SPECT system operator $\mathscr{H}$ maps object in $\mathbb{L}_2(\mathbb{R}^3)$ to projection data in $\mathbb{E}^{\boldsymbol{M}}$. The Poisson-distributed system-measurement noise is denoted by the M-dimensional vector $\boldsymbol{n}$. The images are then reconstructed using the reconstruction operator, denoted by $\mathscr{R}$, yielding the reconstructed images, denoted by the N_{3D}-dimensional vector $\hat{\mathbf{f}}$, where the hat symbol denotes that this is a reconstruction of the object $\boldsymbol{f}$. Thus

$$\hat{\mathbf{f}} = \mathscr{R}\boldsymbol{g} = \mathscr{R}(\mathscr{H}\boldsymbol{f} + \boldsymbol{n}), \tag{1}$$

where, without loss of generalization, we refer to the object as infinite-dimensional vector $\boldsymbol{f}$ to model that the radiotracer distribution is a continuous function. From the reconstructed images, an observer performs the task of detecting perfusion defects. More precisely, the task is to classify the image into defect-absent (H_0) or defect-present (H_1) case. Denote the defect-absent object as $\boldsymbol{f}_b$ and the defect signal as $\boldsymbol{f}_s$. The two hypotheses for the defect-detection task are given by

$$H_0 : \hat{\mathbf{f}} = \mathscr{R}\boldsymbol{g} = \mathscr{R}(\mathscr{H}\boldsymbol{f}_b + \boldsymbol{n}) . \tag{2a}$$

$$H_1 : \hat{\mathbf{f}} = \mathscr{R}\boldsymbol{g} = \mathscr{R}(\mathscr{H}(\boldsymbol{f}_b + \boldsymbol{f}_s) + \boldsymbol{n}) . \tag{2b}$$

In MPI SPECT, the perfusion-defect signal is a cold signal, so $\boldsymbol{f}_s$ is negative-valued term.

In a low-dose protocol, the tracer uptake is lower compared to normal-dose protocols. Thus, the projection data, and the corresponding reconstructed images are noisier at low dose, impacting observer performance on the defect-detection task. Denote the reconstructed images at normal dose and low dose by $\hat{\mathbf{f}}_{\mathrm{ND}}$ and $\hat{\mathbf{f}}_{\mathrm{LD}}$, respectively. Our goal is to design a technique to denoise these low-dose images such that the denoised images yield improved performance on the defect-detection task.

2) Proposed DL-Based Task-Specific Denoising Method: We consider the use of DL to design this denoising technique. Consider a deep network parameterized by the parameter vector $\mathbf{\Theta}$, denote the denoising operator by $\mathscr{D}_{\mathbf{\Theta}}$ and the pred predicted normal-dose image as $\hat{\mathbf{f}}_{\mathrm{ND}}^{\mathrm{pred}}$, where the subscript ND refers to the target of the prediction. The denoising operation can be mathematically expressed as follows:

$$\hat{\mathbf{f}}_{\mathrm{ND}}^{\mathrm{pred}} = \mathscr{D}_{\mathbf{\Theta}}\left(\hat{\mathbf{f}}_{\mathrm{LD}}\right). \tag{3}$$

To preserve the features that assist in the detection task while denoising, we propose a hybrid loss function for this deep network that consists of two terms. The first term penalizes the error associated with image fidelity between the actual and predicted normal-dose images. The second term penalizes the loss of features required to perform detection task in the predicted normal-dose images. Denote the fidelity loss term as $\mathscr{L}_{\mathrm{fid}}(\mathbf{\Theta})$ and the task-specific loss term as $\mathscr{L}_{\mathrm{task}}(\mathbf{\Theta})$. The hybrid loss function $\mathscr{L}(\mathbf{\Theta})$ is given by

$$\mathscr{L}(\mathbf{\Theta}) = \mathscr{L}_{\mathrm{fid}}(\mathbf{\Theta}) + \lambda \mathscr{L}_{\mathrm{task}}(\mathbf{\Theta}), \tag{4}$$

where λ denotes a hyperparameter that controls the weights of these loss functions.

Denote the total number of training samples as J and the jth sample of the low-dose image and normal-dose image by $N_{3\mathrm{D}}$-dimensional vectors $\hat{\mathbf{f}}_{\mathrm{LD}}^{j}$ and $\hat{\mathbf{f}}_{\mathrm{ND}}^{j}$, respectively. Also, denote the normal-dose image predicted by the denoising network as $\hat{\mathbf{f}}_{\mathrm{ND}}^{\mathrm{pred},\, j}$ when the low-dose image $\hat{\mathbf{f}}_{\mathrm{LD}}^{j}$ is given as the input to the network. Thus,

$$\hat{\mathbf{f}}_{\mathrm{ND}}^{\mathrm{pred},\, j} = \mathscr{D}_{\mathbf{\Theta}}\left(\hat{\mathbf{f}}_{\mathrm{LD}}^{j}\right). \tag{5}$$

A typical choice to measure the fidelity between the actual and predicted normal-dose images, including in MPI SPECT, is the MSE between these images [30], [31]. Thus, we chose this distance measure as our fidelity-loss term. Consider that we have J patient images in our training set. Denote the number of voxels in each image slice by $N_{2\mathrm{D}}$ and the number of slices as Z, so that $N_{3\mathrm{D}} = N_{2\mathrm{D}} Z$. Then, the fidelity-loss term is given by

$$\mathscr{L}_{\text{fid}}(\boldsymbol{\Theta}) = \frac{1}{JN_{\text{2D}}Z}\sum_{j=1}^{J}\left|\left|\hat{\mathbf{f}}_{\text{ND}}^{j} - \hat{\mathbf{f}}_{\text{ND}}^{\text{pred},j}\right|\right|_2^2. \tag{6}$$

To obtain an expression for the task-specific loss term $\mathscr{L}_{\text{task}}(\boldsymbol{\Theta})$ in (4), we recognize that the detection task on MPI-SPECT images is performed by human observers. Thus, a mathematical term that preserves features used by human observers while performing detection tasks will intuitively assist in improving performance on the detection task. In this context, there is substantial literature on mathematical model observers that emulate human-observer performance [32], [33], [34], [35], [36]. Further, multiple experiments in human vision have shown that the human visual system processes data using frequency-selective channels [18]. By processing the features extracted from these channels, referred to as anthropomorphic channels, studies have shown that model observers can mimic human-observer performance [32], [37], [38]. Of most relevance to this article, this has also been validated in studies with MPI SPECT on the task of detecting perfusion defects [24], [25]. Thus, a denoising technique that preserves features extracted by these channels may assist with improving observer performance on detection tasks.

Motivated by these studies, we design the task-specific loss term to preserve features that are derived by applying these anthropomorphic channels to the images. Typically, these channels are applied to the 2-D image slices. Thus, first, the profiles of the channels are centered on the defect location and the inner product of the channels and the to-be-processed 2-D image slices are computed to yield the feature value. Mathematically, denote an image slice by the N_{2D}-dimensional vector $\hat{\mathbf{f}}_{2D}$, denote the number of channels by $\boldsymbol{C}$ and the N_{2D}-dimensional column vector corresponding to the c^{th} channel by $\hat{\boldsymbol{u}}_c$. By concatenating the $\boldsymbol{C}$. chanl vectors, we obtain an $N_{\text{2D}} \times \boldsymbol{C}$ matrix $\boldsymbol{U}$. Denote the shift operator that centers the channel profiles to the signal location by $\mathscr{S}$. The shift operation on $\boldsymbol{U}$ can be represented by a multiplication of shift matrix $\boldsymbol{S}$ with the channel matrix $\boldsymbol{U}$. The application of the shifted channel matrix on the centered image slice yields a $\boldsymbol{C}$-dimensional vector, referred to as the channel vector and denoted by $\boldsymbol{v}$

$$\boldsymbol{v} = (\boldsymbol{SU})^T\hat{\mathbf{f}}_{\text{2D}}. \tag{7}$$

The task-specific loss term $\mathscr{L}_{\text{task}}(\boldsymbol{\Theta})$ penalizes the MSE between the channel vectors of the actual and predicted normal-dose image. To obtain the channel vector for the jth patient sample, we first perform acyclic 2-D shifting for each channel so that the center of the channel profile and centroid of the defect coincide. Sce different patients will have defects at different locations, denote the shift matrix for the jth patient as $\boldsymbol{S}^j$. Also, denote the sth slice of the normal-dose image $\hat{\mathbf{f}}_{\text{ND}}^{j}$ and the predicted normal-dose image $\hat{\mathbf{f}}_{\text{ND}}^{\text{pred},j}$ by $\hat{\mathbf{f}}_{\text{ND},2D,s}^{j}$ and $\hat{\mathbf{f}}_{\text{ND},2D,s}^{\text{pred},j}$, respectively. The task-specific loss term $\mathscr{L}_{\text{task}}(\boldsymbol{\Theta})$ is then given by

$$\mathscr{L}_{\text{task}}(\boldsymbol{\Theta}) = \frac{1}{JC(s_2 - s_1 + 1)} \times \sum_{j=1}^{J} \sum_{s=s_1}^{s_2} \left\| \left(\boldsymbol{S}^j \boldsymbol{U}\right)^{\mathrm{T}} \left(\hat{\mathbf{f}}_{\mathrm{ND},2D,s}^{j} - \hat{\mathbf{f}}_{\mathrm{ND},2D,s}^{\mathrm{pred}}\right) \right\|_2^2, \tag{8}$$

where s_1 and s_2 denote the index of the start and end slices where the channels are applied, respectively.

B. Implementation

We developed an encoder–decoder architecture to minimize the loss function given by (4). The encoder–decoder architecture with multiple resolution levels was chosen motivated by the architectures previously proposed for denoising low-dose MPI-SPECT images [21], [30]. The schematic of the architecture is shown in Fig. 1. The details of the network architecture are provided in the Supplementary material and in Rahman et al. [39]. The input and output to the network are the low-dose short-axis volume, $\hat{\mathbf{f}}_{\mathrm{LD}}$ and the denoised (predicted normal-dose) short-axis volume, $\hat{\mathbf{f}}_{\mathrm{ND}}^{\mathrm{pred}}$, respectively. The encoder extracts local spatial features from the low-dose image and generates a set of lower-dimensional latent features, which are used to reconstruct the denoised low-dose volume. Skip connections were used to add features learned in the encoder to the features generated by the decoder. Dropout was used to prevent overfitting. We trained the network by minimizing the hybrid loss in (4) using the ADAM algorithm [40].

III. Evaluation

We objectively evaluated the proposed method in an Institutional Review Board (IRB)-approved retrospective study conducted on clinical MPI-SPECT studies. We followed best practices for the evaluation of AI algorithms in nuclear medicine (RELIANCE guidelines) [41].

A. Data Collection and Curation

We collected data from MPI studies ($N = 4118$) conducted at clinical normal-dose level at Washington University School of Medicine between January 2016 and January 2021. The clinical protocol was a one-day stress/rest protocol and the mean injected activity for the stress images was 10 mCi in patients weighing under 250 pounds and 12 mCi for those weighing over 250 pounds at normal-dose level. 1295 MPI studies contained the binned SPECT projection data and CT images along with patient sex and anonymized clinical reports. The access to projection data allowed us to simulate the low-dose acquisition using binomial sampling, following a similar approach as in [42], which preserved the Poisson distribution in the low-dose projections [18]. More specifically, to obtain the low-dose count in a projection bin, we conducted independent Bernoulli trials for accepting each of the n normal-dose counts in that projection bin with probability p, where p denotes the fraction corresponding to the low-dose level. Essentially, this is equivalent to sampling from a Binomial distribution $\mathscr{B}(n, p)$. For Binomial sampling, we used MATLAB’s default

Mersenne Twister algorithm for pseudo-random number generation. We considered low-dose levels of 25%, 12.5%, and 6.25%. In generating the low-dose levels, we assumed that the fractional myocardial tracer uptake is linearly related to the injected dose. Thus, the count levels in myocardial wall were 25%, 12.5%, and 6.25% of normal-dose count level. At these low-dose levels, the performance on detection task is significantly different than normal-dose images [21], and task performance is dominated by system noise compared to anatomic variability in patient populations [43]. Thus, choosing these dose levels provided a regime to study the efficacy of the proposed method in improving task performance over low-dose images.

For training and evaluation of the proposed method, both the knowledge of presence of defect and the defect centroid were needed. Although presence of defect could be read from the clinical reports, findings in these reports often suffer from reader variability. Moreover, the defect centroid is typically unavailable. To address this issue, we only used the normal (defect-absent) MPI studies ($N = 795$) and inserted synthetic defects using a defect-insertion approach described later (Section III-A1) to create the defect-present images. For defect insertion, segmentation of the left ventricle (LV) wall was needed, but this wall could not be segmented reliably for some cases. Also, in some other cases, the images contained artifactual (apparent) defects. In clinical practice, these artifactual defects are typically ruled out using other patient data, such as the rest scans, polar maps, and projection scans. However, in our observer study, only the stress images are used for the detection task. Thus, we excluded these two sets of cases ($N = 457$) and only used the remaining normal cases ($N = 338$). The datasets were from two scanners, namely, the GE Discovery 670 Pro NaI and the GE Discovery 670 CZT. These two systems have different detectors, namely, NaI and CZT, each of which have different energy and position resolutions (as listed in the supplementary material and in [39]). The data-collection process is illustrated in Fig. 2.

1) Defect Insertion Approach: To insert the defect, we first segmented the LV wall using the reoriented short-axis normal-dose image using SEGMENT software [44], [45]. From the centroid of the LV wall, a 2-D cone region with a specific extent was located. For anterior-wall defect, the cone region was between 80° and $(80 - \theta)°$ where θ denotes the defect extent and was assigned values of 30° and 60°. For determining the angles, the x-axis was assumed to be along the rows of the reoriented image and the origin was the centroid of the LV wall. For inferior-wall defect, the cone region was between −80° and $(-80 + \theta)°$. In the slice containing the LV centroid, the LV wall that lies inside this cone region was considered as the defect mask. The same cone region was used in adjacent slices to create the 3-D defect. A 42-mm defect in the long-axis direction (apex-to-base direction) was considered. We used the mean LV uptake as reference to define defects with specific severities. The defect signal, with specific severity and extent, was then subtracted from the reconstructed image of the defect-absent case to create an initial defect-present image. Next, to create the hybrid dataset with inserted defects, we employed a strategy similar to that proposed by Narayanan et al. [46]. Briefly, we used SIMIND, a well-validated Monte-Carlo simulation software [47], [48] to generate the intermediate projection data corresponding to the defect-absent image and the initial defect-present image. These intermediate projection data were then used to calculate a scale factor. The clinical projection data from defect-absent cases were scaled using this scale factor to create the final defect-present projection data. The scale factor was in general unity apart from the regions near the heart, and even there, for low-severity signals, the scale factor were close to one.

2) Reconstruction and Post-Processing: We used a clinical reconstruction protocol based on the ordered subset expectation maximization (OSEM) algorithm implemented with CASTOR [49] to reconstruct the normal-dose and low-dose images. The reconstruction compensated for attenuation and collimator-detector response. Scatter compensation was not performed. The number of subsets and the iterations in the OSEM algorithm was selected based on the protocol used in the clinic. 3-D Butterworth filtering with filter order of 5 and cutoff frequency of 0.44 cycles/cm was applied to the low-dose and normal-dose images, which were then reoriented to the short axis using linear interpolation. From this reoriented image, we extracted a 48 × 48 × 48 volume where the center of the volume coincided with the center of LV. For better-dynamic range, we set the range of the pixel values to $[0, x_{\mathrm{LV}}]$ where x_{LV} is the maximum value inside the LV wall.

B. Network Training

The training set consisted of 2944 cases. These were obtained from 184 normal MPI studies. A total of 12 synthetic defect types were generated for each normal study, where the defect types were defined in terms of their extent, severity, and position in the LV wall. The defects were inserted in the anterior and inferior walls, had extents of 30° and 60°, and severities of 10%, 17.5%, and 25%. We inserted these 12 defect types in each of the 184 normal studies to generate the defect-present population ($N = 2208$). The defect-absent population was obtained by replicating the 184 normal studies a total of four times, corresponding to the four different defect extents and locations. Thus, the defect-absent population consisted of $N = 184 \times 4 = 736$ samples. These two populations, totaling $N = 2944$ cases, were used to train the network.

In the training phase, to extract the channel vectors from defect-present images, as per (8), we shifted each channel profile in $\boldsymbol{U}$ to be centered to the defect centroid. The channel vectors for the corresponding defect-absent images were obtained by shifting the channel profiles in $\boldsymbol{U}$ to the centroid of the location where the synthetic defect was inserted. With these shifted channel profiles, we extracted the corresponding channel vectors from both predicted and normal-dose images and used these vectors to calculate the task-specific loss term in (8). We performed a four-fold cross-validation to optimize the network. The training was performed on an NVIDIA TESLA V100 GPU with 32 GB of RAM. We trained separate networks for each dose level and a range of λ values. To select the optimized λ value for each dose level, we used a separate validation set obtained from 40 normal cases. Using the same strategy as for the training set, 20 of these cases were used to create the defect-present population of 20 × 12 = 240 samples. For a specific low-dose level, we denoised the images in the validation set using pretrained networks corresponding to different λ values. Using observer studies, as will be described in Section III-C, for each dose level, the value of λ that maximized performance on the detection task in a validation dataset was selected as the optimal λ.

C. Testing Procedure

The test set consisted of $N = 2052$ cases. These were generated using $N = 114$ normal MPI studies. Of these, 61 normal studies were used as the defect-absent population. To create the defect-present population, synthetic defects were inserted in the 53 normal studies. In addition to the 12 defect types, we also introduced six new defects with 45° extent to create out-of-distribution defect types in the test set. These new defects had severities and locations as the usual defects. Therefore, the test set consisted of 18 types of defects. Thus, for the observer study, the test defect-absent population consisted of 61 × 18 = 1098 samples and the test defect-present population consisted of 53 × 18 = 954 samples.

We evaluated the performance of the proposed method on the clinical task of detecting perfusion defects and using task-agnostic fidelity-based figures of merit. Performance was compared to low-dose images that were not denoised. We refer to this as the low-dose protocol. To assess the impact of using our task-specific denoising strategy, we also compared performance to images that were denoised using a commonly used DL-based denoising method [30] that was trained with a loss function that used only the fidelity term (setting $\lambda = 0$ in (4)). We refer to this method as the task-agnostic DL-based denoising (TADL) method. Comparing DEMIST with TADL method allowed assessing the impact of incorporating the task-specific term into the loss function on observer performance.

To objectively evaluate the proposed method on the task of detecting perfusion defects, we considered an anthropomorphic channelized Hotelling observer (CHO) [37] as a surrogate for the human observer. For clinical application, ideally the performance of the proposed method on the defect-detection task should be evaluated using human-observer studies by trained radiologists. However, such studies are time-consuming, expensive, and tedious. To address this challenge, model observers, such as the CHO [37], have been developed. Most importantly, CHOs with rotationally symmetric frequency channels have been validated to emulate human-observer performance on the task of detecting location-known perfusion defects in MPI SPECT [24], [25]. Thus, we used the CHO with these channels as our observer. We follow the same procedure as in [25] to define the rotationally symmetric frequency channels. Briefly, the start frequency and bandwidth of first channel was 0.1838 cycles/cm. The subsequent channels were adjacent to the previous one and had double the start frequency and bandwidth as the previous one.

We selected the 2-D short-axis slice and two adjacent slices from each MPI-SPECT image that contained the defect centroid for conducting the observer studies. From the centroid-containing slice, consistent with previous studies [38], we extracted a 32 × 32 region such that the defect centroid was at the center of the extracted region. This same 2-D region was also extracted from the two adjacent slices. Pixels values of each extracted region were mapped to the range [0, 255]. We then applied anthropomorphic rotationally symmetric frequency channels to each slice to compute the channel vectors. The channel vectors of defect-present and defect-absent populations were used to learn the template of the CHO using a leave-one-out approach. Following that, the test statistics were computed and used to perform the ROC analysis. Stratified analyses based on sex, defect severity, defect extent and scanner type were also performed. A schematic describing the process to obtain the CHO test statistic is shown in Fig. 3.

D. Figures of Merit

ROC analysis was performed on the test statistics derived with the CHO using the pROC package in R [50]. The area under the empirical ROC curve (AUC) was used as the figure of merit. Confidence intervals were calculated using Delong's method [51], which accounts for variability across cases. The AUC values were computed for the normal-dose and low-dose images and those denoised with DEMIST and TADL. To test the statistical significance of difference in AUC values between two methods, we used Delong's test as implemented within the pROC package [50]. To account for multiple hypothesis testing (DEMIST versus low-dose, DEMIST versus TADL and TADL versus low-dose), we used Bonferroni correction [52]. A corrected p value < 0.05 was used to infer a statistically significant difference. For quantitative evaluation based on image fidelity, we considered two widely used fidelity-based figures of merit: 1) RMSE and 2) SSIM.

IV. Results

A. Evaluation on the Task of Perfusion Defect Detection

Fig. 4 shows the AUC values obtained with the low-dose protocol, DEMIST, and TADL methods at all the considered low-dose levels, and with the normal-dose protocol. At all dose levels, DEMIST significantly outperformed low-dose protocol as well as the TADL method. The p-values of all the statistical tests presented in these results are included in the Supplementary material and in Rahman et al. [39]. We do note that the proposed method yields inferior performance on detection task compared to normal-dose protocol, an observation that we will discuss in the Discussions section.

Fig. 5 qualitatively shows the impact of the DEMIST and TADL methods on four representative cases. We observe in these cases that with the TADL method, even though the background looks less noisy compared to low-dose protocol, the defect tends to wash out. This observation is consistent with the findings reported in previous studies [21], [26]. In contrast, with DEMIST, the defect is visibly clearer even as the background looks less noisy compared to low-dose protocol. These representative cases provide an intuitive explanation for the improved performance of the DEMIST method.

Fig. 6(a) and (b) show the AUC values obtained with male and female populations, respectively. We observed that, for both sexes, the proposed method yielded a significant improvement in performance on the detection task at all dose levels compared to low-dose protocol. Moreover, in 5 out of 6 settings (3 dose levels × 2 sexes), DEMIST yielded significant improvement in detection-task performance compared to TADL method. Further, again, the TADL method generally did not improve (and in some cases degraded) performance compared to the low-dose protocol.

Figs. 7 and 8 show the AUC values as a function of defect extent and severity at different dose levels, respectively. We observe that, at all dose levels, the DEMIST method significantly improved observer performance for all considered defect extents and severity compared to low-dose protocol. Moreover, the DEMIST method significantly improved observer performance compared to TADL method in 15 out of 18 settings (3 dose levels × 6 defect types). Again, the TADL method was generally observed to not improve performance compared to low-dose protocol.

Fig. 9 shows the AUC values obtained for stratified analysis based on scanner models. In our study, data were collected across two scanners, namely, "GE Discovery NM/CT 670 Pro NaI" and "GE Discovery NM/CT 670 Pro CZT." For conciseness, we refer to these two scanners as NaI and CZT scanner, respectively. We observe from Fig. 9 that the DEMIST method significantly outperformed low-dose protocol in 4 out of 6 settings (3 dose levels × 2 scanners) and the TADL method in 4 out of 6 settings. We also observed that the performance of the TADL method deteriorated by comparison with the low-dose protocol in some settings. These findings demonstrate the advantage of the proposed method across different scanner types.

B. Quantitative Evaluation Based on Fidelity-Based Figures of Merit

The SSIM and RMSE metrics based on the entire image volume are presented in Table I for the proposed DEMIST method, TADL method and low-dose protocol. We observed that DEMIST yielded improved performance compared to low-dose protocol. Moreover, in general, both the proposed DEMIST method and the TADL method yielded very similar RMSE and SSIM values.

We also computed the RMSE inside the LV wall for low-dose images and images denoised with DEMIST and TADL for defect-absent cases. The normal-dose images were considered as reference for calculating this RMSE inside LV wall. The results, as shown in Table II, show the improvement in RMSE in the LV wall as obtained by DEMIST compared to the low-dose protocol.

V. Discussion

In this work, we proposed a method to denoise low-dose MPI-SPECT images while preserving features that assist in performing detection task by incorporating a task-specific loss term. We then evaluated our method on the clinical task of detecting perfusion defects in MPI-SPECT using a retrospective clinical study. The result in Fig. 4 shows that applying this method resulted in significantly improved defect-detection performance over just using low-dose images, as well as low-dose images denoised using the TADL method. These results provide evidence that incorporating this task-specific loss term can significantly improve observer performance beyond low-dose images and using a commonly used TADL method consistently across a range of defect characteristics. To the best of our knowledge, this is the first time that a DL-based denoising method for MPI SPECT has shown improved performance on the task of detecting perfusion defects in an anthropomorphic model-observer study.

To mathematically interpret the improved performance of the DEMIST method, we conducted an analysis similar to Yu et al. [21]. More specifically, we analyzed the effect of denoising on the first and second-order statistics of the channel vectors of the test set for both DEMIST and TADL. The analysis was performed for each defect type separately.

Denote the mean difference channel vector between defect-present and defect-absent cases as $\Delta\bar{\boldsymbol{v}}$ and the channel-vector covariance matrix as $\boldsymbol{K}_v$. The SNR of the CHO is given by

$$\mathrm{SNR}^2 = \Delta\bar{\boldsymbol{v}}^{-\mathrm{T}}\boldsymbol{K}_v^{-1}\Delta\bar{\boldsymbol{v}}. \tag{9}$$

If the test statistics of defect-absent and defect-present cases are normally distributed, AUC and SNR of the observer are monotonically related [18] and thus, the analysis of observer SNR yields insights on detection-task performance.

Consider that the reconstructed images have been reoriented and windowed with defect centroid at the center. Denote the mean difference reconstructed image between defect-present and defect-absent cases by $\Delta\bar{\hat{\mathbf{f}}}$. Thus, $\Delta\bar{\boldsymbol{v}} = (\boldsymbol{SU})^T\Delta\bar{\hat{\mathbf{f}}}$. As per (9), both the mean difference of the channel vector $\Delta\bar{\boldsymbol{v}}$ and covariance matrix $\boldsymbol{K}_v$ affect observer performance. Eigenanalysis of the covariance matrix provides a mechanism to analyze the combined effect of these two terms [21] on the observer SNR. Denote the mth eigenvector and eigenvalue of $\boldsymbol{K}_v$ by $\boldsymbol{u}_m$ and γ_m, respectively. We can express $\Delta\bar{\boldsymbol{v}}$ in terms of these eigenvectors as follows:

$$\Delta\bar{\boldsymbol{v}} = \sum_{m=1}^{C}\alpha_m\boldsymbol{u}_m, \tag{10}$$

where the coefficient $\alpha_m = \boldsymbol{u}_m^{\mathrm{T}}\Delta\bar{\boldsymbol{v}}$. Further, the SNR of the CHO is given by [21]

$$\mathrm{SNR}^2 = \sum_{m=1}^{C}\frac{\alpha_m^2}{\gamma_m} \tag{11}$$

Thus, assessing the impact of denoising on α_m and γ_m provides an interpretable approach to evaluate the effect of denoising on observer performance. Fig. 10 shows this analysis for two defect types with 6.25% dose level. We first plotted the mean difference reconstructed image $\Delta\bar{\hat{\mathbf{f}}}$ and mean difference channel vector $(\Delta\bar{\boldsymbol{v}})$ between defect-present and defect-absent cases [Fig. 10(a)–(f)]. We observe that the DEMIST method preserved the mean difference originally present in the normal-dose image for this defect type. However, as in Yu et al. [21], we observed that the TADL method reduced this mean difference, negatively impacting observer performance. Fig. 10(g) and (h) show the values of α_m and γ_m as a function of m, respectively. We observed that γ_m reduced for both DEMIST and TADL method compared to low-dose images, which would positively impact observer performance. However, with the TADL method, the values of α_m were lower compared to low-dose images, which leads to limited observer performance on detection task. In contrast, with the DEMIST method, the values of α_m do not reduce (and in some cases increase) compared to low-dose images, resulting in an overall improvement in performance on the defect-detection task. The γ_m values in Fig. 10 are listed in the supplementary material.

The DEMIST method consists of a hyperparameter λ. th penalizes the loss of task-specific features while performing denoising based on the loss function in (4). To qualitatively demonstrate the effect of this parameter, we present a representative result in Fig. 11. To generate this result, we denoised an MPI-SPECT image in the test set acquired at low-dose level of 6.25% with trained DEMIST networks associated with varying λ values. We observe that, for this example, assigning a higher weight to the task-specific loss term (as achieved by increasing λ) leads to improved defect visibility in the denoised image.

This improvement indicates that the incorporation of task-specific loss term preserves features used by human observers for performing detection tasks. These results also illustrate that the λ parameter can be interpreted as a term that controls the smoothness in the image. A lower value of λ results in an increased weight for the fidelity term, and is observed to lead to increased blur in the image, which then translates to the defect being washed out.

Stratified analysis based on patient sex, defect extent and defect severity showed that the proposed method continued to show improved performance compared to low-dose and TADL methods on the task of detecting perfusion defects (Figs. 6–8). We note here the difference in performance for male and female patients (Fig. 6). Since, the defect properties were similar across male and female patients, the difference could be attributed to the anatomical variations and myocardial activity uptake level. Further investigations are required to study these effects and this presents an area of future study.

We note in Figs. 4–8 that while DEMIST yielded improved performance on detection task compared to low-dose and TADL approaches, there is room for improvement compared to normal-dose protocol. To further improve performance, a more advanced network architecture [53] with the proposed task-specific loss term could be used. Also, given the heterogeneity in patient characteristics, increasing the amount of training data may make the method generalize well to test data and thus improve performance [54]. However, there is a possibility of fundamental information loss that might not be retrievable even if we increase the amount of training data. This topic requires further investigation. Moreover, we considered three low-dose levels but there may be other low-dose levels for which the proposed method may yield performance that is similar to normal-dose protocol.

In this article, we developed the task-specific denoising method in the context of cardiac SPECT. However, the method is general and could be applied to other medical imaging modalities where the task of interest is detecting abnormalities. Other applications could include reducing administered radiation dose in oncological PET images and reducing acquisition time for oncological magnetic resonance (MR) images. Another future research direction is to advance the underlying idea of DEMIST to tasks other than detection. DEMIST was developed and evaluated for detection tasks and not for other tasks, such as quantification or joint detection and quantification. However, the method could be advanced for other tasks where the task performance depends on mathematical features extracted from images.

Our study has several limitations. The first limitation is that DEMIST was validated with model observers and not human observers. While we considered a CHO-based model observer that has been shown to emulate human-observer performance, conducting this study with human observers would provide a more rigorous validation of the method. Additionally, the signal location was known to the anthropomorphic observer, but in clinical settings, this location is not known. Furthermore, given the location-known settings, we could not assess the performance of the method on falsely detecting defects at other locations. Localization ROC studies with human observers will enable us to validate whether the proposed method can improve human-observer performance on the task of perfusion-defect detection with unknown defect location. Reliable performance in a human-observer study would provide confidence for the clinical translation of this method. Here, we point out that to test the robustness of the method to different channelized observers, we also conducted the evaluation with another observer, namely, the channelized multi-template observer [55]. Our findings, which are provided in the Supplementary material and in [39], show that even with this observer, DEMIST significantly outperformed low-dose protocol and TADL method. This finding shows the robustness of the method to different observers.

A second limitation of this study is that the DEMIST method was trained with data where the defect-present cases contained synthetic inserted defects. This was because, during training, the knowledge of the presence of defect and the location of defect centroid in defect-present cases was required. However, due to the scaling of the defect-absent projection data during the defect insertion, the Poisson distribution may not be preserved in the final defect-present projection data. Ideally, thus, DEMIST should be trained using data with real perfusion defects. However, determining the ground truth regarding the presence of defects and their centroid is challenging. To address a similar issue of lack of ground truth while training a network to delineate tumors in PET images, Leung et al. [56] pretrained a network with multiple synthetic images where the tumor boundaries were known exactly, and then fine-tuned with a small number of clinical images. A similar strategy of pretraining DEMIST with multiple synthetic-defect images and then fine-tuning this network with a small number of training images where the defect centroid is obtained manually presents an area of future study. Another limitation was that we considered defects in only two regions. Increasing the number of defect locations, including septal and lateral walls of the LV, would provide further insights on the robustness of the proposed method. Furthermore, in practical scenarios, there could be multiple defect locations in the same case. The proposed DEMIST method can be extended by extracting channel vectors from each of these locations. Additionally, the method was evaluated with only single-center data. However, the results motivate evaluation of the data across multiple centers to assess the generalizability of this technique across centers. Finally, the method was developed for nongated MPI SPECT images. Another area of future research is advancing this method to gated MPI SPECT [57], [58]. One challenge here is identifying the center of the defect. The proposed method can be advanced to account for this issue by extracting channel vectors for a neighborhood of possible defect centers.

VI. Conclusion

A detection-task-specific deep-learning-based method (DEMIST) was proposed to denoise low-dose MPI-SPECT images with the goal of improving performance on the clinical task of detecting perfusion defects compared to low-dose images. For this purpose, we introduced a task-specific loss term in our loss function that penalizes the loss of anthropomorphic channel features. According to the RELIANCE guidelines [41], our evaluation study yields the following claim: a deep-learning-based detection-task-specific denoising method for MPI-SPECT improved performance in images acquired at 6.25%, 12.5%, and 25% dose levels on the task of detecting inserted location-known perfusion defects with a significance level of 5% as evaluated in a retrospective clinical study with single-center multiscanner data and with an anthropomorphic channelized Hotelling observer. The results provide strong evidence to evaluate DEMIST with human observers. Open-source code for the proposed method is available at https://github.com/AshequrRahman/demist-tf.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgment

The authors thank the Research Infrastructure Service (RIS) at Washington University for providing computational resources.

This work was supported in part by the National Institute of Biomedical Imaging and Bioengineering of the National Institute of Health under Grant R01-EB031051, Grant R56-EB028287, and Grant R01-EB031962; and in part by the NSF CAREER Award under Grant 2239707.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by Washington University Institutional Review Board under Application No. 201905164.

REFERENCES

[1]. Zaret BL and Beller GA, “Clinical nuclear cardiology: State of the art and future directions,” JAMA, vol. 305, no. 16, pp. 1712–1713, 2010.

[2]. Henzlova MJ, Duvall WL, Einstein AJ, Travin MI, and Verberne HJ, “ASNC imaging guidelines for SPECT nuclear cardiology procedures: Stress, protocols, and tracers,” J. Nucl. Card, vol. 23, no. 3, pp. 606–639, 2016.

[3]. Beller GA, “Importance of consideration of radiation doses from cardiac imaging procedures and risks of cancer,” J. Nucl. Card, vol. 17, pp. 1–3, Jan./Feb. 2010.

[4]. Cohen MC, “Radiation reduction: How low can you go?” J. Nucl. Card, vol. 17, pp. 524–5, Mar. 2010.

[5]. Yuan J-W, Wang Y-T, and Lu C-Z, “Coronary arteriography in the diagnosis results and prognosis analysis of suspected coronary artery disease in patients with normal SPET myocardial perfusion imaging,” Hell. J. Nucl. Med, vol. 18, no. 3, pp. 215–21, 2015. [PubMed: 26637500]

[6]. Bousse A. et al. , “A review on low-dose emission tomography post-reconstruction denoising with neural network approaches,” IEEE Trans. Radiat. Plasma Med. Sci, early access, Jan. 2, 2024, doi: 10.1109/TRPMS.2023.3349194.

[7]. Xie H. et al. , “Deep-learning-based few-angle cardiac SPECT reconstruction using transformer,” IEEE Trans. Radiat. Plasma Med. Sci, vol. 7, no. 1, pp. 33–40, Jan. 2023. [PubMed: 37397179]

[8]. Chen X. et al. , "DuDoSS: Deep-learning-based dual-domain sinogram synthesis from sparsely sampled projections of cardiac SPECT," Med. Phys, vol. 50, no. 1, pp. 89–103, 2023. [PubMed: 36048541]

[9]. Chen X. et al. , "Direct and indirect strategies of deep-learning-based attenuation correction for general purpose and dedicated cardiac SPECT," Eur. J. Nucl. Med. Mol. Imag, vol. 49, no. 9, pp. 3046–3060, 2022.

[10]. Yu Z, Rahman MA, Abbey CK, Siegel BA, and Jha AK, "Development and task-based evaluation of a scatter-window projection and deep learning-based transmission-less attenuation compensation method for myocardial perfusion SPECT," in Proc. SPIE Int. Soc. Opt. Eng, vol. 12463, 2023, p. 124631E.

[11]. Yu Z, Rahman MA, Schindler T, Laforest R, and Jha AK, "A physics and learning-based transmission-less attenuation compensation method for SPECT," in Proc. SPIE Int. Soc. Opt. Eng, vol. 11595, 2021, p. 1159512.

[12]. Ramon AJ, Yang Y, Pretorius PH, Johnson KL, King MA, and Wernick MN, "Initial investigation of low-dose SPECT-MPI via deep learning," in Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Proc. (NSS/MIC), 2018, pp. 1–3.

[13]. Aghakhan Olia N. et al. , "Deep learning–based denoising of low-dose SPECT myocardial perfusion images: Quantitative assessment and clinical performance," Eur. J. Nucl. Med. Mol. Imag, vol. 49, no. 5, pp. 1508–1522, 2022.

[14]. Sun J, Du Y, Li C, Wu T-H, Yang B, and Mok GS, "Pix2Pix generative adversarial network for low dose myocardial perfusion SPECT denoising," Quant. Imag. Med. Surg, vol. 12, no. 7, p. 3539, 2022.

[15]. Sun J. et al. , "Deep learning-based denoising in projection-domain and reconstruction-domain for low-dose myocardial perfusion SPECT," J. Nucl. Cardiol, vol. 30, pp. 970–985, Aug. 2022. [PubMed: 35982208]

[16]. Barrett HH, "Objective assessment of image quality: Effects of quantum noise and object variability," J. Opt. Soc. Amer. A, vol. 7, no. 7, pp. 1266–1278, 1990. [PubMed: 2370589]

[17]. Jha AK et al. , "Objective task-based evaluation of artificial intelligence-based medical imaging methods: Framework, strategies, and role of the physician," PET Clin, vol. 16, no. 4, pp. 493–511, 2021. [PubMed: 34537127]

[18]. Barrett HH and Myers KJ, Foundations of Image Science. Hoboken, NJ, USA: Wiley, 2013.

[19]. Pretorius PH et al. , "Observer studies of image quality of denoising reduced-count cardiac single photon emission computed tomography myocardial perfusion imaging by three-dimensional Gaussian post-reconstruction filtering and deep learning," J. Nucl. Cardiol, vol. 30, no. 6, pp. 2427–2437, 2023. [PubMed: 37221409]

[20]. Krupinski EA, "The important role of task-based model observers and related techniques in medical imaging," J. Nucl. Card, vol. 28, pp. 638–640, Apr. 2021.

[21]. Yu Z. et al. , "Need for objective task-based evaluation of deep learning-based denoising methods: A study in the context of myocardial perfusion SPECT," Med. Phys, vol. 50, no. 7, pp. 4122–4137, 2023. [PubMed: 37010001]

[22]. Li K, Zhou W, Li H, and Anastasio MA, "Assessing the impact of deep neural network-based image denoising on binary signal detection tasks," IEEE Trans. Med. Imag, vol. 40, no. 9, pp. 2295–2305, Sep. 2021.

[23]. Prabhat K, Zeng R, Farhangi MM, and Myers KJ, "Deep neural networks-based denoising models for CT imaging and their efficacy," in Proc. SPIE Med. Imag, vol. 11595, 2021, pp. 105–117.

[24]. Sankaran S, Frey EC, Gilland KL, and Tsui BM, "Optimum compensation method and filter cutoff frequency in myocardial SPECT: A human observer study," J. Nucl. Med, vol. 43, no. 3, pp. 432–438, 2002. [PubMed: 11884505]

[25]. Wollenweber S, Tsui B, Lalush D, Frey E, LaCroix K, and Gullberg G, "Comparison of hotelling observer models and human observers in defect detection from myocardial SPECT imaging," IEEE Trans. Nucl. Sci, vol. 46, no. 6, pp. 2098–2103, Dec. 1999.

[26]. Ongie G, Sidky EY, Reiser IS, and Pan X, "Optimizing model observer performance in learning-based CT reconstruction," in Proc. SPIE Med. Imag, vol. 12035, 2022, pp. 55–59.

[27]. Li K, Li H, and Anastasio MA, “A task-informed model training method for deep neural network-based image denoising,” in Proc. SPIE Med. Imag, vol. 12035, 2022, pp. 249–255.

[28]. Han M, Shim H, and Baek J, “Low-dose CT denoising via convolutional neural network with an observer loss function,” Med. Phys, vol. 48, no. 10, pp. 5727–5742, 2021. [PubMed: 34387360]

[29]. Rahman MA, Yu Z, Siegel BA, and Jha AK, “A task-specific deep-learning-based denoising approach for myocardial perfusion SPECT,” in Proc. SPIE Med. Imag, 2023, p. 1246719.

[30]. Ramon AJ, Yang Y, Pretorius PH, Johnson KL, King MA, and Wernick MN, “Improving diagnostic accuracy in low-dose SPECT myocardial perfusion imaging with convolutional denoising networks,” IEEE Trans. Med. Imag, vol. 39, no. 9, pp. 2893–2903, Sep. 2020.

[31]. Hwang D. et al. , “Improving the accuracy of simultaneously reconstructed activity and attenuation maps using deep learning,” J. Nucl. Med, vol. 59, no. 10, pp. 1624–1629, Oct. 2018. [PubMed: 29449446]

[32]. Abbey CK and Barrett HH, “Human- and model-observer performance in ramp-spectrum noise: Effects of regularization and object variability,” J. Opt. Soc. Amer. A Opt. Image Sci. Vis, vol. 18, no. 3, pp. 473–488, Mar. 2001. [PubMed: 11265678]

[33]. Burgess A, Wagner R, Jennings R, and Barlow HB, “Efficiency of human visual signal discrimination,” Science, vol. 214, no. 4516, pp. 93–94, 1981. [PubMed: 7280685]

[34]. Burgess AE and Colborne B, “Visual signal detection. IV. Observer inconsistency,” J. Opt. Soc. Amer. A, vol. 5, no. 4, pp. 617–627, 1988. [PubMed: 3404312]

[35]. Burgess A, “Statistically defined backgrounds: Performance of a modified nonprewhitening observer model,” J. Opt. Soc. Amer. A, vol. 11, no. 4, pp. 1237–1242, 1994.

[36]. Loo L-N, Doi K, and Metz CE, “A comparison of physical image quality indices and observer performance in the radiographic detection of nylon beads,” Phys. Med. Biol, vol. 29, no. 7, p. 837, 1984. [PubMed: 6473512]

[37]. Myers KJ and Barrett HH, “Addition of a channel mechanism to the ideal-observer model,” J. Opt. Soc. Amer. A, vol. 4, no. 12, pp. 2447–2457, 1987. [PubMed: 3430229]

[38]. Frey EC, Gilland KL, and Tsui BM, “Application of task-based measures of image quality to optimization and evaluation of three-dimensional reconstruction-based compensation methods in myocardial perfusion SPECT,” IEEE Trans. Med. Imag, vol. 21, no. 9, pp. 1040–1050, Sep. 2002.

[39]. Rahman MA, Yu Z, Abbey CK, Siegel BA, and Jha AK “Supplementary: DEMIST: A deep-learning-based task-specific denoising approach for myocardial perfusion SPECT,” github. 2023. [Online]. Available: https://github.com/ajha4/DEMIST/blob/main/demist_supp_v1_june9.pdf

[40]. Kingma DP and Ba J, “Adam: A method for stochastic optimization,” 2014, arXiv:1412.6980.

[41]. Jha AK et al. , “Nuclear medicine and artificial intelligence: Best practices for evaluation (the RELAINCE guidelines),” J. Nucl. Med, vol. 63, no. 9, pp. 1288–1299, 2022. [PubMed: 35618476]

[42]. Juan Ramon A. et al. , “Investigation of dose reduction in cardiac perfusion SPECT via optimization and choice of the image reconstruction strategy,” J. Nucl. Card, vol. 25, no. 6, pp. 2117–2128, 2018.

[43]. He X, Links JM, and Frey EC, “An investigation of the trade-off between the count level and image quality in myocardial perfusion SPECT using simulated images: The effects of statistical noise and object variability on defect detectability,” Phys. Med. Biol, vol. 55, no. 17, p. 4949, 2010. [PubMed: 20693615]

[44]. Heiberg E, Sjögren J, Ugander M, Carlsson M, Engblom H, and Arheden H, “Design and validation of segment-freely available software for cardiovascular image analysis,” BMC Med. Imag, vol. 10, no. 1, pp. 1–13, 2010.

[45]. Soneson H, Ubachs JF, Ugander M, Arheden H, and Heiberg E, “An improved method for automatic segmentation of the left ventricle in myocardial perfusion SPECT,” J. Nucl. Med, vol. 50, no. 2, pp. 205–213, 2009. [PubMed: 19164235]

[46]. Narayanan MV, King MA, Leppo J, Dahlbert S, Pretorius PH, and Gifford HC, “Optimization of regularization of attenuation and scatter-corrected ^{99m}Tc cardiac SPECT studies for defect detection using hybrid images,” IEEE Trans. Nucl. Sci, vol. 48, no. 3, pp. 785–789, Jun. 2001.

[47]. Ljungberg M. and Strand S-E, "A Monte Carlo program for the simulation of scintillation camera characteristics," Comput. Methods Program. Biomed, vol. 29, no. 4, pp. 257–272, 1989.

[48]. Toossi MB, Islamian JP, Momennezhad M, Ljungberg M, and Naseri S, "SIMIND Monte Carlo simulation of a single photon emission CT," J. Med. Phys, vol. 35, no. 1, p. 42, 2010. [PubMed: 20177569]

[49]. Merlin T. et al. , "CASToR: A generic data organization and processing code framework for multi-modal and multi-dimensional tomographic reconstruction," Phys. Med. Biol, vol. 63, no. 18, 2018, Art. no. 185005.

[50]. Robin X. et al. , "pROC: An open-source package for R and S+ to analyze and compare ROC curves," BMC Bioinf, vol. 12, p. 77, Mar. 2011.

[51]. DeLong ER, DeLong DM, and Clarke-Pearson DL, "Comparing the areas under two or more correlated receiver operating characteristic curves: A nonparametric approach," Biometrics, vol. 44, no. 3, pp. 837–845, 1988. [PubMed: 3203132]

[52]. Bland JM and Altman DG, "Multiple significance tests: The Bonferroni method," BMJ, vol. 310, no. 6973, p. 170, 1995. [PubMed: 7833759]

[53]. Fan CM, Liu TJ, and Liu KH, "SUNet: Swin transformer UNet for image denoising," in Proc. IEEE Int. Symp. Circuits Syst, 2022, pp. 2333–2337.

[54]. Chaudhary S, Moon S, and Lu H, "Fast, efficient, and accurate neuroimaging denoising via supervised deep learning," Nat. Commun, vol. 13, no. 1, p. 5165, 2022. [PubMed: 36056020]

[55]. Li X, Jha AK, Ghaly M, Elshahaby FE, Links JM, and Frey EC, "Use of sub-ensembles and multi-template observers to evaluate detection task performance for data that are not multivariate normal," IEEE Trans. Med. Imag, vol. 36, no. 4, pp. 917–929, 2016.

[56]. Leung KH et al. , "A physics-guided modular deep-learning based automated framework for tumor segmentation in PET," Phys. Med. Biol, vol. 65, no. 24, 2020, Art. no. 245032.

[57]. Paul AK and Nabi HA, "Gated myocardial perfusion SPECT: Basic principles, technical aspects, and clinical applications," J. Nucl. Med, vol. 32, no. 4, pp. 179–187, 2004.

[58]. Zhang X, Yang Y, Pretorius PH, and King MA, "Assessment of deep-learning based motion compensation on detection of perfusion defects in cardiac-gated SPECT images," in Proc. IEEE 20th Int. Symp. Biomed. Imag. (ISBI), 2023, pp. 1–4.

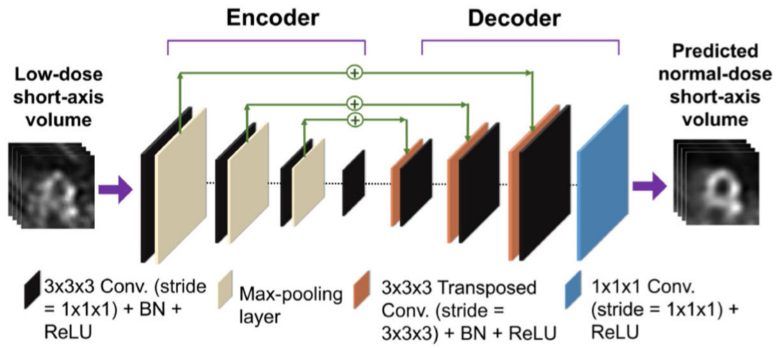


Fig. 1.

Schematic with details of the encoder–decoder denoising network architecture (BN = batch normalization and ReLU = Rectified Linear Unit).

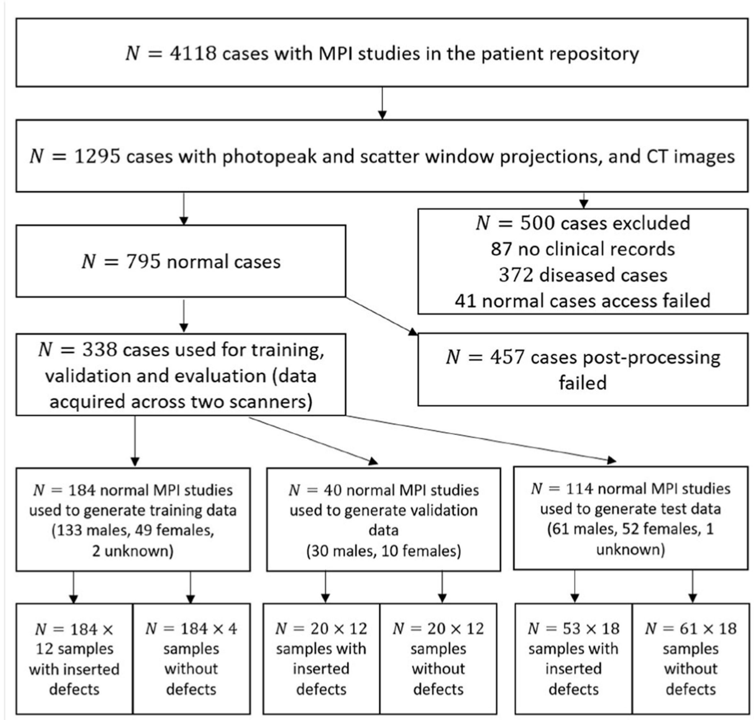


Fig. 2.

Patient data collection from MPI studies and their distribution in various stages of data curation. Sample refers to multiple cases derived from each MPI study.

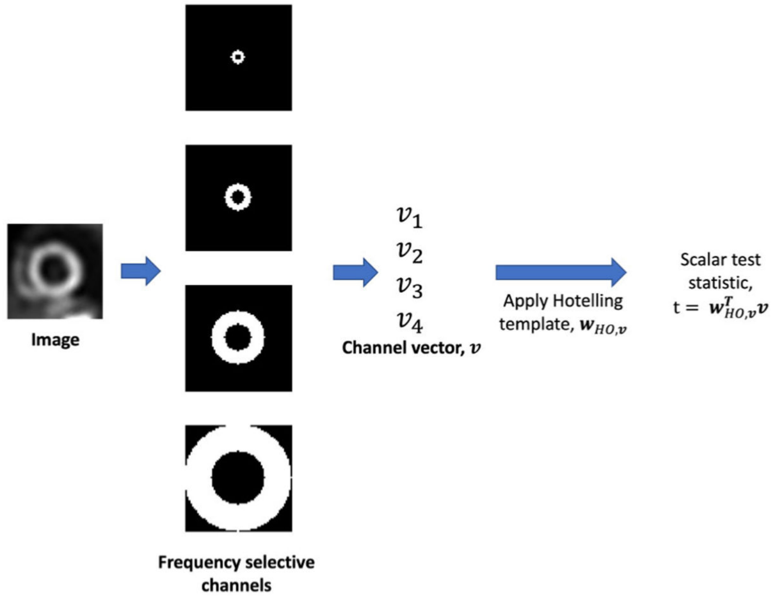


Fig. 3.
Schematic with the detailed process to generate test statistic using CHO with anthropomorphic frequency selective channels.

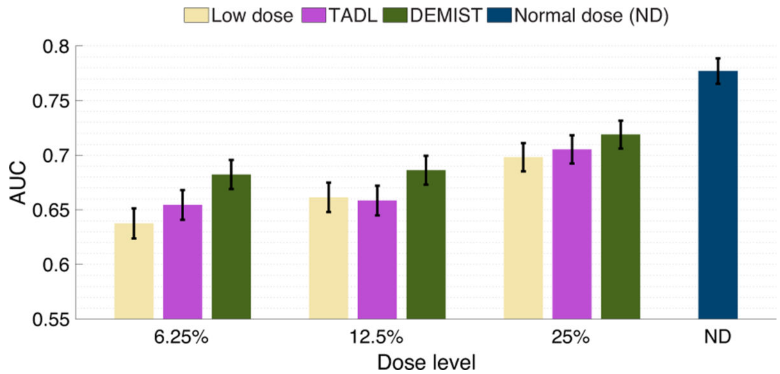


Fig. 4.

AUC values obtained for the normal-dose and low-dose images, and the images denoised using the DEMIST and TADL approaches at various dose levels with CHO. Error bars denote 95% confidence intervals.

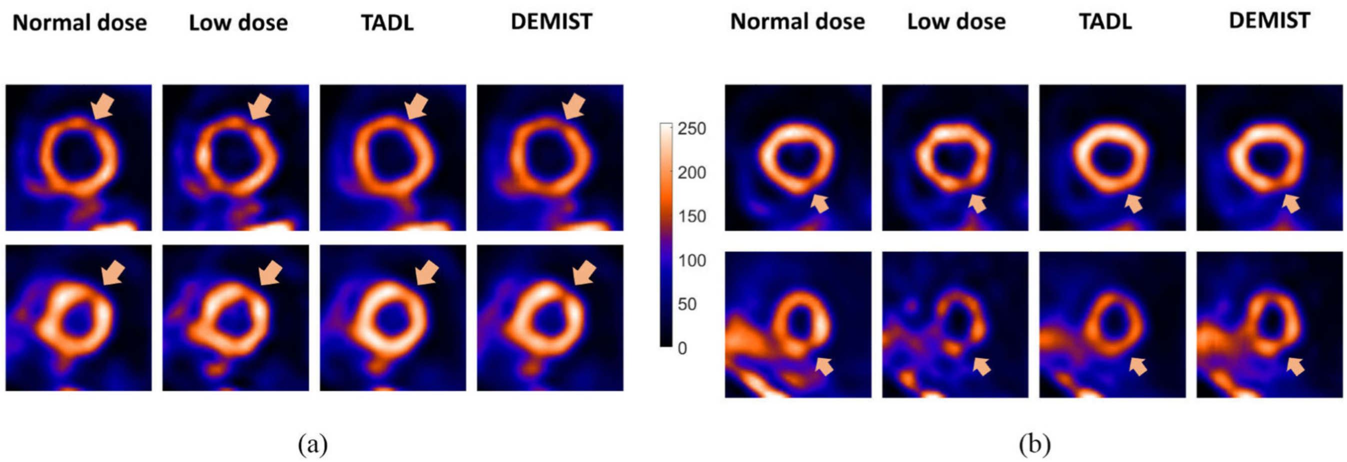


Fig. 5.

Four representative tests cases derived from four different patients, qualitatively showing the performance of TADL method and proposed DEMIST method. The short-axis slice containing the defect centroid is shown in all four cases. For all cases, the low-dose level was set to 12.5%. In (a) and (b), defects were in anterior and inferior wall, respectively. For all four cases, the defects had an extent of 30° and severity of 25%. First, we note that the background appears less noisy compared to low-dose images with both TADL and DEMIST. The defect tends to become less detectable with the TADL (no task-specific loss term). Further, the defect was visually clearer with the proposed DEMIST method.

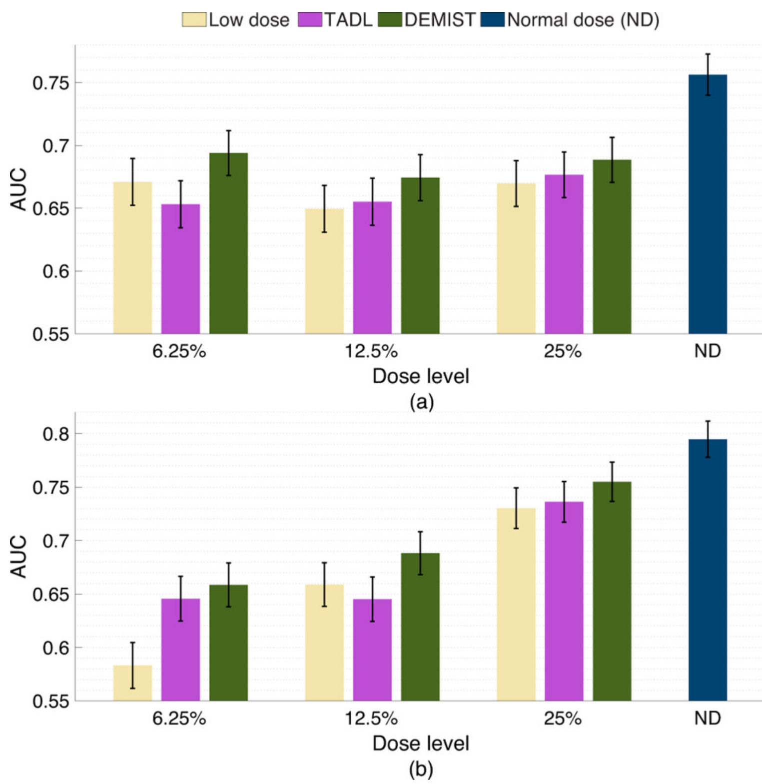


Fig. 6.

AUC values obtained for the different approaches and at various dose levels with (a) male and (b) female patients using CHO. Error bars denote 95% confidence intervals.

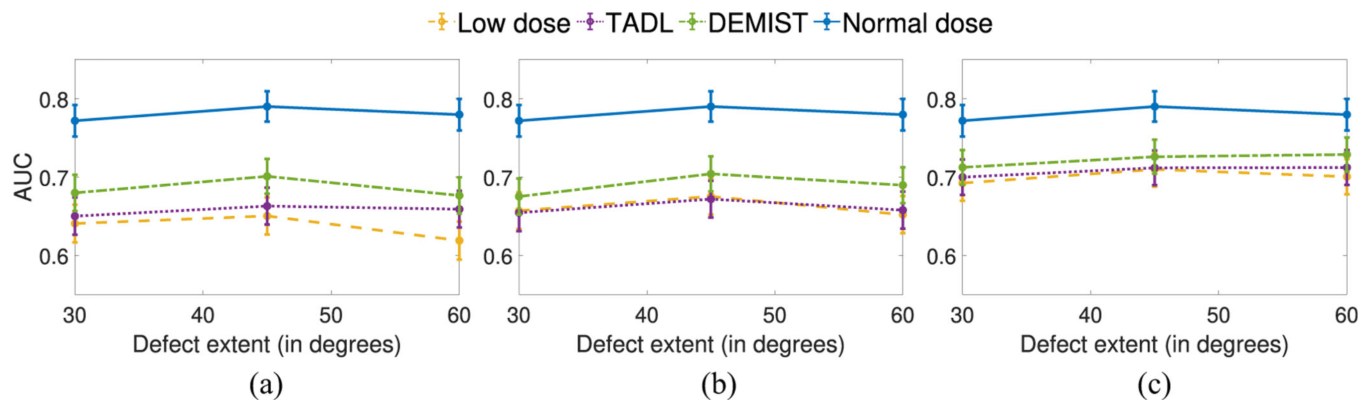
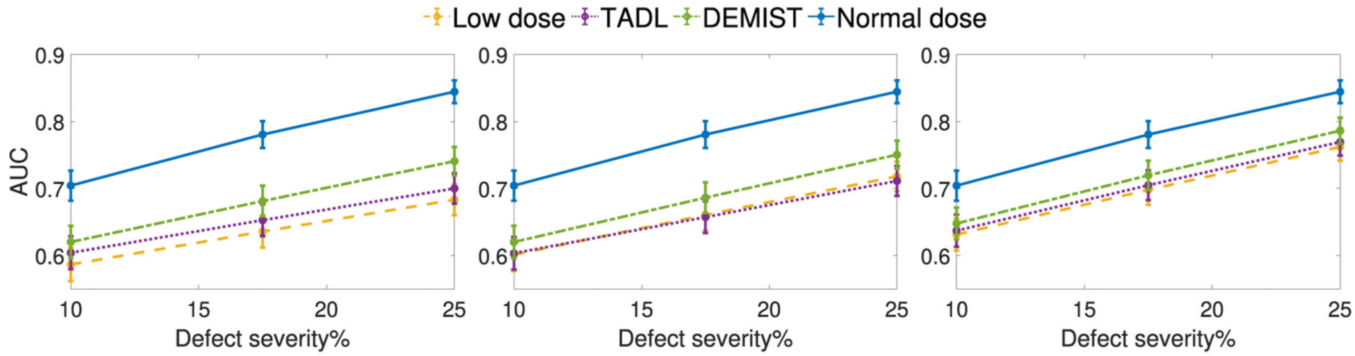


Fig. 7.

AUC values obtained using CHO for the various approaches as a function of different defect extents with (a) 6.25%, (b) 12.5%, and (c) 25% dose levels. Error bars denote 95% confidence intervals.



(a) (b) (c)

Fig. 8.

AUC values obtained using CHO for the various approaches as a function of different defect severities with (a) 6.25%, (b) 12.5%, and (c) 25% dose levels. Error bars denote 95% confidence intervals.

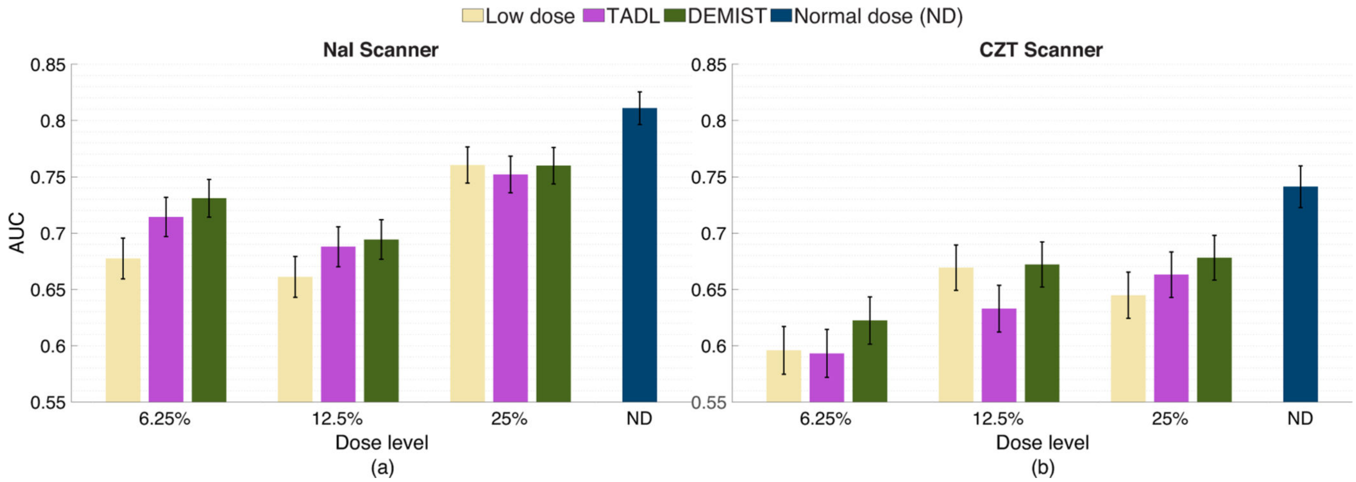


Fig. 9.

AUC values obtained for the considered approaches and at different dose levels with data from the (a) NaI and (b) CZT scanners using CHO. Error bars denote 95% confidence intervals

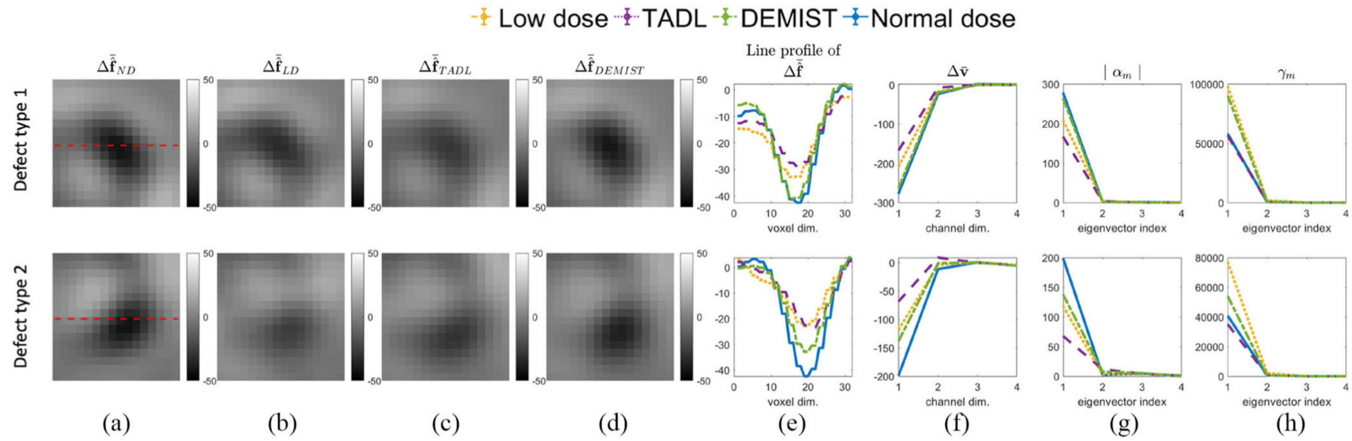


Fig. 10.

Mean difference reconstructed short-axis image between defect-absent and defect-present hypotheses for (a) normal-dose, (b) low-dose, (c) TADL, and (d) DEMIST. The images in (a)–(d) are windowed to a region centered to the defect centroid. (e) Line profile of the mean difference reconstructed image of (a)–(d). The red dashed lines in (a) represent the lines along which the profiles are drawn. (f) Mean difference channel vector $\Delta\bar{\boldsymbol{v}}$ between defect-absent and defect-present hypotheses for various approaches. (g) Absolute value of coefficient α_m and (h) eigenvalue spectra of noise covariance matrix. Low-dose level was set to 6.25% (Defect type 1: 60° extent, 25% severity and anterior wall defect. Defect type 2: 45° extent, 25% severity and inferior wall defect).

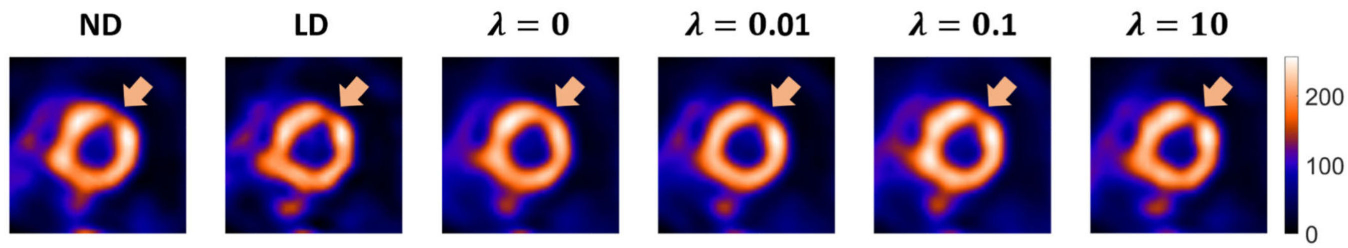


Fig. 11.

From left, normal-dose (ND) image, and low-dose (LD) image acquired at 12.5% dose level and images denoised with proposed method with varying λ as indicated on top of the image. Increasing λ results in recovery of defect visibility. However, the increase in λ also results in a decrease in background smoothness.

TABLE I

RMSE and SSIM Metric for Different Method at Different Dose Levels

Dose level	Metric	Low dose	TADL	DEMIST
6.25%	RMSE	6.87	5.00	5.58
	SSIM	0.77	0.85	0.84
12.5%	RMSE	4.81	4.10	4.01
	SSIM	0.86	0.89	0.89
25%	RMSE	3.16	2.94	2.94
	SSIM	0.93	0.93	0.94

TABLE II

Mean RMSE Inside LV Wall at Various Dose Levels for Different Approaches

Dose level	Low dose	TADL	DEMIST
6.25%	13.23	7.10	7.93
12.5%	9.49	6.02	5.96
25%	6.23	4.78	5.00