An Improved Iterative Neural Network for High-Quality Image-Domain Material Decomposition in Dual-Energy CT

Zhipeng Li[1], Yong Long[1], Il Yong Chun[2]
[1] University of Michigan - Shanghai Jiao Tong University Joint Institute, Shanghai Jiao Tong University, Shanghai 200240, China
[2]School of Electronic and Electrical Engineering, Sungkyunkwan University, Suwon, Gyeonggi 16419, Republic of Korea

Version typeset January 20, 2022

Yong Long and Il Yong Chun are corresponding authors.
Email: yong.long@sjtu.edu.cn; iychun@skku.edu

Abstract

Purpose: Dual-energy computed tomography (DECT) has been widely used in many applications that need material decomposition. Image-domain methods directly decompose material images from high- and low-energy attenuation images, and thus, are susceptible to noise and artifacts on attenuation images. The purpose of this study is to develop an improved iterative neural network (INN) for high-quality image-domain material decomposition in DECT, and to study its properties.
Methods: We propose a new INN architecture for DECT material decomposition. The proposed INN architecture uses distinct cross-material convolutional neural network (CNN) in image refining modules, and uses image decomposition physics in image reconstruction modules. The distinct cross-material CNN refiners incorporate distinct encoding-decoding filters and cross-material model that captures correlations between different materials. We study the distinct cross-material CNN refiner with patch-based reformulation and tight-frame condition.
Results: Numerical experiments with extended cardiac-torso phantom and clinical data show that the proposed INN significantly improves the image quality over several image-domain material decomposition methods, including a conventional model-based image decomposition (MBID) method using an edge-preserving regularizer, a recent MBID method using pre-learned material-wise sparsifying transforms, and a noniterative deep CNN method. Our study with patch-based reformulations reveals that learned filters of distinct cross-material CNN refiners can approximately satisfy the tight-frame condition.
Conclusions: The proposed INN architecture achieves high-quality material decompositions using iteration-wise refiners that exploit cross-material properties between different material images with distinct encoding-decoding filters. Our tight-frame study implies that cross-material CNN refiners in the proposed INN architecture are useful for noise suppression and signal restoration.

This paper has supplementary material. The prefix "S" indicates the numbers in section, equation, and figure in the supplementary material.

Contents

I Introduction

Dual-energy CT (DECT) has been increasingly used in many clinical and industrial applications, including kidney stone characterization[1], iodine quantification[2,3], security inspection[4,5], and nondestructive testing[6]. Compared to conventional single-energy X-ray CT, DECT provides two sets of attenuation measurements at high and low energies. Because the linear attenuation coefficient is material and energy dependent, DECT can characterize different constituent materials in a mixture, known as material decomposition[7]. Decomposed material images provide the elemental material compositions of the imaged object. Researchers have been studying material decomposition or reconstruction with spectral CT[8] and photon-counting CT[9] that can simultaneously acquire more than two spectral measurements.

I.A Literature Review

Model-based image decomposition (MBID) methods incorporate material composition physics, statistical model of measurements, and some prior information of unknown material images. Existing MBID methods for DECT can be classified into direct (projection-to-image domain)[10], projection-domain[11], and image-domain[12] decompositions. Direct decomposition methods perform image decomposition and reconstruction simultaneously, and generate material images directly from collected high and low energy measurements. This type of methods can reduce the cross-talk and beam-hardening artifacts by using an accurate forward model of the DECT system along with priors. However, direct decomposition algorithms need large computational costs, because at each iteration, they apply computationally expensive forward and backward projection operators. Projection-domain methods first decompose high- and low-energy sinograms into sinograms of materials, followed by an image reconstruction method such as filtered back projection (FBP) to obtain material images. Although above two types of methods improve the decomposition accuracy compared to image-domain methods, they usually require accurate system calibrations that use non-linear models[13,14]. In addition, those methods require sinograms or pre-log measurements that are in general not readily available from commercial CT scanners. Image-domain methods do not require projection operators and decompose readily available reconstructed high- and low-energy attenuation images into material images, and are more computationally efficient than direct and projection-domain decomposition methods. However, image-domain methods lack complete DECT imaging model. This may increase noise and artifacts in decomposed material images.

To improve image-domain DECT material decomposition methods, incorporating appropriate prior knowledge or regularizer into decomposition algorithms is critical. Many MBID methods have been proposed from this perspective. Niu *et al.*[12] proposed an iterative decomposition method that incorporates the noise variance of two attenuation images into the least-squares data-fit term. This better suppressed noise and artifacts on decomposed material images than a simple direct matrix inversion method. Xue *et al.*[15] proposed an MBID method that uses the weighted least-squares data-fit model[12] and an edge-preserving (EP) hyperbolar regularizer—called DECT-EP. Recently, there has been growing interest in data-driven methods such as MBID using pre-learned prior operators. Examples include learned synthesis operator/dictionary[16,17] and analysis operator/transform[18,19]. Dictionary learning has been applied to image-domain DECT material decomposition[17] and improved image decomposition compared to non-adaptive MBID methods. We proposed a data-driven method DECT-ST[19] that uses two pre-learned sparsifying transforms (ST) in a prior model to better sparsify the two different materials, and improved the image decomposition accuracy. We also proposed a clustering based cross-material method[20] that assumes correlations between different materials, and followed by a generalized mixed material method[21] that considers both individual properties (e.g., different materials have different densities and structures) and correlations of different material images.

In the past few years, deep regression neural network (NN) methods have been gaining popularity in medical imaging applications, for example, CT image denoising[22,23]. Several deep convolutional NN (dCNN) methods have also been proposed for image-domain DECT material decomposition. Liao *et al.*[24] proposed a cascaded dCNN method to obtain a material image from a single energy attenuation image. The first dCNN roughly maps a single attenuation image to a material image, followed by the other dCNN maps the material image to a high-quality material image. A dCNN method with two input and output channels that directly maps from two high- and low-energy attenuation images to two material images has also been proposed[25] . Different from the first dCNN used in aforementioned cascaded dCNN method[24] that obtains two material images individually, butterfly network[26] decomposes material images with additional CNNs between two attenuation images to perform information exchange. Clark *et al.*[27] investigated the conventional U-Net architecture for image-domain multi-material decomposition. However, the aforementioned methods have the high NN complexity that can increase the overfitting risk particularly when limited training samples are available.

An alternative approach is a so-called iterative NN (INN), which has been successfully applied to diverse imaging problems[28–34]. This approach incorporates iteration-wise image refining NNs into block-wise model-based image reconstruction algorithm. INN improves generalization capability compared to noniterative deep NN by balancing imaging physics and prior information estimated via refining CNNs, particularly when training samples are limited[30,31]. ADMM-Net is a pioneer INN architecture developed by unrolling the alternating direction method of multipliers (ADMM) model-based image reconstruction (MBIR) algorithm[34]; it has been succesfully applied to highly-undersampled MRI[34], low-dose CT[30], etc. BCD-Net is an INN architecture that generalizes the block coordinate descent (BCD) MBIR algorithm using learned convolutional regularizers, while showing better performance over ADMM-Net[30,32]. Its original work[28] uses the identical encoding-decoding architecture, i.e., each filter in decoder is a rotated version of that in encoder, and was successfully applied to highly-undersampled MRI (using single coil). Subsequent works[30,31] use the distinct encoding-decoding architecture for BCD-Net, and successfully applied modified BCD-Net to low-dose CT and low-count PET reconstruction. The Momentum-Net architecture generalizes a block-wise MBIR algorithm that uses momentum and majorizers for fast convergence without needing inner iterations[32]; it has been successfully applied to low-dose[33] and sparse-view[32] CT reconstruction. Different from the aforementioned INN methods that solve image reconstruction problems in low-dose or sparse-view CT, highly-undersampled MRI, and low-count PET, the proposed INN architecture is designed for image-domain material decomposition in DECT. The initial version of this work was presented in a conference[35], where we used an MBID cost function for the model-based image reconstruction module of BCD-Net, and demonstrated that BCD-Net significantly improved image quality over DECT-EP and DECT-ST. The initial BCD-Net work[35] has a single-hidden layer or "shallow" CNN (sCNN) architecture, where sCNN refiner has identical encoding-decoding architecture individually for two different materials (e.g., water and bone). The aforementioned INNs are trained in a supervised manner, whereas the recent study[36] applied a self-supervised image denoising method to an INN.

I.B Contributions

Image-domain material decomposition methods in DECT are susceptible to noise and artifacts (see Section I.A). Our aim is to obtain high-quality decomposed material images in DECT with improved image-domain material decomposition methods. To achieve the goal, the paper proposes an improved BCD-Net architecture. The proposed BCD-Net uses iteration-wise sCNN refiners, where they use *1)* distinct encoding-decoding architecture, i.e., each filter in decoding convolution is distinct from that in encoding convolution, and *2)* cross-material model that captures correlations between different material images. We refer to the previous BCD-Net in the earlier conference work[35] as BCD-Net-sCNN-lc and the proposed BCD-Net in this work as BCD-Net-sCNN-hc, where lc and hc stand for low and high complexity, respectively. In addition, we study the proposed distinct cross-material CNN architecture with the patch-based perspective, empirically showing that learned filters of distinct cross-material CNN refiners at the last BCD-Net iteration approximately satisfy the tight-frame condition. The patch-based reformulation reveals that the proposed CNN architecture has the cross-material property, and specializes to BCD-Net-sCNN-lc[35] refiners. Our tight-frame studies imply that cross-material CNN refiners are useful for noise suppression and signal restoration. The quantitative and qualitative results with extended cardiac-torso (XCAT) phantom and clinical data show that the proposed BCD-Net-sCNN-hc architecture significantly improves the decomposition quality compared to the conventional MBID method, DECT-EP[15], and the following recent image-domain decomposition methods, a noniterative dCNN method and a MBID method, DECT-ST[19], that uses a learned regularizer in an unsupervised way, and BCD-Net-sCNN-lc[35].

I.C Organization

The rest of this paper is organized as follows. Section II describes the proposed BCD-Net architecture for DECT image-domain MBID, studies the distinct cross-material refining sCNN architecture with the patch-based reformulation and the tight-frame condition, and provides training and testing algorithms for proposed BCD-Net architectures. Section III reports results of various decomposition methods on XCAT phantom and clinical data, along with comparisons and discussions. Finally, we make conclusions of this paper, and describe future work in Section IV.

II Methods

This section proposes the BCD-Net-sCNN-hc architecture, studies properties of its refiners, introduces its variations, and describes its training and testing processes.

II.A The Proposed BCD-Net Architecture

Each iteration of BCD-Net for DECT material decomposition consists of an image refining module and an MBID module. See the architecture of the proposed BCD-Net in Figure 1. Each image refining module of proposed BCD-Net has a sCNN architecture that consists of encoding convolution, nonlinear thresholding, and decoding convolution. The MBID cost function uses a weighted least-squares (WLS) data-fit term that models the material composition physics and noise statistics in the measurements, and a regularizer (or a prior term) that uses refined material images from an iteration-wise image refining module. In DECT, decomposing high- and low-energy attenuation images into two material images (water and bone) is the most conventional setup[37], so the section studies the proposed INN method with this perspective.

II.A.1 Image Refining Module

The first box in Figure 1 shows the architecture of proposed iteration-wise distinct cross-material CNNs. The ith image refining module of BCD-Net takes $\{\mathbf{x}_m^{(i-1)} \in \mathbb{R}^N : m = 1, 2\}$, decomposed material images at the $(i-1)$th iteration, and outputs refined material images $\{\mathbf{z}_m^{(i)} \in \mathbb{R}^N : m = 1, 2\}$, for $i = 1, \ldots, I_{\text{iter}}$, where I_{iter} is the number of BCD-Net iterations. Here, $\{\mathbf{x}_1, \mathbf{z}_1\}$, and $\{\mathbf{x}_2, \mathbf{z}_2\}$ denote water and bone images, respectively. We use the following sCNN architecture for each image refining module:

$$(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}) = \mathcal{R}_{\Theta^{(i)}}\left(\mathbf{x}_1^{(i-1)}, \mathbf{x}_2^{(i-1)}\right) = \begin{bmatrix} \sum_{k=1}^K \sum_{n=1}^2 \mathbf{d}_{1,n,k}^{(i)} \circledast \mathcal{T}_{\exp(\alpha_{n,k}^{(i)})}\left(\sum_{m=1}^2 \mathbf{e}_{n,m,k}^{(i)} \circledast \mathbf{x}_m^{(i-1)}\right) \\ \sum_{k=1}^K \sum_{n=1}^2 \mathbf{d}_{2,n,k}^{(i)} \circledast \mathcal{T}_{\exp(\alpha_{n,k}^{(i)})}\left(\sum_{m=1}^2 \mathbf{e}_{n,m,k}^{(i)} \circledast \mathbf{x}_m^{(i-1)}\right) \end{bmatrix}, \tag{1}$$

where $\Theta^{(i)}$ denotes a set of parameters of image refining module at the ith iteration, i.e., $\Theta^{(i)} = \{\mathbf{d}_{m,n,k}^{(i)}, \mathbf{e}_{n,m,k}^{(i)}, \alpha_{n,k}^{(i)} : k = 1, \ldots, K, m = 1, 2, n = 1, 2\}$, $\mathbf{d}_{m,n,k}^{(i)} \in \mathbb{R}^R$ and $\mathbf{e}_{n,m,k}^{(i)} \in \mathbb{R}^R$ are the kth decoding and encoding filters from the nth group of the mth material at the ith iteration, respectively, $\exp(\alpha_{m,k}^{(i)})$ is the kth thresholding value for the mth material at the ith iteration, K is the number of filters in each encoding and decoding structure for each material, and R is the size of filters, $\forall m, n, k, i$. In (1), the element-wise soft thresholding operator $\mathcal{T}_{\mathbf{a}}(\mathbf{b}) : \mathbb{R}^N \to \mathbb{R}^N$ is defined by

$$(\mathcal{T}_{\mathbf{a}}(\mathbf{b}))_j := \begin{cases} b_j - a_j \cdot \operatorname{sign}(b_j), & |b_j| > a_j \\ 0, & |b_j| \leq a_j, \end{cases} \tag{2}$$

for $j = 1, \ldots, N$. We use the exponential function to thresholding parameters $\{\alpha_{n,k}\}$ to avoid thresholding values being negative[30,32]. We will train distinct cross-material CNNs at each iteration to maximize the refinement performance.

The proposed CNN in (1) and the first box in Figure 1 consists of an individual encoding-decoding architecture for each material image, and crossover architectures between different material images. We encode or decode each feature at a hidden layer by two groups of encoding or decoding filters. For example, in Figure 1, input images $\mathbf{x}_1^{(i-1)}$ and $\mathbf{x}_2^{(i-1)}$ convolve with encoding filters $\mathbf{e}_{1,1,K}^{(i)}$ and $\mathbf{e}_{1,2,K}^{(i)}$, respectively (indicated by red and green), and then their thresholded sum gives encoded feature $\mathcal{T}_{\exp(\alpha_{1,K}^{(i)})}(\mathbf{e}_{1,1,K}^{(i)} * \mathbf{x}_1^{(i-1)} + \mathbf{e}_{1,2,K}^{(i)} * \mathbf{x}_2^{(i-1)})$. To decode the feature, we convolve this feature with two decoding filters $\mathbf{d}_{1,1,K}^{(i)}$ and $\mathbf{d}_{2,1,K}^{(i)}$ (indicated by purple and blue). One group of encoding or decoding filters is used to capture a feature of each material image individually, and the other group is used to capture correlations between different material images. When $n = m$, the filters in (1) form the individual encoding-decoding architecture that captures individual properties of the mth material, e.g., filters $\mathbf{e}_{1,1,K}^{(i)}$ and $\mathbf{d}_{1,1,K}^{(i)}$ (indicated by red and purple in Figure 1), whereas when $n \neq m$, these comprise the crossover architecture that exchanges information between two material images, e.g., filters $\mathbf{e}_{1,2,K}^{(i)}$ and $\mathbf{d}_{2,1,K}^{(i)}$ (indicated by green and blue in Figure 1). The crossover architecture is expected to be useful to remove noise and artifacts in material images.

II.A.2 MBID Module

The ith MBID module of BCD-Net in the second box of Figure 1 gives the decomposed material images, $\mathbf{x}^{(i)} = [(\mathbf{x}_1^{(i)})^\top, (\mathbf{x}_2^{(i)})^\top]^\top$, by reducing their deviations from attenuation maps $\mathbf{y} = [(\mathbf{y}_H)^\top, (\mathbf{y}_L)^\top]^\top \in \mathbb{R}^{2N}$ and refined material images $\mathbf{z}^{(i)} = [(\mathbf{z}_1^{(i)})^\top, (\mathbf{z}_2^{(i)})^\top]^\top, \forall i$, where $\mathbf{y}_H \in \mathbb{R}^N$ and $\mathbf{y}_L \in \mathbb{R}^N$ are attenuation maps at high and low energy, respectively. In particular, we reduce the deviation of model-based decomposition $\mathbf{x}^{(i)}$ from attenuation maps $\mathbf{y}$, using decomposition physics and noise statistics in $\mathbf{y}$. We formulate the MBID cost function by combining a WLS data-fit term and a regularizer using $\mathbf{z}^{(i)}$:

$$\mathbf{x}^{(i)} = \operatorname*{argmin}_{\mathbf{x} \in \mathbb{R}^{2N}} \frac{1}{2} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_{\mathbf{W}}^2 + \mathrm{G}(\mathbf{x}), \quad \mathrm{G}(\mathbf{x}) = \frac{\beta}{2} \|\mathbf{x} - \mathbf{z}^{(i)}\|_2^2. \tag{P0}$$

The mass attenuation coefficient matrix $\mathbf{A} \in \mathbb{R}^{2N\times 2N}$ is a Kronecker product of $\mathbf{A}_0$ and identity matrix $\mathbf{I}_N$, i.e., $\mathbf{A} = \mathbf{A}_0 \otimes \mathbf{I}_N$, and the matrix $\mathbf{A}_0 \in \mathbb{R}^{2\times 2}$ is defined as[19]:

$$\mathbf{A}_0 := \begin{bmatrix} \varphi_{1H} & \varphi_{2H} \\ \varphi_{1L} & \varphi_{2L} \end{bmatrix}, \tag{3}$$

in which φ_{mH} and φ_{mL} denote the mass attenuation coefficient of the mth material at high and low energy, respectively. In practice, these four values in matrix $\mathbf{A}_0$ can be calibrated in advance by $\varphi_{mH} = \mu_{mH}/\rho_m$ and $\varphi_{mL} = \mu_{mL}/\rho_m$, where ρ_m denotes the density of the mth material (we use theoretical values 1 g/cm^3 for water and 1.92 g/cm^3 for bone in our experiments), and μ_{mH} and μ_{mL} denote the linear attenuation coefficient of the mth material at high and low energy, respectively. To obtain μ_{mH} and μ_{mL}, we manually select a uniform area in $\mathbf{y}_H$ and $\mathbf{y}_L$ (e.g., water region and bone region) respectively and compute the average pixel value in this area[12]. The weight matrix $\mathbf{W} \in \mathbb{R}^{2N\times 2N}$ represented as $\mathbf{W} = \mathbf{W}_0 \otimes \mathbf{I}_N$ is block-diagonal by assuming the noise in each attenuation image are independent and identically distributed (i.i.d.) over pixels[15]. This noise assumption is widely used in practice[15,38–40]. Here, $\mathbf{W}_0$ is a 2×2 diagonal weight matrix with diagonal elements being the inverse of noise variance at high and low energies. The regularization parameter $\beta > 0$ controls the trade-off between noise and resolution in decompositions.

Based on the structures of matrices $\mathbf{A}$ and $\mathbf{W}$ above, we can separate the $\mathbf{x}$-update problem in (P0) into N subproblems. Then we obtain the following practical closed-form solution of $\mathbf{x}$ at each pixel j:

$$\mathbf{x}_j^{(i)} = (\mathbf{A}_0^\top \mathbf{W}_0 \mathbf{A}_0 + \beta \mathbf{I}_2)^{-1} (\mathbf{A}_0^\top \mathbf{W}_0 \mathbf{y}_j + \beta \mathbf{z}_j^{(i)}), \tag{4}$$

where $\mathbf{x}_j^{(i)} = (x_{1,j}^{(i)}, x_{2,j}^{(i)})^\top$ and $\mathbf{z}_j^{(i)} = (z_{1,j}^{(i)}, z_{2,j}^{(i)})^\top$ denote the water and bone density values of decomposed material images $\mathbf{x}^{(i)}$ and refined material images $\mathbf{z}^{(i)}$ at the jth pixel, respectively, and $\mathbf{y}_j = (y_{H,j}, y_{L,j})^\top$ denotes the high- and low-energy linear attenuation coefficients at the jth pixel, $j = 1, \ldots, N$. Due to small dimensions of matrices $\mathbf{A}_0^\top \mathbf{W}_0 \mathbf{A}_0$ and $\mathbf{I}_2$, the matrix inversion in (4) is efficient; the cost to compute $\{\mathbf{x}_j^{(i)} : \forall j\}$ scales as $O(N)$. Permuting $\{\mathbf{x}_j^{(i)} : \forall j\}$ gives the decomposed material images $\mathbf{x}^{(i)} = (x_{1,1}^{(i)}, \ldots, x_{1,N}^{(i)}, x_{2,1}^{(i)}, \ldots, x_{2,N}^{(i)})^\top$.

II.B Properties of the Proposed CNN Refiner

This section studies some properties of the proposed CNN (1) with the patch perspective. We rewrite (1) with the patch perspective as follows (we omit the iteration superscript indices

(i) for simplicity):

$$\mathcal{R}_\Theta(\mathbf{x}) \text{ in (1)} = \frac{1}{R}\sum_{j=1}^{N} \bar{\mathbf{P}}_j^\top \mathbf{D}\mathcal{T}_{\exp(\boldsymbol{\alpha})}(\mathbf{E}\bar{\mathbf{P}}_j\mathbf{x}), \tag{5}$$

where, $\bar{\mathbf{P}}_j = \mathbf{P}_j \oplus \mathbf{P}_j$, $\mathbf{P}_j \in \mathbb{R}^{R\times N}$ is the patch extraction operator for the jth pixel, $j = 1,\ldots,N$, $\oplus$ denotes the matrix direct sum, $\mathbf{D} \in \mathbb{R}^{2R\times 2K}$ and $\mathbf{E} \in \mathbb{R}^{2K\times 2R}$ are decoding and encoding filter matrices defined by:

$$\mathbf{D} := \begin{bmatrix} \mathbf{D}_{1,1} & \mathbf{D}_{1,2} \\ \mathbf{D}_{2,1} & \mathbf{D}_{2,2} \end{bmatrix} \text{ and } \mathbf{E} := \begin{bmatrix} \mathbf{E}_{1,1} & \mathbf{E}_{1,2} \\ \mathbf{E}_{2,1} & \mathbf{E}_{2,2} \end{bmatrix}, \tag{6}$$

where $\mathbf{D}_{m,n}$ and $\mathbf{E}_{n,m}$ are formed by grouping filters $\{\mathbf{d}_{m,n,k}\}$ and $\{\mathbf{e}_{n,m,k}\}$, respectively, i.e.,

$$\begin{aligned} \mathbf{D}_{m,n} &:= [\mathbf{d}_{m,n,1},\, \mathbf{d}_{m,n,2},\ldots,\mathbf{d}_{m,n,K}], \\ \mathbf{E}_{n,m} &:= [\mathbf{e}_{n,m,1},\, \mathbf{e}_{n,m,2},\,\ldots,\mathbf{e}_{n,m,K}]^\top, \quad m,n = 1,2, \end{aligned}$$

and $\boldsymbol{\alpha} = [\alpha_{1,1},\ldots,\alpha_{1,K},\alpha_{2,1},\ldots,\alpha_{2,K}]^\top \in \mathbb{R}^{2K}$ is a vector containing $2K$ thresholding parameters. We derived (5) using the convolution-to-patch reformulation technique[32]; see Proposition S.1 for more details.

Both of encoding and decoding filter matrices, $\mathbf{E}$ and $\mathbf{D}$, are composed of four smaller block matrices. The refiner of BCD-Net-sCNN-lc[35] uses only block matrices $\mathbf{E}_{1,1}$ and $\mathbf{E}_{1,1}^\top$ as encoding and decoding filters, respectively, for water images, and $\mathbf{E}_{2,2}$ and $\mathbf{E}_{2,2}^\top$ as the encoding and decoding filters, respectively, for bone images. Different from this, the proposed refiner of BCD-Net-sCNN-hc not only uses *distinct* encoding-decoding filters, but also additionally uses off-diagonal block matrices $\{\mathbf{D}_{1,2},\mathbf{D}_{2,1},\mathbf{E}_{1,2},\mathbf{E}_{2,1}\}$ to exploit correlations between the different material images. The crossover architecture captured via $\{\mathbf{D}_{1,2},\mathbf{D}_{2,1},\mathbf{E}_{1,2},\mathbf{E}_{2,1}\}$ models shared structures between water and bone images at the same spatial locations. When trained with some image denoising loss, the crossover architecture with thresholding operator (2) in BCD-Net-sCNN-hc is expected to better refine material images by exchanging shared noisy features between them, compared to the individual encoding-decoding case in BCD-Net-sCNN-lc.

We study the tight-frame property[41] of the proposed cross-material CNN refiners, since learned filters satisfying the tight-frame condition are useful to compact energy of input image and remove unwanted noise and artifacts via thresholding[18,42]. The tight-frame condition for (5) is given by

$$\mathbf{DE} = \mathbf{I}_{2R}. \tag{7}$$

This is implied as follows. Using the patch-perspective reformulation (5), convolutional encoding in (1) can be rewritten as follows: $\sqrt{1/R}[(\mathbf{E}\bar{\mathbf{P}}_1)^\top, \ldots, (\mathbf{E}\bar{\mathbf{P}}_N)^\top]^\top \mathbf{x}$. The tight-frame condition for a refiner that uses this as both encoder and decoder, i.e., (8) in Section II.C, is given as follows[18,42]: $\|\mathbf{x}\|^2 = \mathbf{x}^\top \sum_{j=1}^N \bar{\mathbf{P}}_j^\top \mathbf{E}^\top \mathbf{E}\bar{\mathbf{P}}_j \mathbf{x}/R, \quad \forall \mathbf{x}$. This condition is identical to $\mathbf{E}^\top \mathbf{E} = \mathbf{I}_{2R}$ considering that $\sum_{j=1}^N \bar{\mathbf{P}}_j^\top \bar{\mathbf{P}}_j = R\mathbf{I}_{2N}$ with the periodic boundary condition and sliding parameter 1. If a decoding filter matrix is different from an encoding filter matrix, e.g., (1), then the tight-frame condition can become (7). In Figure 2, we empirically observed for DECT material decomposition that sCNN-hc refiners of BCD-Net at the last iteration approximately satisfy the tight-frame condition.

Figure 3 shows learned filters of BCD-Net-sCNN-lc and BCD-Net-sCNN-hc refiners that use the identical encoding-decoding architecture, i.e., $\mathbf{D} = \mathbf{E}^\top$ in (5), where we display them with four groups, $\mathbf{E}_{1,1}$, $\mathbf{E}_{1,2}$, $\mathbf{E}_{2,1}$, and $\mathbf{E}_{2,2}$ in (6). Filters in diagonal block matrices on the left in Figure 3 include both (short) first-order finite differences and elongated features. In addition, $\mathbf{E}_{1,1}$ includes more elongated structures than $\mathbf{E}_{2,2}$, while $\mathbf{E}_{2,2}$ includes more first-order finite difference like kernels than $\mathbf{E}_{1,1}$ (there are 16 and 23 first-order finite difference like structures in $\mathbf{E}_{1,1}$ and $\mathbf{E}_{2,2}$, respectively). This is potentially because water image includes diverse low-contrast edge features from different soft-tissues, while bone image includes relatively simple high-contrast edge features from bone and air. Many structured kernels in $\mathbf{E}_{1,1}, \mathbf{E}_{1,2}, \mathbf{E}_{2,1}$, and $\mathbf{E}_{2,2}$, on the right in Figure 3 are like first-order finite difference: specifically, $\mathbf{E}_{1,1}$, $\mathbf{E}_{1,2}$, $\mathbf{E}_{2,1}$, and $\mathbf{E}_{2,2}$ have about 10, 17, 17, and 24 first-order finite difference like kernels. Interestingly, the number of first-order finite difference like kernels of $\mathbf{E}_{1,2}$ and $\mathbf{E}_{2,1}$ is intermediate between those of $\mathbf{E}_{1,1}$ and $\mathbf{E}_{2,2}$. This might imply using the conjecture above that cross-materials have less and more diverse edge features than water image and bone image, respectively. What is more, we observed some filters in $\mathbf{E}_{1,2}$ capture similar features as those in $\mathbf{E}_{1,1}$, e.g., filters indicated by red boxes, while some filters in $\mathbf{E}_{1,2}$ capture different features from those in $\mathbf{E}_{1,1}$, e.g., filters indicated by yellow boxes. We also observed similar behavior between $\mathbf{E}_{2,1}$ and $\mathbf{E}_{2,2}$.

II.C Variations of (1)

We specialize (1) to have simpler components. BCD-Net-sCNN-lc is a simpler convolutional encoding-decoding architecture proposed in our recent conference work[35]; it uses following CNN refiner that has identical encoding-decoding architecture independently for two different material images:

$$\mathbf{z}_m^{(i)} = \mathcal{R}_{\Theta_m^{(i)}}(\mathbf{x}_m^{(i-1)}) = \sum_{k=1}^{K} \bar{\mathbf{e}}_{m,m,k}^{(i)} \circledast \mathcal{T}_{\exp(\alpha_{m,k}^{(i)})}\left(\mathbf{e}_{m,m,k}^{(i)} \circledast \mathbf{x}_m^{(i-1)}\right), \quad m = 1, 2, \tag{8}$$

where $\bar{(\cdot)}$ rotates a filter (e.g., it rotates 2D filters by 180°). (1) specializes to (8) by setting $\mathbf{d}_{m,n,k}^{(i)}$ as $\bar{\mathbf{e}}_{n,m,k}^{(i)}$, and $\mathbf{e}_{n,m,k}^{(i)} = \mathbf{d}_{m,n,k}^{(i)} = \mathbf{0}$ for $m \neq n$. One can also use dCNNs instead of the sCNN refiners in (1) and (8). We refer to this method as BCD-Net-dCNN. We investigate the performance of BCD-Net-dCNN (that replaces the refining module in (1) and (8) with a dCNN); see Section III.B.3 later for details of BCD-Net-dCNN.

II.D Training BCD-Net-sCNNs

The training process at the ith iteration requires L input-output image pairs. Input labels are decomposed material images via MBID module, $\{\mathbf{x}_{l,m}^{(i-1)} : l = 1, \cdots, L\}$, and output labels are high-quality reference material images, $\{\mathbf{x}_{l,m} : l = 1, \cdots, L\}$. We use the patch-based training loss of $(1/L)\sum_{l=1}^{L} \|\mathbf{x}_l - \mathcal{R}_\Theta(\mathbf{x}_l^{(i-1)})\|_2^2$, where we derived their bound relation in Proposition S.2 using the convolution-to-patch loss reformulation techniques in a recent work[32]. Patch-based training first extracts reference and noisy material patches from $\{\mathbf{x}_{l,m} : l = 1, \cdots, L\}$ and $\{\mathbf{x}_{l,m}^{(i-1)} : l = 1, \cdots, L\}$ and constructs reference and noisy material data matrices $\widetilde{\mathbf{X}}_m \in \mathbb{R}^{R \times P}$ and $\widetilde{\mathbf{X}}_m^{(i-1)} \in \mathbb{R}^{R \times P}$, respectively, where $P = LN$. (For $\{\mathbf{x}_{l,m}^{(0)} : \forall l, m\}$, we used rough estimates of decomposed images obtained via the direct matrix inversion method (see Section III.A.1).) Then we construct paired multi-material data matrices $\widetilde{\mathbf{X}} \in \mathbb{R}^{2R \times P}$ and $\widetilde{\mathbf{X}}^{(i-1)} \in \mathbb{R}^{2R \times P}$, where each column is formed by stacking vectorized two-dimensional (2D) patches extracted from the same spatial location in different material images. i.e., $\widetilde{\mathbf{X}} = [\widetilde{\mathbf{X}}_1^\top, \widetilde{\mathbf{X}}_2^\top]^\top$ and $\widetilde{\mathbf{X}}^{(i-1)} = [(\widetilde{\mathbf{X}}_1^{(i-1)})^\top, (\widetilde{\mathbf{X}}_2^{(i-1)})^\top]^\top$.

The training loss of BCD-Net-sCNN-hc at the ith iteration is

$$\mathcal{L}(\mathbf{D}, \mathbf{E}, \boldsymbol{\alpha}) := \frac{1}{P}\|\widetilde{\mathbf{X}} - \mathbf{D}\mathcal{T}_{\exp(\boldsymbol{\alpha})}(\mathbf{E}\widetilde{\mathbf{X}}^{(i-1)})\|_{\mathrm{F}}^2, \tag{P1}$$

where $\|\cdot\|_{\mathrm{F}}$ denotes the Frobenius norm of a matrix. The subgradients of $\mathcal{L}(\mathbf{D}, \mathbf{E}, \boldsymbol{\alpha})$ with

Algorithm 1 Training BCD-Net-sCNN-hc

Require: $\{\mathbf{x}_{l,m}, \mathbf{x}_{l,m}^{(0)}, \mathbf{y}_l, \mathbf{A}_l, \mathbf{W}_l : l = 1, \ldots, L, m = 1, 2\}, \beta > 0, I_{\text{iter}} > 0$
  for $i = 1, 2, \cdots, I_{\text{iter}}$ **do**
    Train $\Theta^{(i)}$ via (P1) using $\{\mathbf{x}_{l,m}, \mathbf{x}_{l,m}^{(i-1)} : \forall l, m\}$
    for $l = 1, \ldots, L$ **do**
      Refining: $(\mathbf{z}_{l,1}^{(i)}, \mathbf{z}_{l,2}^{(i)}) = \mathcal{R}_{\Theta^{(i)}}(\mathbf{x}_{l,1}^{(i-1)}, \mathbf{x}_{l,2}^{(i-1)})$ in (1).
      MBID: Obtain $\{\mathbf{x}_{l,m}^{(i)} : \forall l, m\}$ by solving (P0) with (4).
    end for
  end for

respect to $\mathbf{D}$, $\mathbf{E}$, and $\boldsymbol{\alpha}$ for each mini-batch selection are as follows:

$$\frac{\partial \mathcal{L}(\mathbf{D}, \mathbf{E}, \boldsymbol{\alpha})}{\partial \mathbf{D}} = -\frac{2}{B}\left(\mathbf{X} - \mathbf{D}\mathbf{Z}^{(i-1)}\right)\mathbf{Z}^{(i-1)\top} \tag{9}$$

$$\frac{\partial \mathcal{L}(\mathbf{D}, \mathbf{E}, \boldsymbol{\alpha})}{\partial \mathbf{E}} = -\frac{2}{B}\mathbf{D}^\top\left(\mathbf{X} - \mathbf{D}\mathbf{Z}^{(i-1)}\right) \odot \mathbb{1}_{|\mathbf{E}\mathbf{X}^{(i-1)}| > \exp(\boldsymbol{\alpha}\mathbf{1}')} \cdot \mathbf{X}^{(i-1)\top} \tag{10}$$

$$\frac{\partial \mathcal{L}(\mathbf{D}, \mathbf{E}, \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}} = \frac{2}{B}\left\{\mathbf{D}^\top\left(\mathbf{X} - \mathbf{D}\mathbf{Z}^{(i-1)}\right) \odot \exp(\boldsymbol{\alpha}\mathbf{1}') \odot \operatorname{sign}\left(\mathbf{Z}^{(i-1)}\right)\right\}\mathbf{1}, \tag{11}$$

where $\mathbf{X}$, $\mathbf{X}^{(i-1)} \in \mathbb{R}^{2R\times B}$ are mini-batch in which columns are randomly selected from $\widetilde{\mathbf{X}}$ and $\widetilde{\mathbf{X}}^{(i-1)}$, respectively, $\mathbf{Z}^{(i-1)} = \mathcal{T}_{\exp(\boldsymbol{\alpha}\mathbf{1}')}(\mathbf{E}\mathbf{X}^{(i-1)})$, and B is the mini-batch size. Here, $\mathbf{1} \in \mathbb{R}^{B\times 1}$ denotes a column vector of ones, $\mathbb{1}_{(\cdot)}$ is the indicator function (value 0 when condition is violated and 1 otherwise), and $\odot$ is the element-wise multiplication. The derivation details of (9)–(11) are in Section S.I. Once we obtain the learned filters and thresholding values, we apply them to refine material images. These refined images are then fed into the MBID module. Algorithm 1 shows the training process of BCD-Net-sCNN-hc.

Training BCD-Net-sCNN-lc only involves submatrices $\mathbf{E}_{1,1}^{(i)}$ and $\mathbf{E}_{2,2}^{(i)}$, i.e., $\mathbf{E}_{1,2}^{(i)} = \mathbf{E}_{2,1}^{(i)} = \mathbf{D}_{1,2}^{(i)} = \mathbf{D}_{2,1}^{(i)} = \mathbf{0}$, $\mathbf{D}_{1,1}^{(i)} = \mathbf{E}_{1,1}^{(i)\top}$, and $\mathbf{D}_{2,2}^{(i)} = \mathbf{E}_{2,2}^{(i)\top}$ in (P1), and we train it using image pair $(\widetilde{\mathbf{X}}_m, \widetilde{\mathbf{X}}_m^{(i-1)})$, $\forall m, i$. See subgradients for training BCD-Net-sCNN-lc in our earlier conference work[35].

II.E Testing Trained BCD-Nets

At the ith iteration of BCD-Net-sCNN-hc, we apply learned filters and thresholding parameters $\Theta^{(i)}$ to noisy material images $\{\mathbf{x}_m^{(i-1)} : m = 1, 2\}$ to obtain refined material images $\mathbf{z}^{(i)} = \mathcal{R}_{\Theta^{(i)}}(\mathbf{x}_1^{(i-1)}, \mathbf{x}_2^{(i-1)})$, where the definition of $\mathbf{z}^{(i)}$ is given in Section II.A.2. We then feed these refined images into the MBID module to obtain decomposed material images

Algorithm 2 Testing Trained BCD-Net-sCNN-hc

Input: $\{\mathbf{x}_m^{(0)} : m = 1, 2\}, \mathbf{y}, \mathbf{A}, \mathbf{W}, \{\Theta^{(i)} : i = 1, \ldots, I_{\text{iter}}\}, \beta > 0$
Output: $\{\mathbf{x}_m^{(I_{\text{iter}})} : m = 1, 2\}$
for $i = 1, 2, \cdots, I_{\text{iter}}$ **do**
  Refining: $(\mathbf{z}_1^{(i)}, \mathbf{z}_2^{(i)}) = \mathcal{R}_{\Theta^{(i)}}(\mathbf{x}_1^{(i-1)}, \mathbf{x}_2^{(i-1)})$ in (1).
  MBID: Obtain $\{\mathbf{x}_m^{(i)} : m = 1, 2\}$ by solving (P0) with (4).
end for

$\{\mathbf{x}_m^{(i)} : m = 1, 2\}$. After some fixed iterations (where I_{iter} is chosen in training), BCD-Net-sCNN-hc gives the final decomposed images $\{\mathbf{x}_m^{(I_{\text{iter}})} : m = 1, 2\}$. Algorithm 2 summarizes the test process of learned BCD-Net-sCNN-hc. The test process of BCD-Net-sCNN-lc and BCD-Net-dCNN are similar to that of BCD-Net-sCNN-hc.

III Results and Discussions

This section describes experimental setup and reports comparison results with XCAT phantom[43] and clinical DECT head data. We compared the performances of three BCD-Net methods (BCD-Net-sCNN-lc[35], BCD-Net-sCNN-hc, and BCD-Net-dCNN), the conventional direct matrix inversion method, MBID methods using data-driven and conventional non-data-driven regularizers, DECT-ST[19] and DECT-EP[15], and a (noniterative) dCNN method.

III.A Methods for Comparisons

This section describes methods compared with the proposed BCD-Net methods. We will describe their parameters in the next section.

III.A.1 Direct Matrix Inversion

This conventional method solves (P0) with $\mathrm{G}(\mathbf{x}) = 0$ by matrix inversion, i.e., $\mathbf{A}^{-1}\mathbf{y}$. We use direct matrix inversion results as initial material decomposition to DECT-EP and BCD-Nets, i.e., $\{\mathbf{x}^{(0)} = \mathbf{A}^{-1}\mathbf{y}\}$, and noisy input material images to dCNN denoiser.

III.A.2 DECT-EP

This conventional method solves (P0) with a material-wise edge-preserving regularizer that is defined as $\mathrm{G}_{\text{EP}}(\mathbf{x}) = \sum_{m=1}^{2} \beta_m \mathrm{G}_m(\mathbf{x}_m)$, where the mth material regularizer is $\mathrm{G}_m(\mathbf{x}_m) = \sum_{j=1}^{N} \sum_{k \in S} \psi_m(x_{m,j} - x_{m,k})$, and S is a list of indices that correspond to neighboring pixels of a pixel $x_{m,j}$ with $|S| = R_{\text{EP}}, \forall m, j$, where R_{EP} denotes the number of neighbors for each pixel. Here, the potential function is $\psi_m(t) \triangleq \frac{\delta_m^2}{3}(\sqrt{1 + 3(t/\delta_m)^2} - 1)$ with the mth material EP parameter, δ_m. We chose β_m and δ_m for different materials separately to achieve the desired boundary sharpness and strength of smoothness.

III.A.3 DECT-ST

This data-driven method solves (P0) with a regularizer that uses two square material-wise sparsifying transforms trained in an unsupervised way. The regularizer $\mathrm{G}_{\mathrm{ST}}(\mathbf{x})$ is defined as

$$\mathrm{G}_{\mathrm{ST}}(\mathbf{x}) \triangleq \min_{\{\mathbf{z}_{m,j}\}} \sum_{m=1}^{2} \sum_{j=1}^{N} \beta_m \left\{ \|\boldsymbol{\Omega}_m \mathbf{P}_{m,j}\mathbf{x} - \mathbf{z}_{m,j}\|_2^2 + \gamma_m^2 \|\mathbf{z}_{m,j}\|_0 \right\},$$

where $\boldsymbol{\Omega}_1 \in \mathbb{R}^{R_{\mathrm{ST}} \times R_{\mathrm{ST}}}$ and $\boldsymbol{\Omega}_2 \in \mathbb{R}^{R_{\mathrm{ST}} \times R_{\mathrm{ST}}}$ are pre-learned transforms for water and bone, respectively, $\mathbf{P}_{m,j}\mathbf{x}$ and $\mathbf{z}_{m,j}$ denote the jth patch of the mth material image and corresponding sparse vector, respectively, and R_{ST} is the number of pixels in each patch.

III.A.4 dCNN denoiser

The (noniterative) image denoising dCNN method uses two input and output channels; specifically, it takes noisy water and bone images and provides denoised water and bone images. The architecture that maps from noisy material images to true material images corresponds to the second CNN architecture of the cascaded dCNN[24], and that uses two input and two output channels corresponds to the setup of a modified U-Net method[27].

III.B Experimental Setup

III.B.1 Imaging setup for XCAT phantom experiments

We used 1024×1024 material images with pixel size 0.49×0.49 mm^2 of the XCAT phantom in our imaging simulation. We generated noisy (Poisson noise) sinograms of size 888 (radial samples) $\times$ 984 (angular views) using GE LightSpeed X-ray CT fan-beam system geometry corresponding to a poly-energetic source at 80 kVp and 140 kVp with 1.86×10^5 and 1×10^6 incident photons per ray, respectively. We used FBP method to reconstruct 2D high- and low-energy attenuation images of size 512×512 with a coarser pixel size 0.98×0.98 mm^2 to avoid an inverse crime. Figure 4 displays the attenuation images for a test slice.

III.B.2 Data construction

We separated each 1024×1024 slice of the original XCAT phantom into water and bone images according to the table of linear attenuation coefficients for organs provided for the XCAT phantom. We manually grouped fat, muscle, water, and blood into the water density images, and rib bone and spine bone into bone density images. We then downsampled these material density images to size 512×512 by linear averaging to generate ground truths of the decomposed material images. We chose 13 slices from the XCAT phantom, among which $L = 10$ slices were used for training the proposed BCD-Net-sCNNs, and remaining 3 slices were used for testing. Testing phantom images are sufficiently different from training phantom images; specifically, they are at a minimum ≈ 1.5 cm away, i.e., 25 slices. For dCNN, we used $L = 20$ slices of XCAT phantom that includes the 10 slices chosen for training the proposed BCD-Net-sCNNs. In general, dCNNs need many training samples, so we used more image pairs to train dCNN compared to BCD-Net-sCNN-lc and BCD-Net-sCNN-hc.

In addition, using the clinical data, we evaluated the proposed methods and compared them to the methods in Section III.A. The clinical data experiments decomposed a mixture into two constituent materials, water and bone, in each pixel. The patient head data was obtained by Siemens SOMATOM Definition flash CT scanner using dual-energy CT imaging protocols. The protocols of this head data acquisition are listed in Table 1. For dual-energy data acquisition, the dual-energy source were set at 140 kVp and 80 kVp. Figure 8 shows attenuation images of head data. FBP method was used to reconstruct these attenuation images.

III.B.3 Methods setup and parameters

We first obtained the low-quality material images from high- and low-energy attenuation images using direct matrix inversion method, and used these results to initialize DECT-EP method. We used the 8-neighborhood system, $R_{\mathrm{EP}} = 8$. To ensure convergence, we ran DECT-EP with 500 iterations. For XCAT phantom, we set $\{\beta_m, \delta_m : m = 1, 2\}$ as $\{2^8, 0.01\}$ and $\{2^{8.5}, 0.02\}$ for water and bone, respectively; for patient head data, we set them as $\{2^{10.5}, 0.008\}$ and $\{2^{11}, 0.015\}$ for water and bone, respectively.

We pre-learned two sparsifying transform matrices of size $R_{\mathrm{ST}}^2 = 64^2$ with ten slices (same slices as used in training BCD-Net-sCNNs) of true water and bone images of the XCAT phantom, using the suggested algorithm and parameter set (including number of iterations, regularization parameters, transform initialization, etc.) in the original paper[19]. We initialized DECT-ST using decomposed images obtained by DECT-EP method. We tuned the parameters $\{\beta_1, \beta_2, \gamma_1, \gamma_2\}$ and set them as $\{50, 70, 0.03, 0.04\}$ for XCAT phantom, and $\{150, 200, 0.012, 0.024\}$ for patient head data.

For the denoising dCNN architecture, we set the number of layers and number of features in hidden layers as 4 and 64, respectively. We did not use batch normalization and bias because the pixel values of different training/testing images are of the same scale. We learned the dCNN denoiser $\mathcal{R}$ with the standard loss in image denoising, $\mathcal{L}(\mathcal{R}) = \frac{1}{L}\sum_{l=1}^{L} \|\mathbf{x}_l - \mathcal{R}(\mathbf{x}_l^{(0)})\|_2^2$, with Adam using 200 epochs and batch size 1. We observed with the clinical data that selected dCNN architecture gives better decomposed image quality, compared to its variants with 8 layers and/or the different mode that maps high- and low-energy attenuation images to two material images (this mode corresponds to a series of papers[25–27]).

We trained a 100-iteration BCD-Net-sCNN-hc and a 100-iteration BCD-Net-sCNN-lc with image refining CNN architectures in (1) and (8), respectively. For BCD-Net-sCNN-hc, we trained cross-material CNN refiners in (1) with about 1×10^6 paired stacked multi-material patches. We trained $8K = 512$ filters of size $R = 8 \times 8$ at each iteration. For BCD-Net-sCNN-lc, we trained convolutional refiners in (8) for each material with about 1×10^6 paired patches. We trained $K = 64$ filters of size $R = 8 \times 8$ for each material at each iteration. We initialized all filters with values randomly generated from a Gaussian distribution with a zero mean and standard deviation of 0.1. We found in training that thresholding value initialization is important to ensure stable performances. For BCD-Net-sCNN-lc, we set initial thresholding parameters before applying the exponential function as log(0.88) and log(0.8) for water and bone, respectively; for BCD-Net-sCNN-hc, we set them as log(0.88). The regularization parameter β balances data-fit term and the prior estimate from image refining module. To achieve the best image quality and decomposition accuracy, we set β as 600 and 6400 for BCD-Net-sCNN-lc and BCD-Net-sCNN-hc, respectively (note that different BCD-Net architectures have different refining performance). We train NNs of BCD-Net-sCNN-hc and BCD-Net-sCNN-lc with Adam[44] using the default hyper-parameters and tuned learning rate of 3×10^{-4}. We applied the learning rate schedule that decreases learning rates by a ratio of 90% every five epochs. We set batch size and number of epochs as $B = 10000$ and 50, respectively. For patient head data, we used the learned filters and thresholding values with XCAT phantom. The attenuation maps of XCAT phantom and clinical head data were generated by different energy spectrum and dose, and the clinical head data is much more complex than the XCAT phantom (see Figures 4 and 8). We thus set different regularization parameter β for the patient head data to achieve the best image quality; specifically, we set β as 3000 and 12000 in testing BCD-Net-sCNN-lc and BCD-Net-sCNN-hc, respectively.

We trained a 100-iteration BCD-Net-dCNN, where we replaced image refining CNN architecture of BCD-Net-sCNN-hc with the aforementioned denoising dCNN architecture. We used the same training dataset used in training the non-iterative dCNN method. We also used Adam optimization and identical settings (learning rate and regularization parameter β) as those of BCD-Net-sCNN-hc. We set batch size and number of epochs as 1 and 10, respectively. We observed with three test phantom samples that BCD-Net-dCNN becomes overfitted around 40th iteration; see Figure S.1. We thus used the results at the 40th iteration for test phantom samples. For the patient head data, we used 40-iteration BCD-Net-dCNN learned with XCAT phantom. We set β as 2400 after fine tuning to achieve the best image quality.

III.B.4 Evaluation metrics

In the quantitative evaluations with the XCAT phantom, we computed root-mean-square error (RMSE) for decomposed material images within a region of interest (ROI). We set the ROI as a circle region that includes all the phantom tissue. For a decomposed material density image $\hat{\mathbf{x}}_m$, the RMSE in density (g/cm^3) is defined as $\sqrt{\sum_{j=1}^{N_{\mathrm{ROI}}} (\hat{x}_{m,j} - x^\star_{m,j})^2 / N_{\mathrm{ROI}}}$, where $x^\star_{m,j}$ denotes the true density value of the mth material at the jth pixel location, and N_{ROI} is the number of pixels in a ROI. The ROI is indicated in red circle in Figure 5(a).

For the patient head data, we evaluated each method with *1)* contrast-to-noise ratio (CNR) that measures the contrast between tissue of interest (TOI) and local background region, and *2)* noise power spectrum (NPS)[45] that measures noise properties, in decomposed water images. CNR is defined as $\mathrm{CNR} = (\mu_{\mathrm{TOI}} - \mu_{\mathrm{BKG}})/\sigma_{\mathrm{BKG}}$, where μ_{TOI} and μ_{BKG} are mean values in a TOI and local background region, respectively, and σ_{BKG} is standard deviation between pixel values in a local background region. We selected three TOI-local background sets in muscle and fat areas; see red and blue regions in Figure 5(b). The NPS is defined as $\mathrm{NPS} = |\mathrm{DFT}\{f\}|^2$, where f denotes the noise of a ROI of decomposed water image (the patient head data does not have the ground-truth, so we subtract the mean value from the pixel values to approximate noise[45]), and $\mathrm{DFT}\{\cdot\}$ applies the 2D discrete Fourier transform (DFT) to 2D image. We selected three ROIs with uniform intensity and of size 30×30 in decomposed water image, and measured NPS within these ROIs; see the positions of three ROIs in Figure 5(c).

We used the most conventional measures for image quality assessment in tomography research. In XCAT phantom experiments with available ground-truth material images, we calculated RMSE values for each method. In clinical data experiments, we used the CNR measure that is the most widely-used alternative to RMSE in tomography research particularly when ground-truths are unavailable.

III.C Comparisons Between Different Methods with XCAT Phantom Data

Table 2 summarizes the RMSE values of material images decomposed by different methods for three different test slices. BCD-Net-sCNN-lc significantly decreases RMSE for material images compared to direct matrix inversion, DECT-EP, and DECT-ST. For all test samples, BCD-Net-sCNN-hc achieves significantly lower RMSE values compared to BCD-Net-sCNN-lc, implying the superiority of the distinct cross-material CNN architecture in (1) over the identical encoding-decoding architecture in (8). BCD-Net-sCNN-hc and dCNN methods achieve comparable errors: BCD-Net-sCNN-hc achieves an average $0.4 \times 10^{-3}\,\mathrm{g/cm^3}$ improvement for water images over dCNN, while dCNN achieves an average $0.2 \times 10^{-3}\,\mathrm{g/cm^3}$ improvement for bone images over BCD-Net-sCNN-hc. Compared to BCD-Net-dCNN, BCD-Net-sCNN-hc gives higher average RMSE for bone images, and the same average RMSE for water images. Compared to dCNN, BCD-Net-dCNN achieves RMSE improvements for both water and bone images, implying that dCNN denoisers combined with MBID modules in an iterative way can further decrease RMSE values. Figure 6 shows the RMSE convergence behavior of BCD-Net-sCNN-hc: it decreases monotonically. (See its fixed point convergence guarantee in the work[32].)

Figure 7 shows the #1 material density images of direct matrix inversion, DECT-EP, DECT-ST, dCNN, BCD-Net-sCNN-lc, BCD-Net-sCNN-hc, BCD-Net-dCNN, and ground truth. DECT-EP reduces severe noise and artifacts in direct matrix inversion decompositions. DECT-ST, dCNN, and BCD-Net-sCNN-lc significantly improve the image quality compared to DECT-EP, but still have some obvious artifacts. Compared to dCNN, BCD-Net-dCNN further reduces noise and artifacts and shows better recovery of the areas at the boundaries of water and bone; however, BCD-Net-dCNN still blurs soft-tissue regions. Compared to DECT-ST, dCNN, BCD-Net-sCNN-lc, and BCD-Net-dCNN, BCD-Net-sCNN-hc shows significantly better noise and artifacts reduction while improving the sharpness of edges in soft-tissue regions. These improvements are clearly noticeable in the zoom-ins of water images. Decomposed material images for another two test slices are included in Figures S.3–S.4.

III.D Comparisons Between Different Methods with Patient Data

Figure 8 shows decomposed material density images by different methods and high- and low-energy attenuation images for clinical head data. DECT-EP reduces severe noise and artifacts in direct matrix inversion results, but it is difficult to distinguish edges in many soft tissue regions. DECT-ST and dCNN suppress noise and improve the edges in soft tissues compared to DECT-EP, but both still have poor contrast in many soft tissue regions. BCD-Net-sCNN-lc and BCD-Net-dCNN further improve the contrast in soft tissue regions compared to DECT-ST and dCNN. However, BCD-Net-sCNN-lc has bright artifacts—see the bottom-right zoom-in in water image—and BCD-Net-dCNN leads to indistinguishable bone marrow structures—see the bottom-left zoom-ins in water and bone images. BCD-Net-sCNN-hc better removes noise and artifacts, provides clearer image edges and structures, and recovers subtle details, compared to the other methods aforementioned. One clearly noticeable improvement is captured in the bottom-right zoom-ins in water images, where BCD-Net-sCNN-hc not only improves edge sharpness and contrast in soft tissue, but also suppresses bright artifacts. Inside the red circle 1 in water images, BCD-Net-sCNN-hc and BCD-Net-dCNN preserve a "dark spot" that exists in attenuation images, whereas DECT-EP, DECT-ST, dCNN, and BCD-Net-sCNN-lc all missed it. The structure of the dark spot is an artery that contains diluted iodine solution caused by angiogram. The linear attenuation coefficient of iodine is much closer to bone than soft-tissue. During decomposition, most of the iodine is grouped into the bone image, while in the water image there are only some pixels with tiny values, thus it is a dark spot. Moreover, the marrow structures obtained by BCD-Net-sCNN-hc have sharper edges (inside red circle 2) than the other methods.

Table 3 summaries the CNR values for the three different TOI-local background sets in the decomposed water images via different methods. BCD-Net-sCNN-hc achieves significantly higher CNR compared to the other methods for all the three TOI-local background sets, and the performance degrades in the following order: BCD-Net-dCNN, BCD-Net-sCNN-lc, dCNN, DECT-ST, DECT-EP, direct matrix inversion. In particular, BCD-Net-sCNN-hc achieves 1.70 improvement in CNR in average over BCD-Net-dCNN, and BCD-Net-dCNN achieves 3.14 improvement in CNR in average over dCNN.

Figure 9 compares the magnitude of NPS from different methods. Across all frequencies, the NPS magnitude of BCD-Net-sCNN-hc is significantly smaller than that of direct matrix inversion, DECT-EP, DECT-ST, and dCNN. The overall low-frequency noise of BCD-Net-sCNN-hc is also significantly less than that of the aforementioned methods. What is more, BCD-Net-sCNN-hc achieves fewer vertical and horizontal frequency strips with lower intensity compared to BCD-Net-sCNN-lc and BCD-Net-dCNN, especially in the ROI #1 and #3. The aforementioned NPS comparisons demonstrate the superiority of the proposed BCD-Net-sCNN-hc in removing noise and artifacts inside soft tissue regions. We observed similar trends in averaged NPS measures using multiple noise realizations; see Figure S.2.

Similar to XCAT phantom results, the dCNN denoiser and BCD-Net-dCNN give less appealing material images of the clinical head data, compared to the proposed BCD-Net-sCNN-hc. We conjecture that the following reasons may limit the dCNN denoising performance: lack of considering decomposition physics and/or limited training samples and diversity. Although BCD-Net-dCNN incorporates decomposition physics, due to too high NN complexity (compared to the diversity of the training data), the image quality for both phantom and patient head data are still unsatisfactory. The proposed method, BCD-Net-sCNN-hc, resolves the issues of dCNN and BCD-Net-dCNN by using both MBID cost minimization and shallow CNN refiner at each iteration. The clinical head data shows that the proposed BCD-Net-sCNN-hc successfully reduces noise/artifacts and preserves subtle details that exist in attenuation images in Figure 8.

III.E Computational Complexity Comparisons

The computational cost of DECT-EP, DECT-ST, and the proposed BCD-Net-sCNNs scale as $O(R_{\mathrm{EP}} N I_{\mathrm{EP}})$, $O((R_{\mathrm{ST}})^2 N I_{\mathrm{ST}})$, and $O(RKNI_{\mathrm{iter}})$, respectively, where I_{EP} and I_{ST} are the number of iterations for DECT-EP and DECT-ST, respectively. The computational cost of the chosen dCNN architecture in Section III.A.4 and BCD-Net-dCNN scale as $O(R_{\mathrm{dCNN}} K_{\mathrm{dCNN}} N((C-2)K_{\mathrm{dCNN}}+4))$ and $O(R_{\mathrm{dCNN}} K_{\mathrm{dCNN}} N((C-2)K_{\mathrm{dCNN}}+4)I_{\mathrm{dCNN}})$, respectively, where R_{dCNN}, K_{dCNN}, and C are kernel size, the number of features, and the number of convolutional layers of dCNN denoiser, respectively, and I_{dCNN} is the number of BCD-Net-dCNN iterations. In all experiments, we used $R_{\mathrm{EP}} = 8$ and $I_{\mathrm{EP}} = 500$ for DECT-EP, $R_{\mathrm{ST}} = 64$ and $I_{\mathrm{ST}} = 1000$ for DECT-ST, $R_{\mathrm{dCNN}} = 3^2$, $K_{\mathrm{dCNN}} = 64$, and $C = 4$ for dCNN denoiser, $I_{\mathrm{dCNN}} = 40$ for BCD-Net-dCNN, and $R = K = 8^2$ and $I_{\mathrm{iter}} = 100$ for the proposed BCD-Net-sCNN-hc. The big-O analysis reveals that the computational cost of 100-iteration of the proposed BCD-Net-sCNN-hc is larger than 500-iteration DECT-EP and the chosen dCNN denoiser, 87% cheaper than that of 40-iteration BCD-Net-dCNN, and 90% cheaper than that of 1000-iteration DECT-ST.

III.F Discussions for Generalization Performance of dCNN, BCD-Net-dCNN, and BCD-Net-sCNN-hc

To study the generalization performance of dCNN, BCD-Net-dCNN, and BCD-Net-sCNN-hc, we calculated the average RMSE values from training and test samples, and their difference. Table 4 reports the RMSE gap between decomposed images in training and test via dCNN, BCD-Net-dCNN, and BCD-Net-sCNN-hc. BCD-Net-dCNN has smaller RMSE gap for both water and bone images, compared to dCNN that lacks decomposition physics. We conjecture that including MBID modules in an iterative way can improve the generalization performance of dCNN denoisers. This result is well aligned with the recent work[46] demonstrating that combining deep NNs, imaging physics, and sparisty-promoting regularizer gives the stable performance against perturbations. BCD-Net-sCNN-hc has smaller RMSE gap for both water and bone images, compared to BCD-Net-dCNN. At each BCD-Net iteration, the number of trainable parameters are $2K(4R+1)$ and $R_{\mathrm{dCNN}}K_{\mathrm{dCNN}}((C-2)K_{\mathrm{dCNN}}+4)$ for BCD-Net-sCNN-hc and BCD-Net-dCNN, respectively; specifically, they are 32,896 and 76,032 using the parameter sets in Section III.E. We conjecture that sCNN-hc refiner with lower NN complexity can improve the generalization performance over dCNN refiner.

IV Conclusions

Image-domain decomposition methods are readily applicable to commercial DECT scanners, but susceptible to noise and artifacts on attenuation images. To improve MBID performance, it is important to incorporate accurate prior knowledge into sophisticatedly designed MBID. The proposed INN architecture, BCD-Net-sCNN-hc, successfully achieves accurate MBID by providing accurate prior knowledge via its iteration-wise refiners that exploit correlations between different material images with distinct encoding-decoding filters. Our study with patch-based reformulation reveals that learned filters of distinct cross-material CNN refiners can approximately satisfy the tight-frame condition and useful for noise suppression and signal restoration. On both XCAT phantom and patient head data, the proposed BCD-

Net-sCNN-hc reduces the artifacts at boundaries of materials and improves edge sharpness and contrast in soft tissue, compared to a conventional MBID method, DECT-EP, a recent unsupervised MBID method, DECT-ST, and a noniterative dCNN method. We also show that BCD-Net-sCNN-hc improves the image quality over BCD-Net-dCNN, especially for patient head data, potentially due to its lower refiner complexity over that of BCD-Net-dCNN. For choosing refiner architecture in BCD-Net, we suggest considering the number of trainable parameters with the size/diversity of training data.

There are a number of avenues for future work. Our first future work is to investigate a three-material decomposition BCD-Net architecture in DECT; see its potential benefit in Section S.III and Figures S.5–S.7. Second, to further improve the MBID model, we plan to train the weight matrix $\mathbf{W}_0$ in (P0) in a supervised way with proper loss function designs, rather than statistically estimating it. By extending the patch-perspective interpretations, we will develop an "explainable" deeper refiner that might further improve the MBID performance of BCD-Net. Third, to accommodate the non-trivial tuning process of β in (P0), we plan to learn it from training datasets. Finally, to further improve the generalization capability of the proposed INN architecture, we will additionally incorporate a sparsity-promoting regularizer into the proposed framework, similar to the recent work[46].

V Acknowledgement

The authors thank Dr. Tianye Niu, Shenzhen Bay Laboratory, for providing clinical DECT images for our experiments.

VI Conflict of Interest Statement

The authors have no relevant conflicts of interest to disclose.

VII Data Availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

[1] N. Hokamp, S. Lennartz, J. Salem, et al. Dose independent characterization of renal stones by means of dual energy computed tomography and machine learning: an ex-vivo study. *European Radiology*, 30(3):1397–1404, 2020.

[2] M. C. Jacobsen, E. N. Cressman, E. P. Tamm, et al. Dual-energy CT: lower limits of iodine detection and quantification. *Radiology*, 292(2):414–419, 2019.

[3] Y. Li, G. Shi, S. Wang, S. Wang, and R. Wu. Iodine quantification with dual-energy CT: phantom study and preliminary experience with VX2 residual tumour in rabbits after radiofrequency ablation. *British Journal of Radiology*, 86(1029):143–151, 2013.

[4] Y. Liu, J. Cheng, Z. Chen, and Y. Xing. Feasibility study: Low-cost dual energy CT for security inspection. In *Proc. IEEE Nuc. Sci. Symp. Med. Im. Conf.*, pages 879–882, 2010.

[5] L. Martin, A. Tuysuzoglu, W. C. Karl, and P. Ishwar. Learning-based object identification and segmentation using dual-energy CT images for security. *IEEE Trans. Im. Proc.*, 24(11):4069–4081, 2015.

[6] Philip Engler and William D. Friedman. Review of dual-energy computed tomography techniques. *Materials Evaluation*, 48(5):623–629, 1990.

[7] P. R. S. Mendonca, P. Lamb, and D. Sahani. A flexible method for multi-material decomposition of dual-energy CT images. *IEEE Trans. Med. Imag.*, 33(1):99–116, 2014.

[8] W. Wu, Q. Wang, F. Liu, Y. Zhu, and H. Yu. Block matching frame based material reconstruction for spectral CT. *Phys. Med. Biol.*, 64(23):235011, 2019.

[9] W. Wu, D. Hu, K. An, S. Wang, and F. Luo. A high-quality photon-counting CT technique based on weight adaptive total-variation and image-spectral tensor factorization for small animals imaging. *IEEE Transactions on Instrumentation and Measurement*, 70(25):427–31, 2020.

[10] Y. Long and J. A. Fessler. Multi-material decomposition using statistical image reconstruction for spectral CT. *IEEE Trans. Med. Imag.*, 33(8):1614–1626, August 2014.

[11] J. Noh, J. A. Fessler, and P. E. Kinahan. Statistical sinogram restoration in dual-energy CT for PET attenuation correction. *IEEE Trans. Med. Imag.*, 28(11):1688–1702, November 2009.

[12] T. Niu, X. Dong, M. Petrongolo, and L. Zhu. Iterative image-domain decomposition for dual-energy CT. *Med. Phys.*, 41(4):041901, April 2014.

13 M. M. Goodsitt, E. G. Christodoulou, and S. C. Larson. Accuracies of the synthesized monochromatic CT numbers and effective atomic numbers obtained with a rapid kVp switching dual energy CT scanner. *Med. Phys.*, 38(4):2222–2232, April 2011.

14 M. Daniele, T.B. Daniel, M. Achille, and C. N. Rendon. State of the art: Dual-Energy CT of the abdomen. *Radiology*, 271(2):327–342, May 2014.

15 Y. Xue, R. Ruan, X. Hu, et al. Statistical image-domain multi-material decomposition for dual-energy CT. *Med. Phys.*, 44(3):886–901, 2017.

16 I. Y. Chun and J. A. Fessler. Convolutional dictionary learning: Acceleration and convergence. *IEEE Trans. Im. Proc.*, 27(4):1697–1712, April 2018.

17 W. Wu, H. Yu, P. Chen, et al. DLIMD: Dictionary learning based image-domain material decomposition for spectral CT. May 2019. Online: https://arxiv.org/abs/1905.02567.

18 I. Y. Chun and J. A. Fessler. Convolutional analysis operator learning: Acceleration and convergence. *IEEE Trans. Im. Proc.*, 29:2108–2122, 2020.

19 Z. Li, S. Ravishankar, Y. Long, and J. A. Fessler. Image-domain material decomposition using data-driven sparsity models for dual-energy CT. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 52–56, April 2018.

20 Z. Li, S. Ravishankar, and Y. Long. Image-domain multi-material decomposition using a union of cross-material models. In *Proc. Intl. Mtg. on Fully 3D Image Recon. in Rad. and Nuc. Med*, pages 1107210–1–1107210–5, 2019.

21 Z. Li, S. Ravishankar, Y. Long, and J. A. Fessler. DECT-MULTRA: Dual-energy CT image decomposition with learned mixed material models and efficient clustering. *IEEE Trans. Med. Imag.*, 39(4):1223–1234, 2020.

22 D. Wu, K. Kim, G. Fakhri, and Q. Li. A cascaded convolutional neural network for X-ray low-dose CT image denoising. August 2017. Online: http://arxiv.org/abs/1705.04267.

23 E. Froustey K. H. Jin, M. T. McCann and M. Unser. Deep convolutional neural network for inverse problems in imaging. *IEEE Trans. Im. Proc.*, 26(9):4509–4522, 2017.

24 Y. Liao, Y. Wang, S. Li, et al. Pseudo dual energy CT imaging using deep learning-based framework: basic material estimation. In *Proc. SPIE*, volume 10573, page 105734N, March 2018.

25 Y. Xu, B. Yan, J. Zhang, J. Chen, L. Zeng, and L. Wang. Image decomposition algorithm for dual-energy computed tomography via fully convolutional network. *Comput. Math. Methods Med.*, September 2018.

26 W. Zhang, H. Zhang, L. Wang, et al. Image domain dual material decomposition for dual-energy CT using butterfly network. *Med. Phys.*, 46(5):2037–2051, May 2019.

27 D. P. Clark, M. Holbrook, and C. T. Badea. Multi-energy CT decomposition using convolutional neural networks. In *Medical Imaging 2018: Physics of Medical Imaging*, volume 10573, page 105731O, October 2018.

28 I. Y. Chun and J. A. Fessler. Deep BCD-Net using identical encoding-decoding CNN structures for iterative image recovery. In *Proc. IEEE Wkshp. on Image, Video, Multidim. Signal Proc.*, pages 1–5, 2018.

29 I. Y. Chun, H. Lim, Z. Huang, and J. A. Fessler. Fast and convergent iterative signal recovery using trained convolutional neural networkss. In *Proc. Allerton Conf. on Commun., Control, and Comput.*, pages 155–159, Allerton, IL, October 2018.

30 I. Y. Chun, X. Zheng, Y. Long, and J. A. Fessler. BCD-Net for low-dose CT reconstruction: Acceleration, convergence, and generalization. *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 31–40, October 2019.

31 H. Lim, I. Y. Chun, Y. K. Dewaraja, and J. A. Fessler. Improved low-count quantitative PET reconstruction with an iterative neural network. *IEEE Trans. Med. Imag.*, May 2020. DOI: 10.1109/TMI.2020.2998480.

32 I. Y. Chun, Z. Huang, H. Lim, and J. A. Fessler. Momentum-Net: Fast and convergent iterative neural network for inverse problems. early access in *IEEE Trans. Pattern Anal. Mach. Intell.*, Jul. 2020. DOI: 10.1109/TPAMI.2020.3012955.

33 S. Ye, Y. Long, and I. Y. Chun. Momentum-Net for low-dose CT image reconstruction. accepted to *Asilomar Conf. on Signals, Syst., and Comput.*, August 2020. Online: http://arxiv.org/abs/2002.12018.

[34] Y. Yang, J. Sun, H. Li, and Z. Xu. Deep ADMM-Net for compressive sensing MRI. In *Advances in Neural Information Processing Systems 29*, pages 10–18, December 2016.

[35] Z. Li, I. Y. Chun, and Y. Long. Image-domain material decomposition using an iterative neural network for dual-energy CT. In *Proc. IEEE Intl. Symp. Biomed. Imag.*, pages 651–655, April 2020.

[36] W. Fang, D. Wu, K. Kim, M. K. Kalra, R. Singh, L. Li, and Q. Li. Iterative material decomposition for spectral CT using self-supervised Noise2Noise prior. *Phys. Med. Biol.*, 66(15):155013, June 2021.

[37] C. Maass, M. Baer, and M. Kachelriess. Image-based dual energy CT using optimized precorrection functions: A practical new approach of material decomposition in image domain. *Med. Phys.*, 36(8):3818–3829, 2009.

[38] Y. Xue, Y. Jiang, C. Yang, Q. Lyu, J. Wang, C. Luo, L. Zhang, C. Desrosiers, K. Feng, X. Sun, X. Hu, K. Sheng, and T. Niu. Accurate multi-material decomposition in dual-energy CT: A phantom study. *IEEE Transactions on Computational Imaging*, 5(4):515–529, 2019.

[39] W. Wu, P. Chen, S. Wang, V. Vardhanabhuti, F. Liu, and H. Yu. Image-domain material decomposition for spectral CT using a generalized dictionary learning. *IEEE Transactions on Radiation and Plasma Medical Sciences*, 5(4):537–547, 2021.

[40] W. Wu, H. Yu, P. Chen, F. Luo, F. Liu, Q. Wang, Y. Zhu, Y. Zhang, J. Feng, and H. Yu. Dictionary learning based image-domain material decomposition for spectral CT. *Phys. Med. Biol.*, 65(24):245006, 2020.

[41] S. F. D. Waldron. *An introduction to finite tight frames*. Springer, 2018.

[42] J.-F. Cai, H. Ji, Z. Shen, and G.-B. Ye. Data-driven tight frame construction and image denoising. *Appl. Comput. Harmon. Anal*, 37(1):89–105, 2014.

[43] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. W. Tsui. Realistic CT simulation using the 4D XCAT phantom. *Med. Phys.*, 35(8):3800–3808, August 2008.

[44] D. P. Kingma and J. L. Ba. Adam: A method for stochastic optimization. In *Proc. ICLR*, pages 1–15, May 2015.

45 M. Petrongolo and L. Zhu. Noise suppression for dual-energy CT through entropy minimization. *IEEE Trans. Med. Imag.*, 34(11):2286–2297, 2015.

46 W. Wu, D. Hu, W. Cong, et al. Stabilizing deep tomographic reconstruction, 2021. Online: http://arxiv.org/abs/2008.01846.

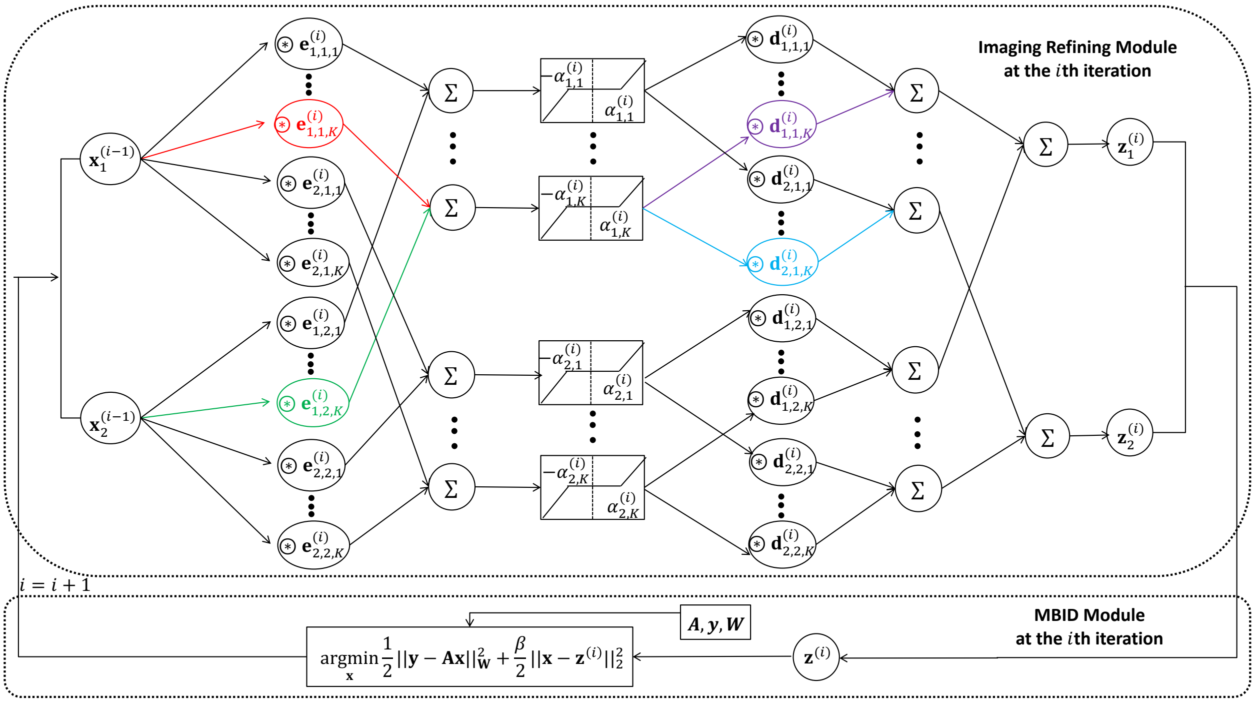


Figure 1: The proposed BCD-Net architecture at the ith iteration, for $i = 1, \ldots, I_{\text{iter}}$.

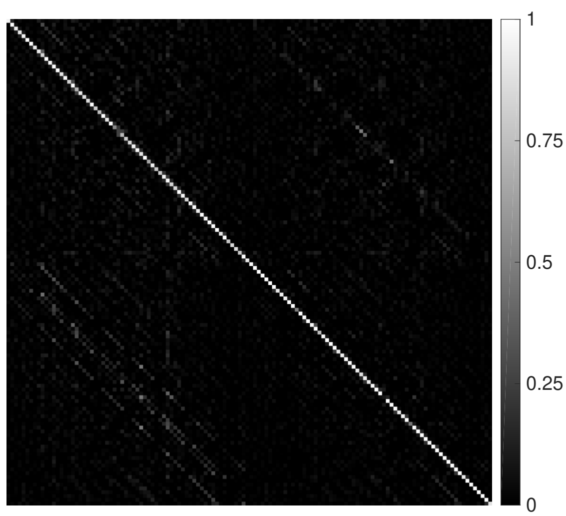


Figure 2: $\mathbf{D}^{(100)}\mathbf{E}^{(100)}$ of BCD-Net-sCNN-hc.

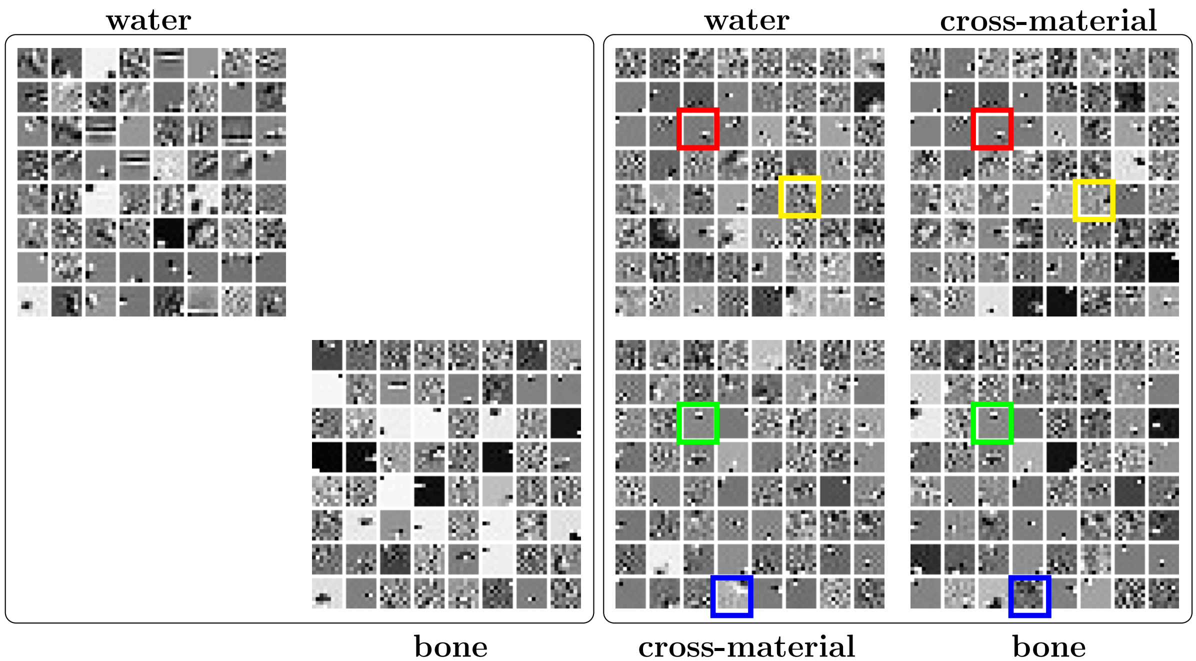


Figure 3: Left and right are learned filters of BCD-Net-sCNN-lc and BCD-Net-sCNN-hc at the last iteration that uses identical encoding-decoding architecture (i.e., $\mathbf{D} = \mathbf{E}^\top$), respectively. Top-left, top-right, bottom-left, and bottom-right correspond to $\mathbf{E}_{1,1}$, $\mathbf{E}_{1,2}$, $\mathbf{E}_{2,1}$, and $\mathbf{E}_{2,2}$, respectively. Four pairs of filters (indicated by four different colors) are selected as examples to show similar or different structures between off-diagonal and diagonal blcok matrices; filters indicated by red or green boxes show similar structures, while blue or yellow boxes show different structures.

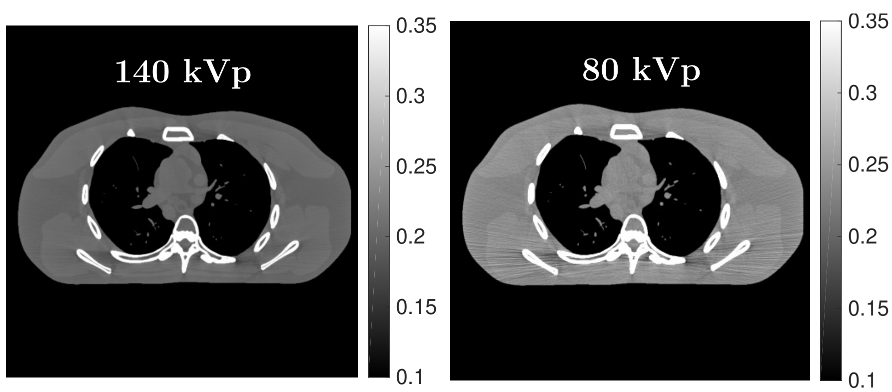
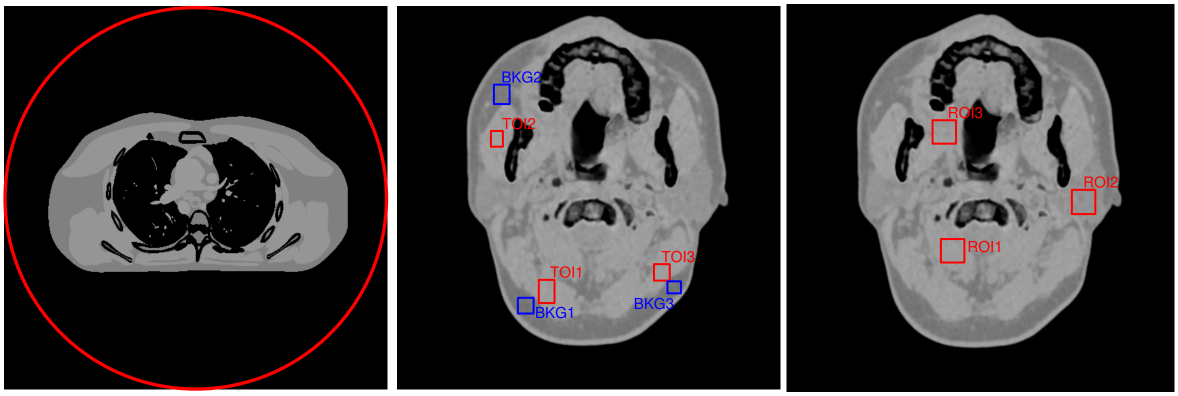


Figure 4: The attenuation images (zoomed-in) for a test slice at high and low energies, respectively.



(a) (b) (c)

Figure 5: (a) ROI used for RMSE calculation for XCAT phantom data. (b) Three selected TOIs in muscle (indicated by red rectangles) and corresponding local background regions in fat (indicated by blue rectangles) on the decomposed water image of head data. (c) Three selected ROIs for NPS calculation for the decomposed water image of head data.

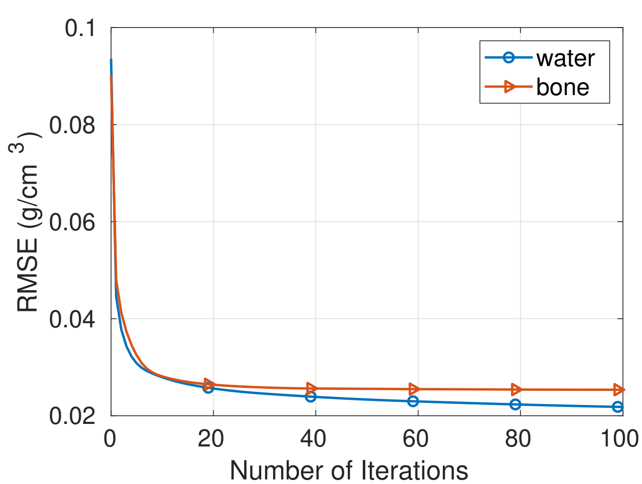


Figure 6: RMSE convergence behaviors of BCD-Net-sCNN-hc (averaged RMSE values across three test slices of XCAT phantom).

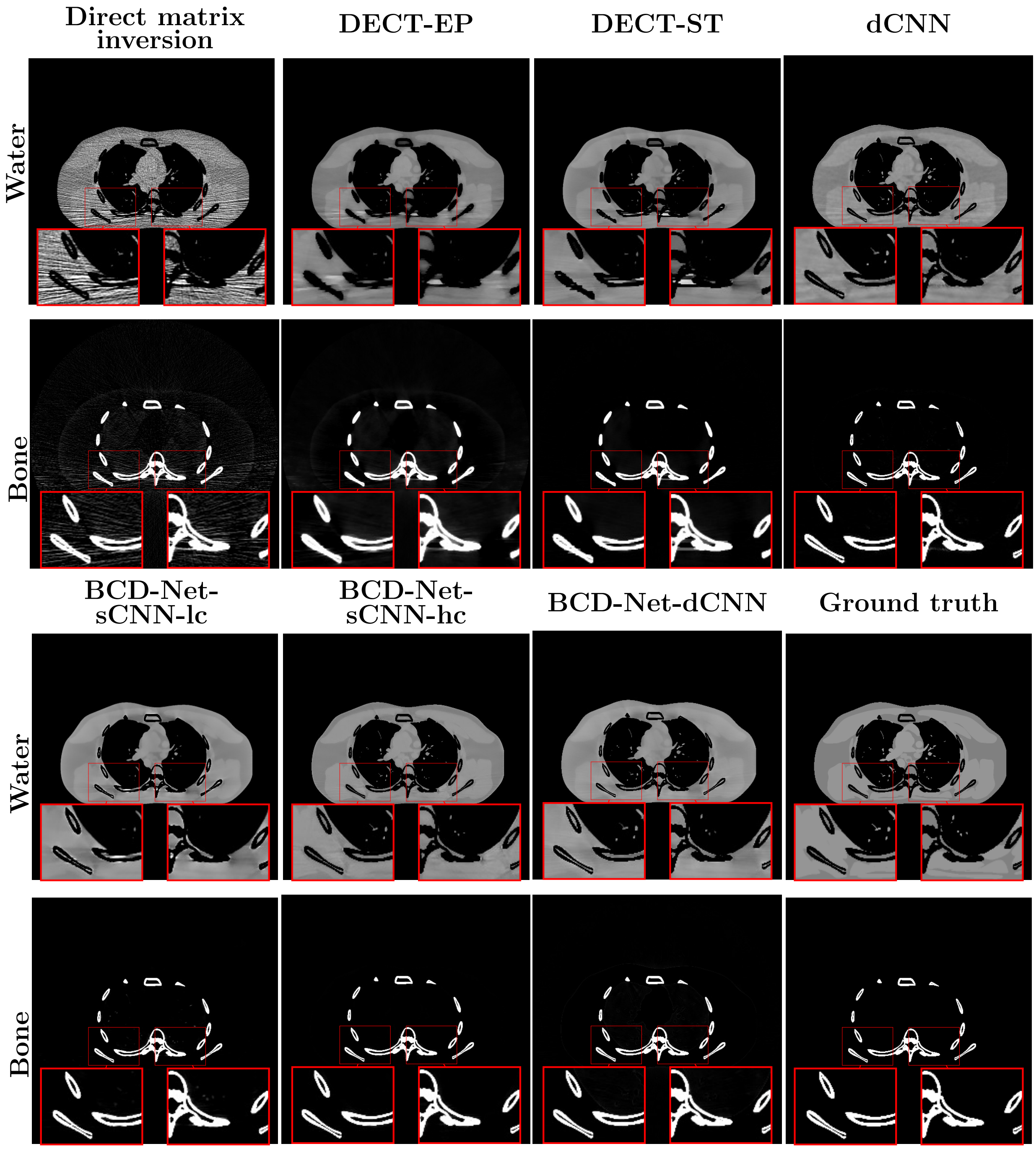


Figure 7: Comparison of decomposed images from different methods (XCAT phantom test slice #1). Water and bone images are shown with display windows [0.7 1.3] g/cm^3 and [0 0.8] g/cm^3, respectively.

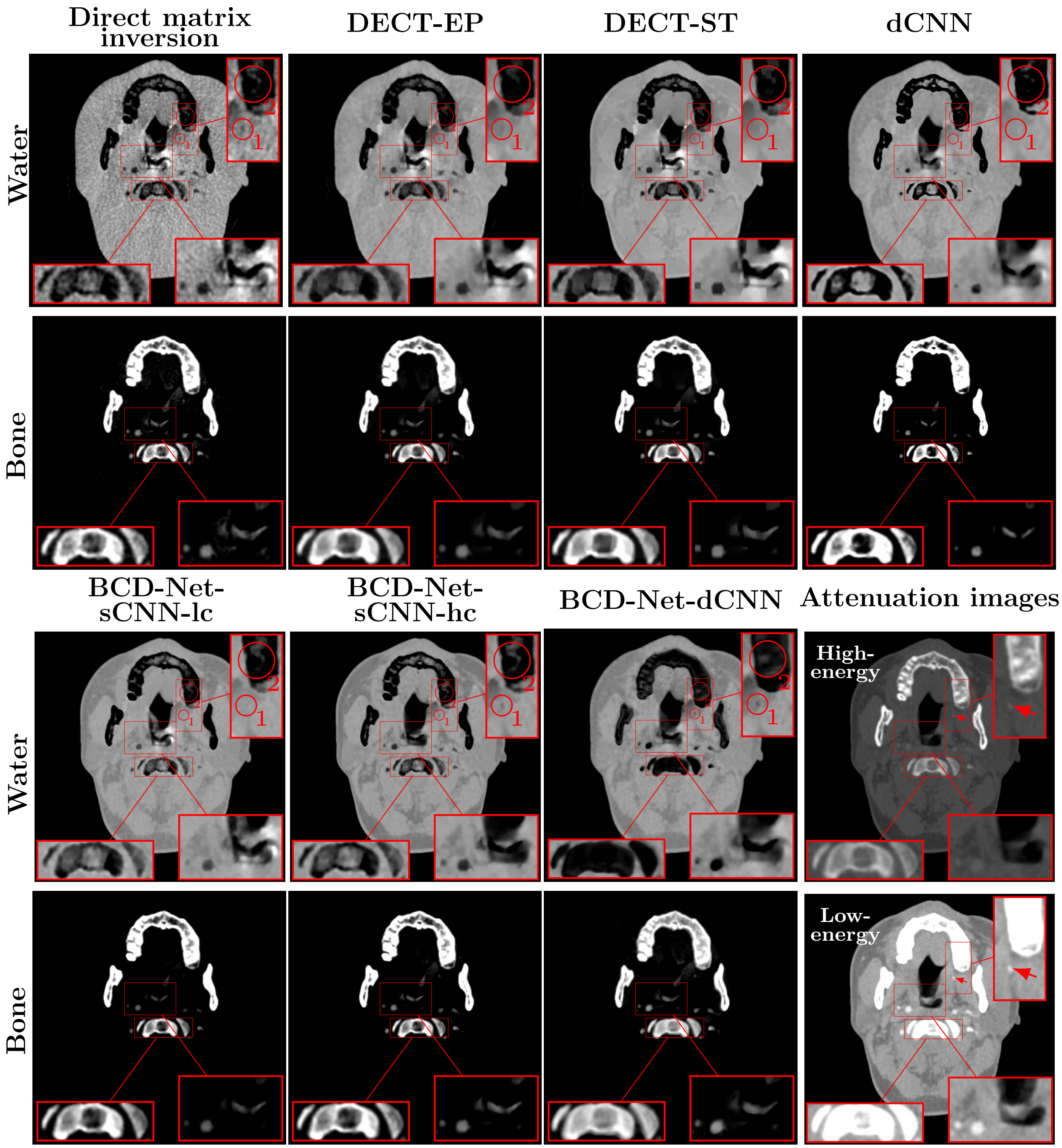


Figure 8: Comparison of decomposed images from different methods (clinical head data). Water and bone images are displayed with windows [0.5 1.3] g/cm^3 and [0.05 0.905] g/cm^3, respectively. High- and low-energy attenuation images are displayed with window [0.1 0.35] cm^{-1}.

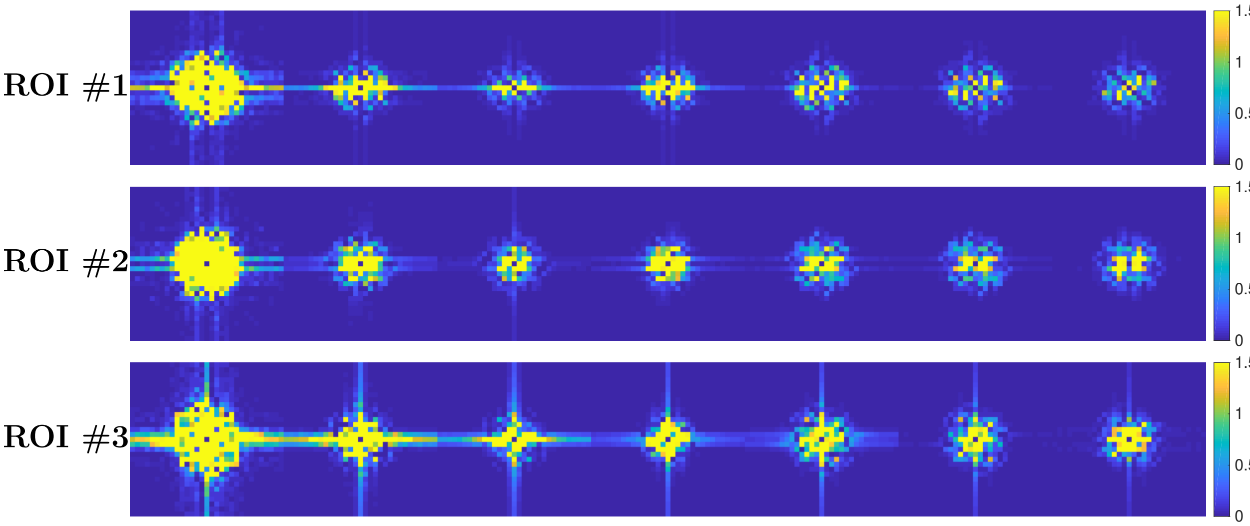


Figure 9: Left to right: NPS measured within ROIs of decomposed water images obtained by direct matrix inversion, DECT-EP, DECT-ST, dCNN, BCD-Net-dCNN, BCD-Net-sCNN-lc, and BCD-Net-sCNN-hc. The first to the third rows show the NPS of the first to third ROI in Figure 5(c), respectively, with display windows [0 1.5] g^2/cm^6.

Table 1: Data acquisition parameters applied in head data acquisition.

Scanner	Head Data	
	High-energy	Low-energy
Peak Voltage (kVp)	140	80
X-ray Tube Current (mA)	364	648
Exposure Time (s)	0.285	
Current-exposure Time Product (mAs)	103.7	184.7
Noise STD (mm^{-1})	1.57×10^{-4}	3.61×10^{-4}
Helical Pitch	0.7	
Gantry Rotation Speed (circle/second)	0.28	

Table 2: RMSE of decomposed material density images obtained by different methods for three different test slices of XCAT phantom. The unit for RMSE is $10^{-3}\,\mathrm{g/cm^3}$.

Methods	Test #1		Test #2		Test #3		Average	
	water	bone	water	bone	water	bone	water	bone
Direct matrix inversion	91.2	89.0	70.4	69.9	119.2	111.9	93.6	90.3
DECT-EP	60.0	68.5	59.5	63.3	69.9	75.9	63.1	69.2
DECT-ST	54.2	60.3	52.1	54.1	62.5	66.3	56.3	60.2
dCNN	21.9	24.3	19.8	20.8	24.9	30.2	22.2	25.1
BCD-Net-sCNN-lc	44.4	39.1	37.0	33.4	47.2	48.8	42.9	40.4
BCD-Net-sCNN-hc	23.0	25.3	20.2	23.2	22.2	27.6	**21.8**	25.3
BCD-Net-dCNN	22.7	23.4	22.0	22.6	20.7	22.0	**21.8**	**22.7**

Table 3: CNR of decomposed water density images obtained by different methods for clinical head data.

	TOI-local BKG #1	TOI-local BKG #2	TOI-local BKG #3	Average
Direct matrix inversion	-0.05	-0.21	0.05	-0.06
DECT-EP	0.14	-0.28	0.63	0.16
DECT-ST	1.97	0.18	3.44	1.86
dCNN	5.08	4.92	4.46	4.82
BCD-Net-sCNN-lc	6.83	8.45	5.39	6.89
BCD-Net-sCNN-hc	10.01	11.48	7.49	**9.66**
BCD-Net-dCNN	8.16	9.44	6.29	7.96

Table 4: RMSE of decomposed density images from training and test samples via dCNN, BCD-Net-dCNN, and BCD-Net-sCNN-hc. RMSE gap is the difference between test RMSE and training RMSE. The unit for RMSE is $10^{-3}\,\mathrm{g/cm^3}$.

	Methods	dCNN		BCD-Net-dCNN		BCD-Net-sCNN-hc	
		water	bone	water	bone	water	bone
	Training	18.4	21.6	18.7	19.4	21.5	22.8
RMSE	Test	22.2	25.1	21.8	22.7	21.8	25.4
	Gap	3.8	3.5	3.1	3.3	0.3	2.6

List of Figures:

- Figure 9: Left to right: NPS measured within ROIs of decomposed water images obtained by direct matrix inversion, DECT-EP, DECT-ST, dCNN, BCD-Net-dCNN, BCD-Net-sCNN-lc, and BCD-Net-sCNN-hc. The first to the third rows show the NPS of the first to third ROI in Figure 5(c), respectively, with display windows [0 1.5] g^2/cm^6.

- Figure S.1: RMSE plot of BCD-Net-dCNN for Test #1, Test #2, and Test #3, respectively.

- Figure S.2: (a) Five selected ROIs indicated for $\overline{\text{NPS}}$ calculation for the decomposed water image of XCAT phantom. (b) Left to right: NPS measured within ROIs of decomposed water images obtained by direct matrix inversion, DECT-EP, DECT-ST, dCNN, BCD-Net-dCNN, BCD-Net-sCNN-lc, and BCD-Net-sCNN-hc. The first to the fifth rows in (b) show the $\overline{\text{NPS}}$ of the first to fifth ROIs, respectively, with display windows [0 0.6] g^2/cm^6.

- Figure S.3: Comparison of decomposed images from different methods (XCAT phantom test slice #2). Water and bone images are shown with display windows [0.7 1.3] g/cm^3 and [0 0.8] g/cm^3, respectively.

- Figure S.4: Comparison of decomposed images from different methods (XCAT phantom test slice #3). Water and bone images are displayed with windows [0.7 1.3] g/cm^3 and [0 0.8] g/cm^3, respectively.

- Figure S.5: Comparison of three decomposed images from regularized direct matrix inversion ($\lambda = 1 \times 10^{-5}$), BCD-Net-sCNN-hc, and ground truth. Fat, muscle, and bone images are shown with display windows [0 2] g/cm^3, [0 2] g/cm^3, and [0 0.5] g/cm^3, respectively.

- Figure S.6: RMSE convergence behaviors of three-material decomposition BCD-Net-sCNN-hc.

- Figure S.7: Comparisons of decomposed bone images (display window [0 0.5] g/cm^3) and their error maps (display window [0 0.3] g/cm^3) from dual- and three-material decomposition BCD-Net-sCNN-hc architectures.