SeNAS-Net: Self-Supervised Noise and Artifact Suppression Network for Material Decomposition in Spectral CT

Xu Ji, Yuchen Lu, Yikun Zhang, Xu Zhuo, Shengqi Kan, Weilong Mao, Gouenou Coatrieux, Jean-Louis Coatrieux, Guotao Quan, Yan Xi, et al.

SeNAS-Net: Self-supervised Noise and Artifact Suppression Network for Material Decomposition in Spectral CT

Xu Ji, Yuchen Lu, Yikun Zhang, Xu Zhuo, Shengqi Kan, Weilong Mao, Gouenou Coatrieux, Jean-Louis Coatrieux, Guotao Quan, Yan Xi, Shuo Li, Tianling Lyu and Yang Chen, *Senior Member, IEEE*

***Abstract*—For material decomposition in spectral computed tomography, the x-ray attenuation coefficient of an unknown material can be decomposed as a combination of a group of basis materials, in order to analyze its material properties. Material decomposition generally leads to amplification of image noise and artifacts. Meanwhile, it is often difficult to acquire the ground truth values of the material basis images, preventing the application of supervised learning-based noise reduction methods. To resolve such problem, we proposed a self-supervised noise and artifact suppression network for spectral computed tomography. The proposed method consists of a projection-domain self-supervised denoising network along with physics-driven constraints to mitigate the secondary artifacts, including a noise modulation item to incorporate the anisotropic noise amplitudes in the projection domain, a sinogram mask image to suppress streaky artifacts and a data fidelity loss item to further mitigate noise and to improve signal accuracy. The performance of the proposed method was evaluated based on both numerical experiment tests and laboratory experiment tests. Results demonstrated that the proposed method has promising performance in noise and artifact suppression for material decomposition in spectral computed tomography. Comprehensive ablation studies were performed to demonstrate the function of each physical constraint.**

***Index Terms*—Spectral CT, Material decomposition, Neural network, Self-supervised Learning, Photon Counting CT**

This work was supported in part by the National Key Project of Research and Development Plan under Grant 2022YFC2408500, in part by the National Natural Science Foundation of China under Grants T2225025 and 62301145, in part by the Key Research and Development Programs in Jiangsu Province of China under Grants BE2021703 and BE2022768, and in part by Jiangsu Province Science Foundation for Youths (BK20220825). Corresponding author: Yang Chen; Co-corresponding author: Tianling Lyu.

Xu Ji and Yang Chen are with the Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, Southeast University, Nanjing, China, and also with the Laboratory of Image Science and Technology, the School of Computer Science and Engineering, Southeast University, Nanjing, China, and also with the Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education, Nanjing, China. (email: xuji@seu.edu.cn; chenyang.list@seu.edu.cn)

Yuchen Lu, Yikun Zhang, Xu Zhuo, Shengqi Kan and Weilong Mao are with the Laboratory of Image Science and Technology, the School of Computer Science and Engineering, Southeast University, Nanjing, China. (email:yclu@seu.edu.cn; yikun@seu.edu.cn; zhuoxu@seu.edu.cn; 2443204956@qq.com; 220212122@seu.edu.cn)

Gouenou Coatrieux is with IMT Atlantique, Inserm, Brest, France. (email: gouenou.coatrieux@telecombretagne.eu)

Jean-Louis Coatrieux is with Laboratoire Traitement du Signal et de l'Image, Universite de Rennes 1, Rennes, France. (email: jean-louis.coatrieux@univ-rennes1.fr)

Guotao Quan is with the CT RPA Department, United Imaging Healthcare Co., Ltd., Shanghai, China. (email: guotao.quan@united-imaging.com)

Yan Xi is with Jiangsu First-Imaging Medical Equipment Co., Ltd., Jiangsu,, China. (email: yanxi@first-imaging.com)

Shou Li is with the Department of Computer and Data Science and Department of Biomedical Engineering, Case Western Reserve University, US. (email: shuo.li11@case.edu)

Tianling Lyu is with the Research Center of Augmented Intelligence, Zhejiang Lab, Hangzhou, China. (email: lyutianling@zhejianglab.com)

I. Introduction

FOR single-energy computed tomography (CT) scan (CT scans with a single x-ray input spectrum), different materials may demonstrate the same measured CT number [i.e., measured x-ray linear attenuation coefficients (LACs)] so that the materials can not be differentiated and quantified. However, since different materials demonstrate different dependencies of LACs on x-ray energy, they can be differentiated based on CT scan measurements under multiple spectra, which is referred to as the spectral CT imaging technique [1]–[3]. The reconstructed CT images corresponding to different spectra can be utilized to acquire the material properties.

One method to extract the material property information is through material decomposition: with two or more materials as a group of basis materials (e.g., water and iodine), the acquired sinograms (i.e., data before the CT reconstruction) or CT images are decomposed to generate the decomposition coefficients [4]–[7]. Material decomposition methods can be divided into two major categories: projection domain decomposition [1], [4], [8], [9] and image domain decomposition [10]–[15]. The projection domain material decomposition acquires line integrals of different materials by performing a non-linear transformation of the sinograms [1]. In principle, projection domain material decomposition can eliminate beam hardening artifacts since the polychromatic projection process is incorporated in the non-linear transformation. However, such decomposition method requires paired multi-spectral sinograms which is sometimes unavailable for clinical spectral CT systems. In this case, the image domain decomposition is a more versatile method, which generates the material basis images from a linear transformation of the reconstructed spectral CT images. Such linear transform is based on the assumption that the reconstructed spectral CT images can be approximated as mono-energetic CT images under different energy levels, which are free of artifacts induced by polychromaticity. If the assumption is violated, these artifacts would propagate to the decomposed basis images, thus influencing the quantification accuracy (e.g., accuracy of measurement of iodine concentrations in human body) [10].

Besides the consideration on the decomposition accuracy, both the projection domain and image domain material decomposition suffer from the noise amplification problem: the decomposed basis images demonstrate higher noise levels than the original CT images. Due to such decomposition noise, application of spectral CT has been restricted to quantification for regions of interest comprising many pixels with similar pixel values while quantitative information related to the tiny structures (e.g., contrast-enhanced vessels) are hard to acquire [16]. Previous studies proposed different methods to mitigate the noise of the decomposition process. Conventional methods utilized the negative statistical correlation between the basis images to reduce noise. Kalender et al. estimated the noise of one basis image, which was scaled and added to another basis image to reduce noise [16]. Petrongolo et al. reduced decomposed basis image noise by using a basis transformation in image domain based on entropy minimization [17]. Jiang et al. expanded such method to multi-material decomposition [18]. Besides conventional methods, iterative methods reduce the noise by introducing different regularizers. Some of the iterative methods directly outputs the basis CT images from raw projection domain data, which are also referred to as one-step inversion methods. Long et al. proposed a multi-material decomposition method based on the x-ray transmission physics accompanied by an edge-preserving regularizer [19]. Iterative calculations can also be performed in the image domain. Niu et al. proposed an iterative image-domain material decomposition method to reduce the noise based on the statistical correlation between the basis images [12]. Xue et al. developed a statistical image-domain decomposition method with a negative logarithm-likelihood term and an edge-preserving regularization [15]. The major drawback of iterative methods is that the computational time is relatively long due to the iterative calculations.

With the development of deep learning algorithms, multiple studies demonstrate the advantage of incorporating deep neural networks in the material decomposition process to reduce noise and artifacts [20]–[25]. Most of the deep learning-based methods utilize supervised learning, which requires the ground truths of the material basis images. These ground truths, however, are often unavailable for experimental or clinical systems. To resolve such limitation, several studies introduced a self-supervised denoising method for noise reduction studies. Lehtinen et al. found that the optimal network parameters remain the same when zero-mean noise is added to the clean network labels. Based on such finding, Noise2Noise method was proposed where neural network was trained by mapping one noisy image to another noisy image without the need of clean training labels [26]. Fang et al. utilized the denoised image from the Noise2Noise method as a prior to reduce the noise of the decomposed material basis images [27]. When noisy image pairs taken from the same scene are not available, an alternative self-supervised training strategy named Neighbor2Neighbor (Ne2Ne) utilizes random downsampling operators to create noisy data pairs [28]. Wu et al. applied such method along with guided filtering to low dose CT imaging in the reconstructed CT image domain [29]. However, the reconstruction algorithm can lead to complex correlations between pixels in the CT images [30]; since the original Ne2Ne method requires the neighboring pixels to be free of correlations [28], the performance of the Ne2Ne method may be compromised.

This work incorporates the idea of the Ne2Ne method to realize self-supervised noise and artifact suppression for material decomposition in spectral CT. The noise reduction is performed in the projection domain in order to disentangle the complex noise correlations from the CT reconstruction. However, direct application of the Ne2Ne method in the projection domain without additional constraints can lead to degradation of image quality (which will be shown later in the ablation study). This work proposes several physics-driven constraints together to mitigate secondary artifacts caused by the projection domain operations:

- A sinogram mask (generated from forward projection of the thresholded CT image) is added as network input to suppress streak artifacts;
- A noise modulation item is added to incorporate the anisotropic noise levels within each sinogram;
- A data fidelity loss item incorporating the polychromatic x-ray interaction model is utilized to further suppress noise and to maintain signal accuracy.

To the best of our knowledge, this work is the first self-supervised deep-learning method to realize noise and artifact suppression for projection-domain material decomposition in spectral CT. The proposed self-supervised noise and artifact reduction network (SeNAS-Net) is validated by both numerical experiments and physical experiments and is compared with a conventional denoising method [17] and several supervised denoising methods. The proposed method demonstrates promising performance in noise and artifact suppression for spectral CT without the need of training labels. In addition, comprehensive ablation studies are performed to demonstrate the function of each physics-driven constraint.

II. Methods

A. Theoretical workflow of the proposed self-supervised material decomposition noise and artifact suppression method

The proposed self-supervised material decomposition noise suppression method utlizes the Ne2Ne denoising strategy for noise suppression in the projection domain. To mitigate the secondary artifacts from operations in the projection domain, a mask image as well a projection domain constraint item is incorporated in the network. The flowchart and the network structure of the proposed method are summarized in Fig. 1 and Fig. 2. Details of the proposed methods are given as follows.

1) Review of the Ne2Ne denoising method: We will first briefly review the Ne2Ne denoising method. For a denoising network f with parameters $\boldsymbol{\theta}$ and noisy input $\vec{x}$, the output of network is given as $f(\vec{x};\boldsymbol{\theta})$. If the clean images $\vec{y}$ corresponding to each $\vec{x}$ are available as training labels, the aim is to solve following optimization problem:

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \mathbb{E}_{\vec{x}} ||f(\vec{x};\boldsymbol{\theta}) - \vec{y}||_2^2, \tag{1}$$

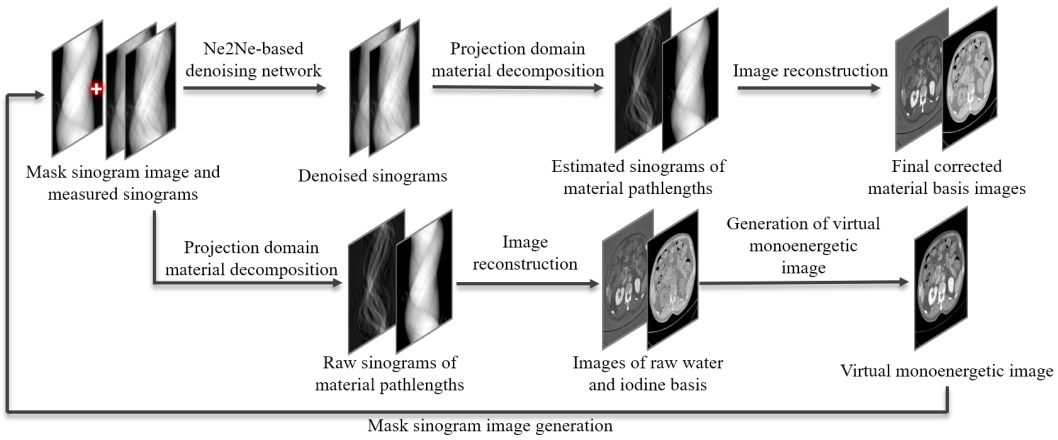

Fig. 1. Overall flowchart of the proposed method. The acquired multi-energy sinograms goes through the Ne2Ne-based denoising network. The denoised sinograms are decomposed and reconstructed to generate the material basis CT images. The network structure is shown in Fig. 2. The mask sinogram image is generated from the VMI based on the conventional material decomposition process, where the detailed workflow is shown in Fig. 5. The loss items for the projection domain are illustrated in Fig. 4.

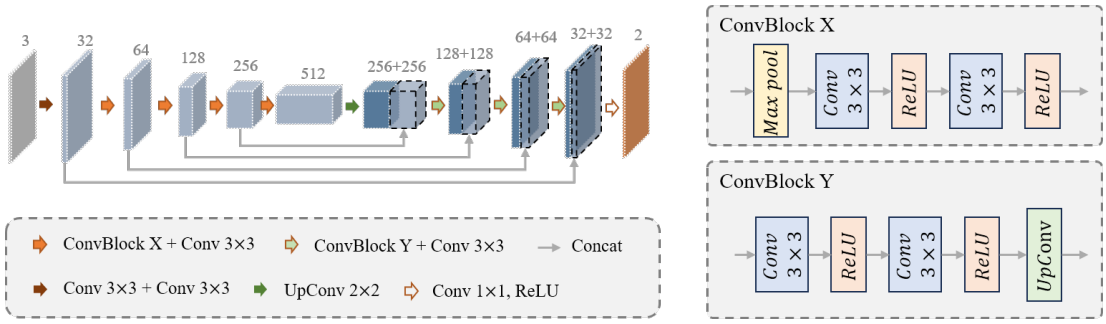

Fig. 2. Structure for the projection domain denoising network to output the unmodulated sinogram noise signal ($f(\boldsymbol{S}_{\text{in}}, \boldsymbol{S}_{\text{M}}; \boldsymbol{\theta})$).

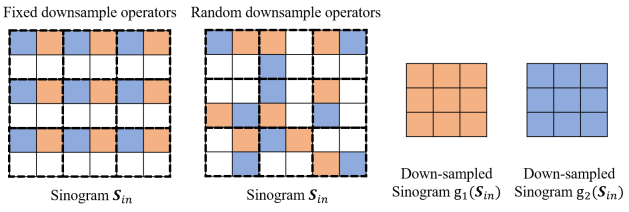

Fig. 3. Illustration of the random downsample operators.

where $\mathbb{E}$ represents the expectation operator and $||\cdot||_2^2$ represents the L2 loss. Due to the fact that clean images $\vec{y}$ are often unavailable, self-supervised denoising methods come into play. The Ne2Ne denoising method relies on the assumption that the neighboring pixels are weakly correlated [28]. It utilized random downsample operators to form the training label to calculate the reconstruction loss $\mathcal{L}_{\text{rec}}$:

$$\mathcal{L}_{\text{rec}} = ||f(g_1(\vec{x})) - g_2(\vec{x})||_2^2, \tag{2}$$

where g_1 and g_2 are random downsample operators, which is illustrated in Fig. 3. Due to the fact that the clean realizations of $g_1(\vec{x})$ and $g_2(\vec{x})$ differ, minimization of $\mathcal{L}_{\text{rec}}$ leads to over-smoothing; therefore, the regularization loss is added:

$$\mathcal{L}_{\text{reg}} = ||f(g_1(\vec{x})) - g_2(\vec{x}) - \{g_1(f(\vec{x}) - g_2(f(\vec{x})\}||_2^2, \tag{3}$$

The parameters of the network is acquired through minimization of the summation of $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{reg}}$:

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \mathbb{E}_{\vec{x}}\{\mathcal{L}_{\text{rec}} + a\mathcal{L}_{\text{reg}}\}, \tag{4}$$

where a is a coefficients changing the weight of $\mathcal{L}_{\text{rec}}$ and $\mathcal{L}_{\text{reg}}$.

2) Projection domain denoising network: A denoising network is applied in the projection domain. The whole network is denoted as F. The input of the network is the measured noisy multi-energy sinograms (e.g., $\boldsymbol{S}_{\text{in}} = \{S_{\text{in},1}, S_{\text{in},2}\}$ for dual-energy CT) and a mask sinogram image $\boldsymbol{S}_{\text{M}}$. The mask sinogram image is to mitigate the secondary artifacts in the image domain caused by the sinogram operations. In this work, the network structure is shown in Fig. 2. The direct output from the network, $f(\boldsymbol{S}_{\text{in}}, \boldsymbol{S}_{\text{M}}; \boldsymbol{\theta})$ (f indicates the network processing and $\boldsymbol{\theta}$ are the network parameters), is the unmodulated output noise signals of the input sinograms. The noise variances of a sinogram can be different for different regions: assuming Poisson statistics, the noise variance of the post-log signal is proportional to $1/N$, where N is the photon detected per pixel when the image object is present. Since N is proportional to $e^{-\boldsymbol{S}}$ ($\boldsymbol{S}$ is the sinogram), the noise standard deviation is proportional to $\sqrt{1/e^{-\boldsymbol{S}}} = e^{\boldsymbol{S}/2}$. Therefore, $f(\boldsymbol{S}_{\text{in}}, \boldsymbol{S}_{\text{M}}; \boldsymbol{\theta})$ is further modulated as:

$$f(\boldsymbol{S}_{\text{in}}, \boldsymbol{S}_{\text{M}}; \boldsymbol{\theta}) \otimes e^{\text{smooth}(\boldsymbol{S}_{\text{in}})/2}, \tag{5}$$

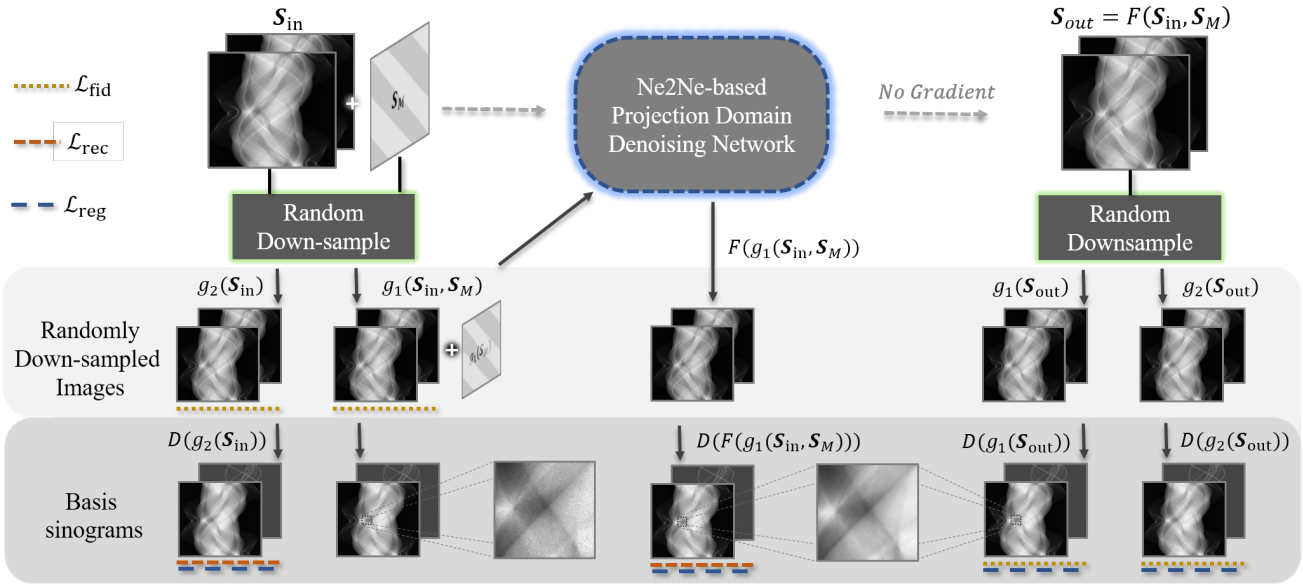

Fig. 4. Details of the data processing workflow and loss items of the projection domain denoising network. Lines with different line styles indicate the items to calculate the specific type of loss (for example, $\mathcal{L}_{\text{rec}}$ is calculated from $D[g_2(\boldsymbol{S}_{\text{in}})]$ and $D[F(g_1(\boldsymbol{S}_{\text{in}}; \boldsymbol{S}_{\text{M}}))]$).

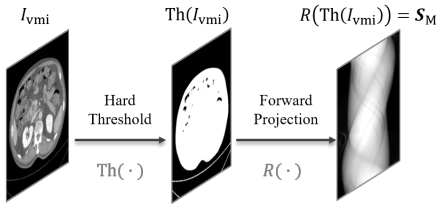

Fig. 5. The mask sinogram image is generate by thresholding the virtual monoenergetic image following a forward projection process.

where $\otimes$ indicates the Hardmard product; "smooth" indicates a low-pass filter to mitigate the noise of S_{in} for better noise amplitude estimation. Here the low-pass filter is a Gaussian filter with a width of 3 pixels.

Operations in the projection domain denoising network leads to secondary artifacts in the reconstructed CT images, especially in regions close to the high contrast boundaries between air the image object (will be demonstrated in the Results section). To mitigate this issue, the mask sinogram image $\boldsymbol{S}_{\text{M}}$ is introduced. The idea is inspired by deep learning-based metal artifact reduction methods, where the trace of the metal is masked in sinograms [31]. The workflow to generate the mask sinogram image is shown in Fig. 5: the raw sinograms are decomposed following a standard projection domain material decomposition process [32] and are reconstructed to generate the noisy material basis images:

$$\boldsymbol{I} = R^{-1}(D(\boldsymbol{S}_{\text{in}})), \tag{6}$$

where R^{-1} is the inverse Radon transform or filtered back-projection operator and D represents the material decomposition operator. The material basis images are scaled and summed to generate the virtual monoenergetic image (VMI). Take two material decomposition as an example, namely $\boldsymbol{I} = \{I_1, I_2\}$. The VMI is calculated as

$$I_{\text{vmi}} = \mu_1(E_0)I_1 + \mu_2(E_0)I_2, \tag{7}$$

where E_0 is selected to be the average of the mean energies of the simulated spectra and μ_i is the energy-dependent linear attenuation coefficients of the i-th material. Then the VMI is thresholded and forward projected to generate the sinogram mask image:

$$\boldsymbol{S}_{\text{M}} = R(\text{Th}(I_{\text{vmi}})), \tag{8}$$

where R indicates the Radon transform for parallel-beam cases or x-ray transform for fan-beam cases and Th indicates a hard thresholding process. The threshold value is selected to be half of the LAC of water. The VMI contains less noise and artifacts compared with the raw reconstructed CT image for each spectrum; therefore, it is utilized to generate the mask image. The final estimated noise $\delta\boldsymbol{S}_{\text{in}}$ in the projection domain is further modulated by $\boldsymbol{S}_{\text{M}}$:

$$\delta\boldsymbol{S}_{\text{in}} = f(\boldsymbol{S}_{\text{in}}, \boldsymbol{S}_{\text{M}}; \boldsymbol{\theta}) \otimes e^{\text{smooth}(\boldsymbol{S}_{\text{in}})/2} \otimes \boldsymbol{S}_{\text{M}}. \tag{9}$$

The noise is subtracted from the input sinograms to generate the noise-free sinograms:

$$\boldsymbol{S}_{\text{out}} = \boldsymbol{S}_{\text{in}} - f(\boldsymbol{S}_{\text{in}}, \boldsymbol{S}_{\text{M}}; \boldsymbol{\theta}) \otimes e^{\text{smooth}(\boldsymbol{S}_{\text{in}})/2} \otimes \boldsymbol{S}_{\text{M}}. \tag{10}$$

The output of the denoising network can be summarized as:

$$\boldsymbol{S}_{\text{out}} = F(\boldsymbol{S}_{\text{in}}, \boldsymbol{S}_{\text{M}}; \boldsymbol{\theta}), \tag{11}$$

where $\boldsymbol{S}_{\text{out}} = \{S_{\text{out},1}, S_{\text{out},2}\}$ is the network output for dual energy CT.

3) Incorporation of the Ne2Ne denoising method: Due to the lack of noise-free multi-energy sinograms, the Ne2Ne method is incorporated to realize self-supervised denoising. The reason in performing the Ne2Ne method in the projection domain is that the noise of the neighboring pixels of the sinograms are weakly correlated (especially for photon counting

CT [33]), compared with the reconstructed CT images. As pointed out in the original Ne2Ne paper, the Ne2Ne method is more effective if the noise amplitude is larger than a threshold value: for a given 8-bit image with value 0-255, noise amplitude should better be larger than 5. The noise amplitude of the measured multi-energy sinograms marginally meet the above criterion. Since the material decomposition process leads to noise amplification, the loss of the Ne2Ne method is calculated from the material basis sinograms from the projection domain material decomposition, rather than the original multi-energy sinograms. Mathematically speaking, $\mathcal{L}_{\rm rec}$ and $\mathcal{L}_{\rm reg}$ are calculated as:

$$\mathcal{L}_{\rm rec} = \mathbb{E}_{\boldsymbol{S}_{\rm in}} ||D[F(g_1(\boldsymbol{S}_{\rm in}, \boldsymbol{S}_{\rm M})) - g_2(\boldsymbol{S}_{\rm in})]||_2^2; \tag{12}$$

$$\mathcal{L}_{\rm reg} = \mathbb{E}_{\boldsymbol{S}_{\rm in}} ||D[F(g_1(\boldsymbol{S}_{\rm in}, \boldsymbol{S}_{\rm M})) - g_2(\boldsymbol{S}_{\rm in}) - \{g_1(F(\boldsymbol{S}_{\rm in}, \boldsymbol{S}_{\rm M})) - g_2(F(\boldsymbol{S}_{\rm in}, \boldsymbol{S}_{\rm M}))\}]||_2^2; \tag{13}$$

where D represents the projection domain material decomposition. For the detailed form of D, we used to up to second-order polynomials following reference [32]. The coefficients can be acquired through a calibration process, which will be discussed in section II-B.

Since second-order polynomials may not perfectly characterize the relationship between the material basis sinograms and the raw multi-energy sinograms, a data fidelity loss item based on the polynomial forward projection process is added:

$$\mathcal{L}_{\rm fid} = \mathbb{E}_{\boldsymbol{S}_{\rm in}} ||h(D[F(\boldsymbol{S}_{\rm in}, \boldsymbol{S}_{\rm M})]; \boldsymbol{\Omega}(E)) - \boldsymbol{S}_{\rm in}||_2^2, \tag{14}$$

where $\boldsymbol{\Omega}(E) = \{\Omega_1, \Omega_2, \ldots, \Omega_m\}$ are the input x-ray spectra and h represents the polynomial forward projection:

$$h(\boldsymbol{L}; \boldsymbol{\Omega}(E)) = \{-\ln \int_E \Omega_i(E) e^{-\sum_{j=1}^n \mu_j(E) L_j} {\rm d}E\}, \quad i = 1, 2, \ldots, m, \tag{15}$$

where $\boldsymbol{L}_{\rm out} = D[F(\boldsymbol{S}_{\rm in}, \boldsymbol{S}_{\rm M})] = \{L_1, L_2, \ldots, L_n\}$ are the decomposed pathlengths for n types of materials, E is the x-ray energy and μ_j are the linear attenuation coefficient of the j-th material. For dual energy measurement, $\boldsymbol{\Omega}(E)$ and $\boldsymbol{L}$ are simplified to be $\boldsymbol{\Omega}(E) = \{\Omega_1, \Omega_2\}$ and $\boldsymbol{L} = \{L_1, L_2\}$. The data processing workflow and loss items for the projection domain denoising network are summarized in Fig. 4.

4) Summary of the proposed method: The total loss $\mathcal{L}_{\rm tot}$ of the denoising network F is given as:

$$\mathcal{L}_{\rm tot} = \alpha \mathcal{L}_{\rm fid} + \beta \mathcal{L}_{\rm reg} + \gamma \mathcal{L}_{\rm rec}. \tag{16}$$

where α, β and γ are the weighting coefficients. In this work, the values of these coefficients are selected to be 1, 1 and 2 respectively. A comprehensive study will be performed later to demonstrate the impact of the coefficients on the results. The aim of the proposed method is to find the value of the network parameters to minimize the total loss:

$$\boldsymbol{\theta} = \arg\min_{\boldsymbol{\theta}} \mathcal{L}_{\rm tot}. \tag{17}$$

B. Experimental Methods

1) Methods for numerical experiments: Numerical experiments were performed to test the performance of the proposed method. High-quality dual energy clinical abdomen CT images (more than 1500 slices in total) from 13 patients were collected in Shanghai Changhai Hospital, China, using a clinical diagnostic CT scanner (Model uCT960, United Imaging Healthcare Co., Ltd., Shanghai, China). The usage of the data was with the approval of the institutional review board and patient consent forms (under approval number 2018-038-01). Water and iodine density images were generated from the dual energy CT images as digital phantoms. The basis images were forward projected to generate the basis sinograms $\boldsymbol{L} = \{L_w, L_b\}$, which were then used to generate the polychromatic x-ray sinograms:

$$S_i = -\ln \int_E \Omega_i(E) e^{-\mu_w(E) L_w - \mu_b(E) L_b} {\rm d}E, \tag{18}$$

where $i = 1, 2$ representing low and high energy measurement. The forward projection was implemented using an in-house CT simulation package named MandoCT [34]. The energy-dependent attenuation coefficients $\mu(E)$ are from the NIST website [35]. The spectra in the simulation study were shown in Fig. 6, with kVp = 80 kV and 140 kV, respectively. The mean energy of the spectra were approximately 50 and 75 keV, respectively. The forward projection process was simulated to be a fan-beam one, with an ideal point x-ray source and a line detector. An ideal photon counting detector with a pixel size of 0.8 mm and a pixel array of 1 by 800 pixels was simulated. The source to detector distance and the source to isocenter distance were simulated to be 800 mm and 1200 mm respectively. The simulated CT scan was a 360 degree full scan with 720 view angles. Poisson noise was added to the pre-log raw forward projection data to simulated the noise in CT data acquisition:

$$N_{\rm m} \sim {\rm Poisson}(N_0 \times e^{-S}), \tag{19}$$

where N_0 is the number of photon counts in the absence of the image object, S is the post-log sinogram and $N_{\rm m}$ is measured post-object photon counts. The measured post-log projection data $p_{\rm m}$ with noise is calculated as:

$$p_{\rm m} = -\ln \frac{N_{\rm m}}{N_0}. \tag{20}$$

In the simulation study, $N_0 = 10^5$ per pixel. The training data and validation data were from 10 patients (7 patients for training data and 3 patients for validation data) while the test data were from the remaining 3 patients. During the training process, data augmentation operations including random cropping, random flipping, and random noise generation were introduced to further expand the dataset. Data corresponding to the same patient were not split to the training dataset and the test dataset. The reconstructed CT images has a dimension of 512×512 and a pixel pitch of 0.8 mm.

Up to second-order polynomials were used to fit the projection domain decomposition relationships:

$$L_w = L_w(S_1, S_2) = a_{1,0}S_1 + a_{0,1}S_2 + a_{2,0}S_1^2 + a_{1,1}S_1S_2 + a_{0,2}S_2^2 \tag{21}$$

$$L_b = L_b(S_1, S_2) = b_{1,0}S_1 + b_{0,1}S_2 + b_{2,0}S_1^2 + b_{1,1}S_1S_2 + b_{0,2}S_2^2 \tag{22}$$

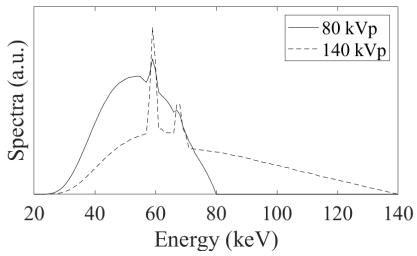

Fig. 6. Spectra for the dual energy CT scan simulation.

To achieve the coefficients for the projection domain material decomposition, materials with water thicknesses ranging from 0 to 30 cm and bone thickness ranging from 0 to 10 cm was simulated. The polychromatic projection values S_i were calculated based on (18). The coefficients were acquired using least-square calculations:

$$\{a_{i,j}\} = \arg\min_{\{a_{i,j}\}} ||L_w - L_w(S_1, S_2)||_2^2 \tag{23}$$

$$\{b_{i,j}\} = \arg\min_{\{b_{i,j}\}} ||L_b - L_b(S_1, S_2)||_2^2 \tag{24}$$

The bone and water basis images were then transformed into iodine and water basis images based on simple linear basis transformation process.

To evaluate the performance of the proposed method on the noise standard deviation, signal accuracy and the spatial resolution on the decomposed basis images, an additional digital quantification phantom was used, as shown in Fig. 7. The simulation of the projection data of the quantification phantom follows those of the simulation of the patient. The reconstructed CT images has a dimension of 512×512 and a pixel pitch of 0.4 mm.

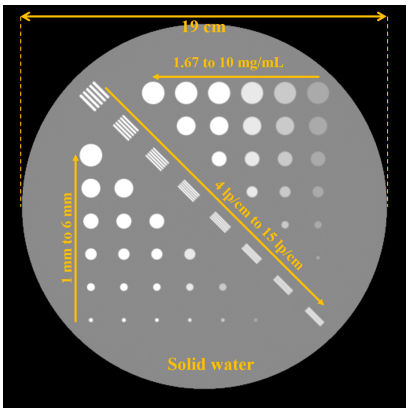

Fig. 7. Digital quantification phantom to evaluate the quantitative performance of the proposed method. The base of the phantom is solid water while the inserts are iodine solutions with different sizes and concentrations.

2) Methods for laboratory experiments: To validate the proposed method using real experimental data, dual energy CT measurements were performed on a benchtop CT system. The x-ray source of the system was from Spellman Cor. (Model XRB160PN192). The detector was an energy integrating detector (Model CareView 300 RF, CareRay Digital Medical Technologies, China) with an active area of 201 mm by 151 mm and a pixel pitch of 98 μm. The pixels were binned 4 by 8 for reconstruction. The source to detector distance was 860 mm and the source to isocenter distance was 1500 mm. The dual energy CT scan was realized by scanning the image object twice: first with 80 kVp tube potential and then with 120 kVp tube potential. Each CT scan was a 360 degree full scan with 1080 view angles. The image object was an in-house quantification phantom. The diameter of the phantom was approximately 7 cm and was filled with water. $CaCl_2$ solutions with different concentrations (10, 25, 50 and 100 mg/mL) were placed in plastic test tubes, which were then inserted into the phantom. The reconstructed CT images has a dimension of 512×512 and a pixel pitch of 0.2 mm. To test the performance of the proposed method for animal objects, additional experiments with mice objects were performed on a prototype small animal CT system. The x-ray source of the system was a microfocus x-ray tube (Model L9421-02. Hamamatsu Photonics, Hama-matsu, Japan). The detector was a photon counting detector (Model XC-Thor, Direction Conversion AB, Sweden) with a active area of 77 mm by 13 mm and a pixel pitch of 100 μm. The source to detector distance was 110 mm and the source to isocenter distance was 220 mm. The CT scan was a 360 degree full scan with 600 view angles. During the CT scan, the tube potential was 40 kVp. The detector thresholds were set to be 25 keV and 33 keV to realize dual energy CT scan with the [25,33] keV as the low energy bin and the [33,40] keV as the high energy bin. The reconstructed CT images has a dimension of 384×384 and a pixel pitch of 0.075 mm. Due to the fact that the image object was small, the projection domain material decomposition was approximated as a linear process without considering the 2nd or higher order items. Since the objects are small, the LACs of the basis materials (water and 100 mg/mL $CaCl_2$ for quantification phantom, soft tissue and bone for mice study) were directly measured from the spectral CT images to form the LAC matrix. The linear decomposition coefficients were obtained by calculating the inverse of the LAC matrix. It should also be noted that in this case, the data fidelity item in (14) degenerates into a linear relationship.

3) Training strategies and computational cost: The network training and testing for both the numerical simulation study and the physical experiment study were implemented using PyTorch toolbox. The environment for the network training and testing was a desktop computer with Intel Core i7-6950X CPU and one NVIDIA TITAN X GPU. The network were trained using the Adam optimizer with a learning rate of 1×10^{-3} and weight decay of 10^{-4}. The networks were trained 200 epochs and the learning rate was reduced to half after each 40 epochs.

4) Quantitative metrics: For the numerical simulation studies, material basis images were evaluated using the root mean squared error (RMSE) and the structural similarity index (SSIM), with the ground truth values of the materials basis images as references. For the quantification phantom, the noise standard deviation and the mean values of the concentrations of the iodine inserts were measured to quantify the denoising performance as well as the signal accuracy. The resolution patterns were used to evaluate the spatial resolution. For the

physical experiments, due to the fact that the ground truths of the material basis images are not available, images are evaluated by visual inspection.

5) Comparison studies and ablation studies: Comparison studies include a conventional material decomposition method based on entropy minimization [17] and three supervised deep learning methods. One supervised deep learning method was implemented by omitting the Ne2Ne modules while directly using the clean material basis images as training labels [Denoted as "Proposed-S" (Proposed-supervised)]. Other aspects of the supervised method were the same with the proposed self-supervised method. The second supervised deep learning method was the Direct-Net, which is a dual-domain neural network for material decomposition [25]. The third method was the RED-CNN method, which is for denoising in the image domain [36]. The input of the network was the grouped reconstructed basis images from projection domain decomposition while the labels were the grouped ground truths of the basis images.

Comprehensive ablation studies were performed to demonstrate the necessity of different components of the proposed method. The ablation studies included: **#1**: All additional constraints were removed and Ne2Ne was directly applied to the decomposed material basis sinograms. **#2**: With all additional constraints except the sinogram mask; **#3**: With all additional constraints except the data fidelity loss item $\mathcal{L}_{\text{fid}}$ in (16);**#4**: With all additional constraints except the sinogram noise modulation item in (5); **#5**: The reconstruction loss and regularization loss for Ne2Ne was calculated from the original sinogram rather than the decomposed sinogram, namely:

$$\mathcal{L}_{\text{rec}} = \mathbb{E}_{\boldsymbol{S}_{\text{in}}} ||F(g_1(\boldsymbol{S}_{\text{in}})) - g_2(\boldsymbol{S}_{\text{in}})||_2^2; \tag{25}$$

$$\mathcal{L}_{\text{reg}} = \mathbb{E}_{\boldsymbol{S}_{\text{in}}} ||F(g_1(\boldsymbol{S}_{\text{in}})) - g_2(\boldsymbol{S}_{\text{in}}) - \{g_1(F(\boldsymbol{S}_{\text{in}})) - g_2(F(\boldsymbol{S}_{\text{in}}))\}||_2^2; \tag{26}$$

This ablation study was to demonstrate the side effect when the noise level was not high enough to meet the criterion of the Ne2Ne method.

We also tested the impact of the values of the coefficients in (16) on the final results. With β remained 1, we altered the value of α from 0.5 to 4 and value of γ from 1 to 16. The results were analyzed quantitatively by comparing those with the results from the original proposed method.

III. Results

A. Results for numerical experiments

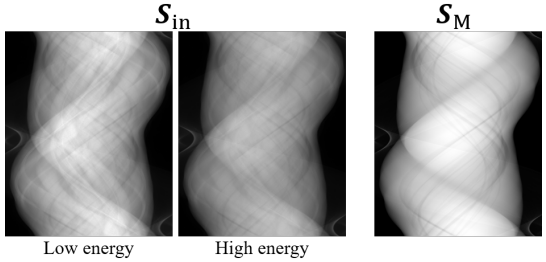

Fig. 8. Sample of the low and high energy sinograms as well as the mask image as the input of the network.

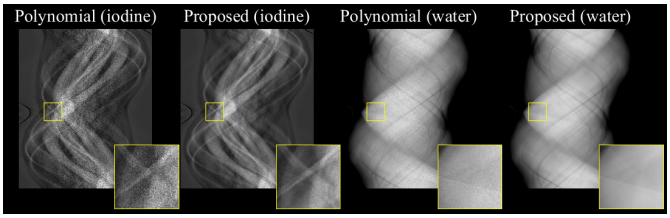

Fig. 9. Comparison of the material basis sinograms from the polynomial projection domain decomposition and after the proposed denoising method for the numerical experiments.

Fig. 8 demonstrates an example of the input low and high energy sinogram as well as the mask sinogram image generated based on the workflow in Fig. 3. The mask image extracted the non-air region of the image slice and was utilized to mitigate the secondary artifacts, whose function will be demonstrated later in the ablation studies. Fig. 9 demonstrates the results of denoised sinograms from the projection domain denoising network. As one can tell, compared with the material basis sinograms calculated from the polynomial fitting process, the proposed method effectively reduces the noise amplitude.

Fig. 10 and Fig. 11 show the results of the proposed method of the digital abdomen phantom as well as the results of the comparison methods. The results from the direct projection domain decomposition using polynomial fitting demonstrates high noise level due to the noise amplification process in the material decomposition calculation, as shown in Fig. 10(a1)-(c1) and Fig. 11(a1)-(c1). The conventional denoising method based on entropy minimization could partially reduce the noise amplitude. However, it leads to signal bias and loss of differentiation of soft tissue contrast. For example, in Fig. 11(b2), the contrast of the water basis image in the zoomed region could not be differentiated. The iodine signal in the zoomed region in Fig. 10(a2) is also darker than the reference image. The three supervised deep learning methods achieve better denoising performance compared with the IDEM method. The performance of the proposed self-supervised method approaches that of the supervised method. For example, for Fig. 10(a6), the proposed method reveals the shape of the renal vein in the water basis image, which is barely visible due to the noise in the decomposed basis image using the polynomial method. The proposed method successfully recovers the soft tissue details in the water basis image: as shown in the zoomed image for the case in Fig. 11(b6), the soft tissue details are well preserved. Such results indicate the effectiveness of the proposed self-supervised network in noise and artifact suppression without the need of training labels.

The quantitative metrics for the digital abdomen phantom are summarized in Table I. The supervised learning method achieves the highest SSIM as well as lowest RMSE. The metrics for the proposed SeNAS-Net are almost the same with the supervised method for the iodine basis images and was slightly inferior than those of the supervised method for the water basis images. The metrics for the conventional IDEM method are much more inferior than neural network-based methods. The quantitative results further confirms the capability of the proposed method in noise and artifact suppression.

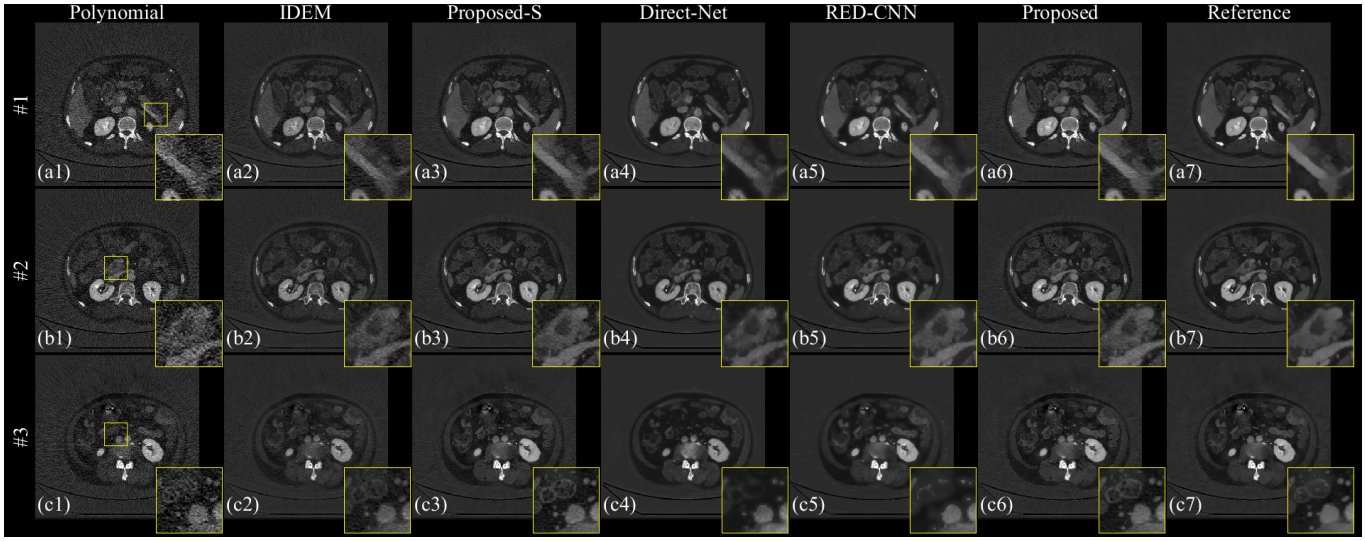

Fig. 10. Comparison of the iodine basis CT images from different methods for the digital abdomen phantom in the simulation study. Display range: [-5 25] mg/mL.

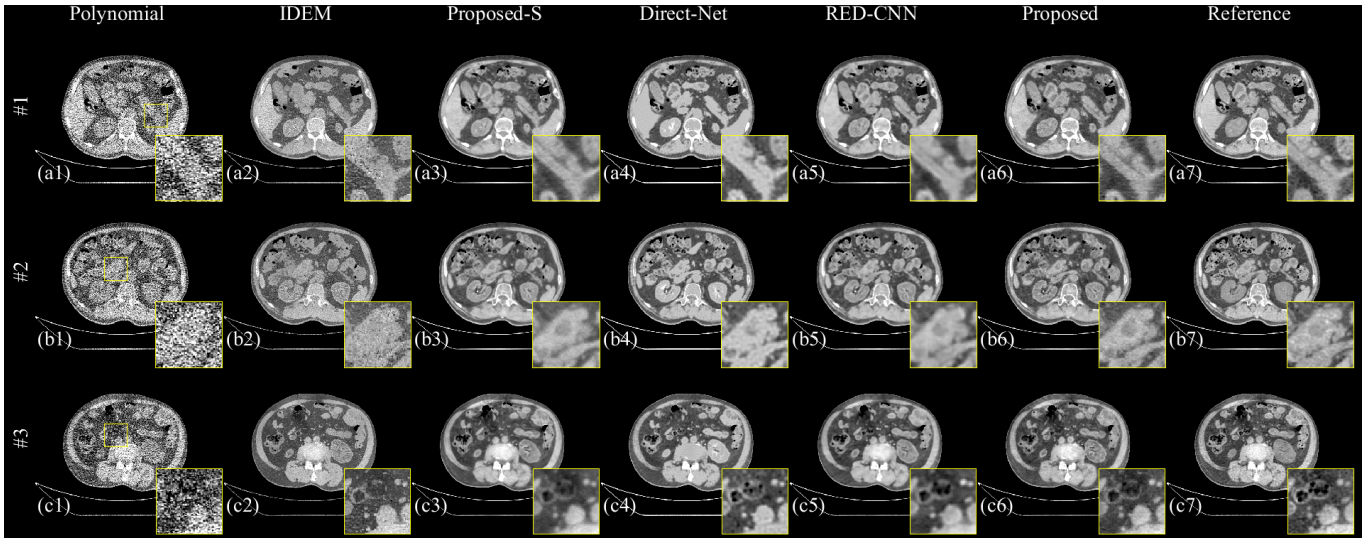

Fig. 11. Comparison of the water basis CT images from different methods for the digital abdomen phantom in the simulation study. Display range: [900 1100] mg/mL for water.

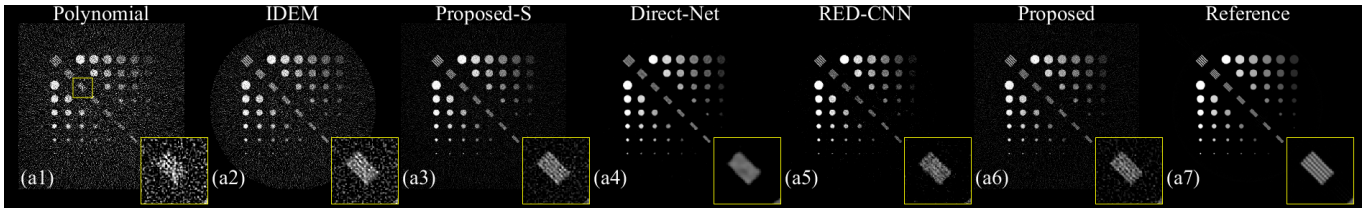

Fig. 12. Comparison of the iodine basis CT images from different methods for the digital quantification phantom. Display range: [0 10] mg/mL.

The reconstructed images of iodine basis of the quantification phantom, processed by different methods, are shown in Fig. 12. The zoomed region demonstrates one of the resolution pattern (7 lp/cm) in the quantification phantom. As one can tell, the proposed method successfully maintains the spatial resolution while suppressing the image noise. Fig. 13 demonstrates the quantitative performance of the proposed method and other comparison methods: the measured concentrations for the 3.3 mg/mL, 6.7 mg/mL and 10 mg/mL iodine inserts are quantified and compared with the ground truth values, as well as the errors. For all inserts, the signal bias of the proposed method is less than 0.3 mg/mL, which indicates that it maintains the signal accuracy.

For the numerical experiments, the training time to conver-

TABLE I
QUANTITATIVE METRICS FOR DIFFERENT METHODS IN THE NUMERICAL STUDY

	Polynomial	IDEM	Proposed-S	Direct-Net	RED-CNN	Proposed
RMSE (iodine) (mg/mL) ↓	0.99 ± 0.06	0.96 ± 0.04	0.39 ± 0.02	0.71 ± 0.07	0.29 ± 0.02	0.41 ± 0.02
RMSE (water) (mg/mL) ↓	61.74 ± 1.76	28.48 ± 0.61	15.50 ± 0.71	18.97 ± 0.79	15.38 ± 0.58	16.98 ± 0.64
SSIM (iodine) ↑	0.944 ± 0.006	0.955 ± 0.004	0.973 ± 0.003	0.964 ± 0.004	0.983 ± 0.002	0.971 ± 0.004
SSIM (water) ↑	0.618 ± 0.018	0.895 ± 0.007	0.970 ± 0.001	0.971 ± 0.001	0.970 ± 0.001	0.962 ± 0.001

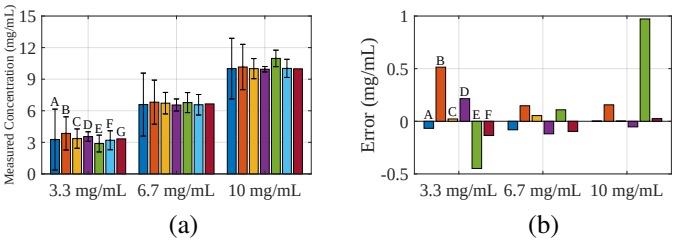

Fig. 13. Bar plots of the quantitative performance of the proposed method as well as other methods for comparison. A: Polynomial, B: IDEM, C: Proposed-S, D: Direction-Net, E: RED-CNN, F: Proposed, G: Reference. (a) Measured concentrations. (b) Errors.

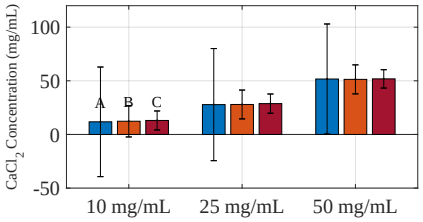

Fig. 15. Comparison of the $CaCl_2$ quantification accuracy of the proposed method and comparison methods. A: Polynomial, B: IDEM, C: Proposed method.

gence was about 50 hours. When the network training was finished, the time to process one image slice was approximately 0.1 s.

B. Results for laboratory experiments

Fig. 14 demonstrates that results of the quantification phantom. For the results using the polynomial method, there exist strong noise in the $CaCl_2$ material basis images. The IDEM method is able to partially mitigate the image noise, but there still exist abnormal dark and bright spots, as shown in the zoomed region in Fig. 14(a2). In contrast, the proposed method demonstrates more robust performance in noise reduction. Results of the quantitative measurements are shown in Fig. 15. The proposed method does not lead to signal bias and the noise standard deviations (error bars) are reduced. For example, for the 50 mg/mL $CaCl_2$ insert, the standard deviation is reduced from 51 mg/mL to 8.5 mg/mL, resulting an increase of signal to noise ratio of approximately 6 times.

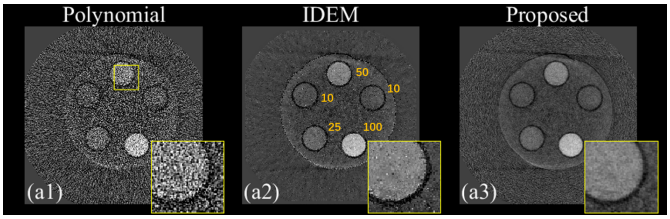

Fig. 14. Comparison of the reconstructed CT images for the $CaCl_2$ material basis from different methods for the laboratory experiment study. The nominal concentrations for each insert are also annotated (unit: mg/mL). Display range: [-20 120] mg/mL.

Fig. 16 shows the comparison of the reconstructed CT images using different methods for the small animal CT scan. The original reconstructed images for the low energy (in even rows) and high energy bin (in odd rows) are also shown as reference images. For the results using the polynomial method, visualization of the details of the soft tissue are compromised due to the amplification of noise in the material decomposition process (Fig. 16(a1)-(d1)). The boundaries of the bone tissue are also obscured due to the image noise. The IDEM method can partially reduce the image noise but the soft tissue details can not be fully recovered (e.g., the lung tissue in Fig. 16(b2), which is similar to the observation for the numerical experiments. There also exists signal bias: for example, the bone signals marked by the arrows of the zoomed regions in Fig. 16(a2) and (c2) are lost. The proposed method effectively reduces the noise of the decomposed CT images. For the bone basis images, the proposed method extracts the bone information while suppressing the background noise, as shown in Fig. 16(a3), which can be told by comparing the decomposed bone basis image and the high energy CT images. For the water basis, the proposed method recovers the soft tissue details in the lung region, as shown in zoomed region in Fig. 16(b3). However, we acknowledge that the performance of the proposed method in this mice study is not as good as that in the numerical simulation study. For example, there remain low-frequency rings artifacts due to the inconsistent energy response of the photon counting detector pixels. Such inconsistency may also impact the noise suppression performance of the proposed method. These non-idealities will be further discussed in the Discussion section.

C. Results for ablation studies

Results of the ablation studies are shown in Fig. 17. For ablation study #1 (direct application of Ne2Ne in the decomposed material basis sinograms without additional constraints), one can observe that the reconstructed basis images demonstrates severe artifacts: there are abnormal blooming and shading regions as well as ring artifacts in Fig. 17(b1). The image reduction is neither as effective as the proposed method, due to the absence of the noise modulation item in (5). Such results indicate the necessity of the additional

TABLE II
QUANTITATIVE METRICS FOR THE ABLATION STUDIES (UNIT FOR RMSE: MG/ML)

	Proposed	Ablation #1	Ablation #2	Ablation #3	Ablation #4	Ablation #5
RMSE (iodine) ↓	0.41 ± 0.02	0.46 ± 0.03	0.40 ± 0.02	0.44 ± 0.02	0.52 ± 0.03	0.69 ± 0.04
RMSE (water) ↓	16.98 ± 0.64	20.67 ± 0.68	18.11 ± 0.43	21.28 ± 0.40	19.91 ± 0.58	38.97 ± 0.95
SSIM (iodine) ↑	0.971 ± 0.004	0.969 ± 0.004	0.971 ± 0.004	0.970 ± 0.004	0.965 ± 0.004	0.957 ± 0.005
SSIM (water) ↑	0.962 ± 0.001	0.950 ± 0.001	0.952 ± 0.001	0.932 ± 0.004	0.943 ± 0.006	0.790 ± 0.012

TABLE III
QUANTITATIVE METRICS FOR THE COMBINATIONS OF DIFFERENT HYPER-PARAMETERS (UNIT FOR RMSE: MG/ML)

	$\alpha = 0.5, \gamma = 2$	$\alpha = 2, \gamma = 2$	$\alpha = 4, \gamma = 2$	$\alpha = 1, \gamma = 1$	$\alpha = 1, \gamma = 4$	$\alpha = 1, \gamma = 16$	Proposed
RMSE (iodine)↓	0.43 ± 0.02	0.48 ± 0.02	0.48 ± 0.02	0.46 ± 0.02	0.58 ± 0.03	0.43 ± 0.02	0.41 ± 0.02
RMSE (water)↓	16.66 ± 0.50	16.60 ± 0.60	16.44 ± 0.51	16.73 ± 0.58	18.45 ± 0.48	25.00 ± 1.13	16.98 ± 0.64
SSIM (iodine) ↑	0.971 ± 0.004	0.967 ± 0.004	0.967 ± 0.004	0.968 ± 0.004	0.963 ± 0.004	0.962 ± 0.003	0.971 ± 0.004
SSIM (water) ↑	0.964 ± 0.001	0.964 ± 0.001	0.965 ± 0.001	0.965 ± 0.001	0.952 ± 0.002	0.903 ± 0.005	0.962 ± 0.001

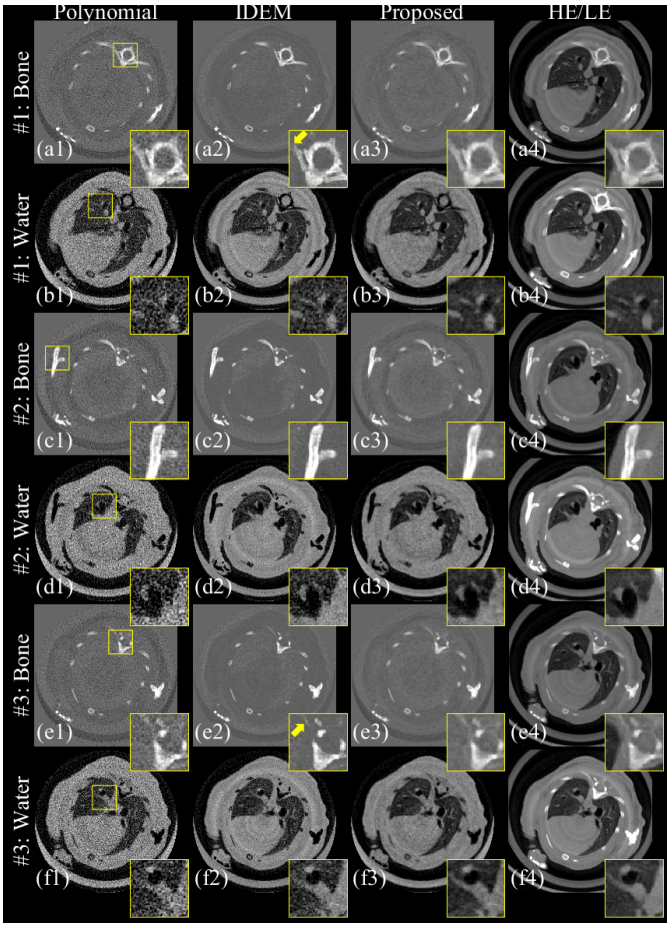

Fig. 16. Comparison of the reconstructed CT images from different methods for the physical experiment study. The original low energy (LE) and high energy (HE) CT images are also shown as references. Display range: [-1 1.5] for bone basis images, [0 2] for water basis images and [0 0.5] cm^{-1} for HE/LE images.

physical constraints of the proposed method. For ablation study #2, if the mask sinogram is not concatenated to the spectral sinograms as inputs, secondary artifacts presented in the reconstructed basis images as radial streak artifacts close to the high contrast region, e.g. air sacs in human abdomen. Such artifacts can be observed in the zoomed region of Fig. 17(b2) and (d2). For ablation study #3, if the data fidelity loss item is removed in (16), there are remaining noise streaks and loss of differentiation of the soft tissue details, as shown in the zoomed region in Fig. 17(b3) and (d3). Meanwhile, there exist stronger signal bias, which can be told from the increase of the RMSE in the quantitative metrics in Table II. For ablation study #4, if the noise modulation item in (5) is not applied, the reconstructed material basis images demonstrate abnormal coarse noise structure, which is due to the error of noise estimation in the projection domain without consideration of the anisotropic noise amplitudes. For ablation study #5, when the loss items related to the Ne2Ne method are calculated in the raw sinogram domain, noise level is not high enough to meet the criterion of the Ne2Ne: as shown in the first column of Fig. 17, the noise level is still high after the denoising network. As a result, the soft tissue details is compromised, as shown in the zoomed region in Fig. 17(d5). The quantitative metrics are summarized in Table II. For the water basis, the metrics of all ablation studies are inferior to that of the proposed method. The metrics of ablation #2 are slightly better than the proposed method for the iodine basis image (but is not statistically significant considering the standard deviations). This is consistent with the visual observation as the streak artifacts for the ablation study #2 is more obvious in the water basis images.

Results of the ablation studies when the values of the hyper-parameters in (16) are tuned are shown in Fig. 18. For the impact of coefficient α (coefficient for the data fidelity loss item), results demonstrate that the water basis images are not very sensitive to the value of α. As α increases, the noise of the iodine basis images slightly increases, which can be told by comparing Fig. 18(a1) - (a3), but the difference is not significant. The quantitative results in Table III confirm the above visual observations. For the impact of coefficient γ (coefficient for the reconstruction loss item), the difference between the case when $\gamma = 1$ and the proposed method ($\gamma = 2$) is not very obvious: there is a slightly better reduction of noise streaks in the water basis image while the noise in the iodine basis image is less suppressed for $\gamma = 1$.

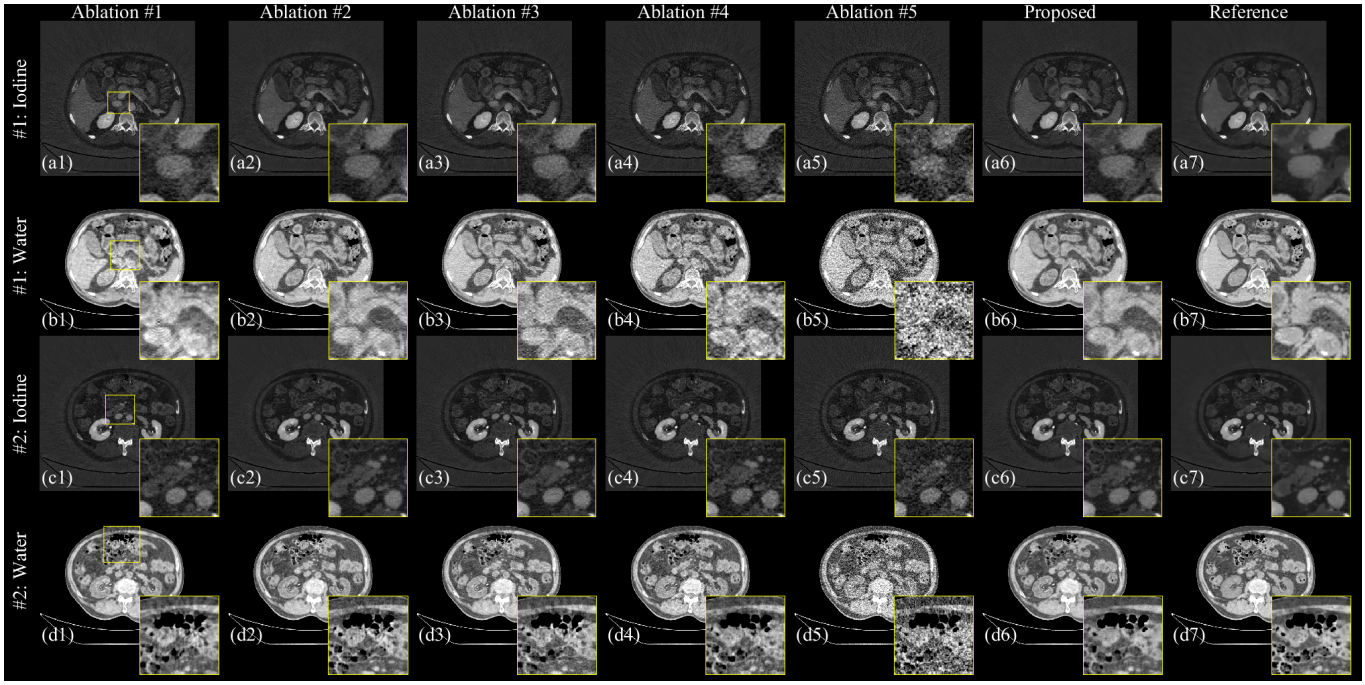

Fig. 17. Results of the ablation studies to demonstrate the impact of the additional constraints. Display range: [-5 25] mg/mL for iodine and [900 1100] mg/mL for water.

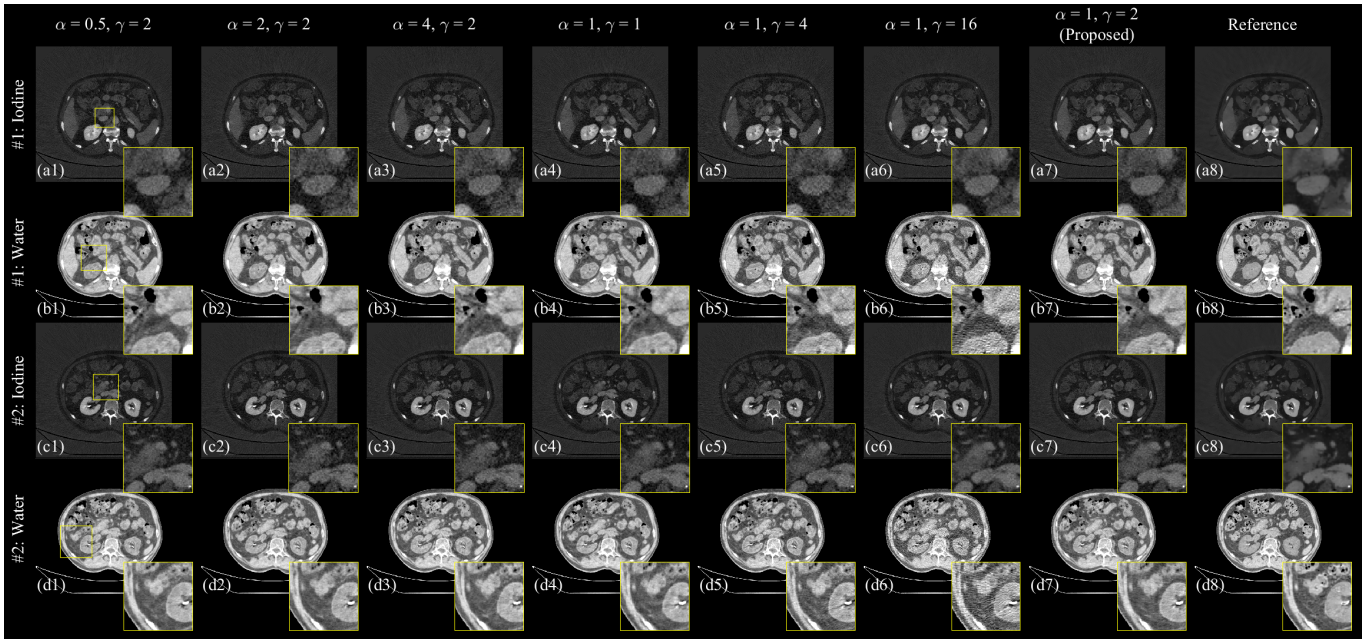

Fig. 18. Results of the ablation study when the hyper-parameters in (16) are tuned. Note that the parameter $\beta = 1$. Display range: [-5 25] mg/mL for iodine and [900 1100] mg/mL for water.

As γ further increases, the noise streaks in the water basis images become more and more severe, which can be told by comparing Fig. 18(b4) - (b6). Such result is also reflected in the quantitative results in Table III, as the RMSE for the water basis images increases when γ increases. The above results indicate that the selection of the parameter values for the proposed method ($\alpha = 1, \beta = 1, \gamma = 2$) is an quasi-optimal one.

IV. Discussion and Conclusion

In this work, a self-supervised noise and artifact suppression method (SeNAS-Net) for spectral CT is proposed. The method incorporates the idea of Neighbor2Neighbor denoising strategy to realize self-supervised learning without the need of training labels. The method is physics-driven by including material decomposition process in the calculation of the reconstruction and regularization loss in the projection domain. Meanwhile,

several additional constraints were incorporated in the network to suppress the secondary artifacts due to the random down-sample operators in the projection domain network. The proposed method is validated using both numerically simulated data and real experimental data. Results demonstrated that the performance of the proposed method is close to a supervised learning method which required the ground truth values of the material basis images as the training label.

The neighbour2neighbour is realized in the projection domain rather than in the image domain in this work, due to the fact that in image domain, the pixels have long range correlation so the prerequisites of the Ne2Ne method are violated. The correlation in the projection domain only comes from the correlation between the detectors pixels, and correlation length is much shorter. Projection domain operations inevitably lead to secondary artifacts in the reconstructed CT images. This work introduces a mask image as the input along with several constraints to mitigate such artifacts. However, there still remain certain discrepancies between the decomposed basis images using the proposed method and the ground truth images. For example, in Fig. 11(b7), the iodine signal in the center of the right kidney erroneously leaked to the water basis image, presenting as a bright dot. Additional constraints may help to mitigates such secondary artifacts and the proposed method may be further improved.

There are several additional limitations related to the study. The data fidelity loss item for the proposed method requires spectral information. Such information may not be easy to acquire for real CT system: even the output spectra of the source can be estimated by measuring the half-value layer, the spectra can be distorted due to the energy responses of the x-ray detectors. For the real experimental data in this experiment, due to the fact that the image object is small, the spectra is approximated to be monoenergetic so the problem is partially resolved. However, even in this case, the detected spectra across different pixels can be different so that ring artifacts remain in the decomposed CT images, as shown in Fig. 16. Pixel-specific spectral calibration may mitigate such problem. For large objects, the spectra need to be relatively precisely modeled to reflect the effect caused by the polychromaticity. Finally, although the work only covers experiments for dual energy CT, the proposed method can be revised for multi-material decomposition in principle. However, several issues may arise for multi-material decomposition. For example, the polynomial coefficients for multi-materials decomposition can be hard to acquire, especially for the high order items, so that the initial decomposition may contain large signal bias. How the proposed method is applied to multi-material decomposition can be an interesting future topic.

REFERENCES

[1] R. E. Alvarez and A. Macovski, "Energy-selective reconstructions in x-ray computerised tomography," *Phys. Med. Biol.*, vol. 21, no. 5, pp. 733–744, Sep. 1976.

[2] T. Yoshizumi, "Dual energy CT in clinical practice," *Med. Phys.*, vol. 38, no. 11, p. 6346, Nov. 2011.

[3] T. Johnson, C. Fink, S. O. Schönberg, and M. F. Reiser, Eds., *Dual energy CT in clinical practice*, ser. Diagnostic Imaging. Berlin, Germany: Springer, Nov. 2010.

[4] L. A. Lehmann, R. E. Alvarez, A. Macovski, W. R. Brody, N. J. Pelc, S. J. Riederer, and A. L. Hall, "Generalized image combinations in dual KVP digital radiography," *Med. Phys.*, vol. 8, no. 5, pp. 659–667, Sep. 1981.

[5] A. D. Laidevant, S. Malkov, C. I. Flowers, K. Kerlikowske, and J. A. Shepherd, "Compositional breast imaging using a dual-energy mammography protocol," *Med. Phys.*, vol. 37, no. 1, pp. 164–174, Jan. 2010.

[6] X. Liu, L. Yu, A. N. Primak, and C. H. McCollough, "Quantitative imaging of element composition and mass fraction using dual-energy CT: three-material decomposition," *Med. Phys.*, vol. 36, no. 5, pp. 1602–1609, May 2009.

[7] P. R. S. Mendonca, P. Lamb, and D. V. Sahani, "A flexible method for multi-material decomposition of dual-energy CT images," *IEEE Trans. Med. Imaging*, vol. 33, no. 1, pp. 99–116, Jan. 2014.

[8] S. J. Riederer and C. A. Mistretta, "Selective iodine imaging using k-edge energies in computerized x-ray tomography," *Med. Phys.*, vol. 4, no. 6, pp. 474–481, Nov. 1977.

[9] E. Y. Sidky, Y. Zou, and X. Pan, "Impact of polychromatic x-ray sources on helical, cone-beam computed tomography and dual-energy methods," *Phys. Med. Biol.*, vol. 49, no. 11, pp. 2293–2303, Jun. 2004.

[10] 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.*, vol. 36, no. 8, pp. 3818–3829, Aug. 2009.

[11] L. Yu, J. A. Christner, S. Leng, J. Wang, J. G. Fletcher, and C. H. McCollough, "Virtual monochromatic imaging in dual-source dual-energy CT: radiation dose and image quality," *Med. Phys.*, vol. 38, no. 12, pp. 6371–6379, Dec. 2011.

[12] T. Niu, X. Dong, M. Petrongolo, and L. Zhu, "Iterative image-domain decomposition for dual-energy CT," *Med. Phys.*, vol. 41, no. 4, p. 041901, Apr. 2014.

[13] W. Zhao, T. Niu, L. Xing, Y. Xie, G. Xiong, K. Elmore, J. Zhu, L. Wang, and J. K. Min, "Using edge-preserving algorithm with non-local mean for significantly improved image-domain material decomposition in dual-energy CT," *Phys. Med. Biol.*, vol. 61, no. 3, pp. 1332–1351, Feb. 2016.

[14] J. Harms, T. Wang, M. Petrongolo, T. Niu, and L. Zhu, "Noise suppression for dual-energy CT via penalized weighted least-square optimization with similarity-based regularization," *Med. Phys.*, vol. 43, no. 5, p. 2676, May 2016.

[15] X. Dong, T. Niu, and L. Zhu, "Combined iterative reconstruction and image-domain decomposition for dual energy CT using total-variation regularization," *Med. Phys.*, vol. 41, no. 5, p. 051909, May 2014.

[16] W. A. Kalender, E. Klotz, and L. Kostaridou, "An algorithm for noise suppression in dual energy CT material density images," *IEEE Trans. Med. Imaging*, vol. 7, no. 3, pp. 218–224, 1988.

[17] M. Petrongolo and L. Zhu, "Noise suppression for dual-energy CT through entropy minimization," *IEEE Trans. Med. Imaging*, vol. 34, no. 11, pp. 2286–2297, Nov. 2015.

[18] Y. Jiang, Y. Xue, Q. Lyu, L. Xu, C. Luo, P. Yang, C. Yang, J. Wang, X. Hu, X. Zhang, K. Sheng, and T. Niu, "Noise suppression in image-domain multi-material decomposition for dual-energy CT," *IEEE Trans. Biomed. Eng.*, vol. 67, no. 2, pp. 523–535, Feb. 2020.

[19] Y. Long and J. A. Fessler, "Multi-material decomposition using statistical image reconstruction for spectral CT," *IEEE Trans. Med. Imaging*, vol. 33, no. 8, pp. 1614–1626, Aug. 2014.

[20] C. T. Badea, M. Holbrook, and D. P. Clark, "Multi-energy CT decomposition using convolutional neural networks," in *Medical Imaging 2018: Physics of Medical Imaging*, G.-H. Chen, J. Y. Lo, and T. Gilat Schmidt, Eds. SPIE, Mar. 2018.

[21] 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.*, vol. 2018, p. 2527516, Sep. 2018.

[22] Z. Chen and L. Li, "Robust multimaterial decomposition of spectral CT using convolutional neural networks," *Opt. Eng.*, vol. 58, no. 01, p. 1, Jan. 2019.

[23] K. C. Zimmerman, G. Sharma, A. K. Parchur, A. Joshi, and T. G. Schmidt, "Experimental investigation of neural network estimator and transfer learning techniques for k-edge spectral CT imaging," *Med. Phys.*, vol. 47, no. 2, pp. 541–551, Feb. 2020.

[24] W. Zhang, H. Zhang, L. Wang, X. Wang, X. Hu, A. Cai, L. Li, T. Niu, and B. Yan, "Image domain dual material decomposition for dual-energy CT using butterfly network," *Med. Phys.*, vol. 46, no. 5, pp. 2037–2051, May 2019.

[25] T. Su, X. Sun, J. Yang, D. Mi, Y. Zhang, H. Wu, S. Fang, Y. Chen, H. Zheng, D. Liang, and Y. Ge, "DIRECT-Net: A unified mutual-domain material decomposition network for quantitative dual-energy CT imaging," *Med. Phys.*, vol. 49, no. 2, pp. 917–934, Feb. 2022.

[26] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2noise: Learning image restoration without clean data," 2018.

[27] 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.*, vol. 66, no. 15, p. 155013, Jul. 2021.

[28] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, "Neighbor2neighbor: Self-supervised denoising from single noisy images," 2021.

[29] Q. Wu, X. Ji, Y. Gu, J. Xiang, G. Quan, B. Li, J. Zhu, G. Coatrieux, J.-L. Coatrieux, and Y. Chen, "Unsharp structure guided filtering for self-supervised low-dose CT imaging," *IEEE Trans. Med. Imaging*, vol. PP, May 2023.

[30] J. Baek and N. J. Pelc, "The noise power spectrum in CT with direct fan beam reconstruction," *Med. Phys.*, vol. 37, no. 5, pp. 2074–2081, May 2010.

[31] W.-A. Lin, H. Liao, C. Peng, X. Sun, J. Zhang, J. Luo, R. Chellappa, and S. K. Zhou, "DuDoNet: Dual domain network for CT metal artifact reduction," 2019.

[32] P. Stenner, T. Berkus, and M. Kachelriess, "Empirical dual energy calibration (EDEC) for cone-beam computed tomography," *Med. Phys.*, vol. 34, no. 9, pp. 3630–3641, Sep. 2007.

[33] X. Ji, R. Zhang, Y. Ge, G.-H. Chen, and K. Li, "Signal and noise characteristics of a CdTe-based photon counting detector: cascaded systems analysis and experimental studies," in *Medical Imaging 2017: Physics of Medical Imaging*, T. G. Flohr, J. Y. Lo, and T. Gilat Schmidt, Eds. SPIE, Mar. 2017.

[34] MandoCT, [EB/OL], https://github.com/SEU-CT-Recon/mandoct Accessed Oct. 10, 2023.

[35] NIST, [EB/OL], https://www.nist.gov/pml/x-ray-mass-attenuation-coefficients, Standard Reference Database 126, Accessed Sept. 2, 2023.

[36] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Med. Imaging*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.