

PAPER

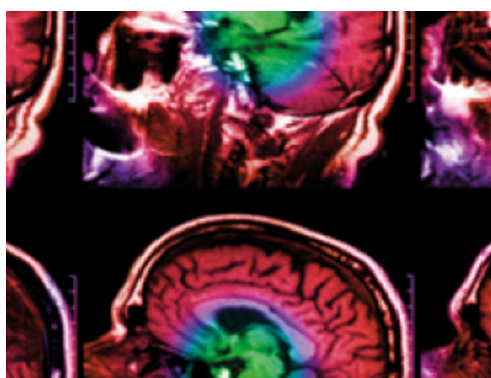
Iterative material decomposition for spectral CT using self-supervised Noise2Noise prior

To cite this article: Wei Fang *et al* 2021 *Phys. Med. Biol.* **66** 155013

View the [article online](#) for updates and enhancements.

You may also like

- [Nonlocal low-rank and sparse matrix decomposition for spectral CT reconstruction](#)
Shanzhou Niu, Gaochang Yu, Jianhua Ma et al.
- [Photon-counting x-ray detectors for CT](#)
Mats Danielsson, Mats Persson and Martin Sjölin
- [An algorithm for constrained one-step inversion of spectral CT data](#)
Rina Foygel Barber, Emil Y Sidky, Taly Gilat Schmidt et al.



IPEM | IOP

Series in Physics and Engineering in Medicine and Biology


Your publishing choice in medical physics,
biomedical engineering and related subjects.

Start exploring the collection—download the
first chapter of every title for free.



PAPER

Iterative material decomposition for spectral CT using self-supervised Noise2Noise prior

Wei Fang^{1,2}, Dufan Wu^{2,3} , Kyungsang Kim^{2,3}, Mannudeep K Kalra⁴, Ramandeep Singh⁴, Liang Li^{1,*} and Quanzheng Li^{2,3,*}¹ Department of Engineering Physics, Tsinghua University, Beijing, 100084, People's Republic of China² Center for Advanced Medical Computing and Analysis, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, United States of America³ Gordon Center for Medical Imaging, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, United States of America⁴ Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Boston, MA 02114, United States of America

* Authors to whom any correspondence should be addressed.

E-mail: liang@tsinghua.edu.cn and li.quanzheng@mgm.harvard.edu**Keywords:** Noise2Noise, spectral CT, material decomposition, denoising, self-supervised deep learning**Abstract**

Compared to conventional computed tomography (CT), spectral CT can provide the capability of material decomposition, which can be used in many clinical diagnosis applications. However, the decomposed images can be very noisy due to the dose limit in CT scanning and the noise magnification of the material decomposition process. To alleviate this situation, we proposed an iterative one-step inversion material decomposition algorithm with a Noise2Noise prior. The algorithm estimated material images directly from projection data and used a Noise2Noise prior for denoising. In contrast to supervised deep learning methods, the designed Noise2Noise prior was built based on self-supervised learning and did not need external data for training. In our method, the data consistency term and the Noise2Noise network were alternatively optimized in the iterative framework, respectively, using a separable quadratic surrogate (SQS) and the Adam algorithm. The proposed iterative algorithm was validated and compared to other methods on simulated spectral CT data, preclinical photon-counting CT data and clinical dual-energy CT data. Quantitative analysis showed that our proposed method performs promisingly on noise suppression and structure detail recovery.

1. Introduction

Spectral computed tomography (SCT), with dual energy CT (DECT) as a special case, can provide a material decomposition function by measuring the attenuation of objects under different spectra (Alvarez and Macovski 1976, Li *et al* 2014). Clinically useful information such as iodine maps (Chandarana *et al* 2011, Pourmorteza *et al* 2016), virtual nonenhanced CT images (Zhang *et al* 2010) and pseudo-monochromatic images (Yu *et al* 2012) can be obtained after applying material decomposition. It has been demonstrated that the material decomposition can benefit many clinical applications such as kidney stone characterization (Primak *et al* 2007), various intrathoracic diseases including pulmonary perfusion in acute and chronic pulmonary thromboemboli (Thieme *et al* 2008, Otrakji *et al* 2016), osteocartilaginous diseases (Rajendran *et al* 2017), bone removal in CT angiography (McCollough *et al* 2015) and intravenous contrast volume reduction (Digumarthy *et al* 2020).

The ability of spectral CT to decompose depends on the physical fact that different materials have different attenuation decreasing curves along the x-ray energy. The energy-dependent attenuation of objects can be expressed as a linear combination of photoelectric absorption and the Compton scattering effect (Alvarez and Macovski 1976, Li *et al* 2018). Or more practically, the attenuation can be represented with a linear combination of the attenuation of several underlying basis materials (Lehmann *et al* 1981, Yao *et al* 2019). However, the challenge for material decomposition of spectral CT is that the generated material-specific images can be very

RECEIVED
25 January 2021REVISED
10 June 2021ACCEPTED FOR PUBLICATION
14 June 2021PUBLISHED
27 July 2021

noisy, which will significantly limit the clinical value of spectral CT, especially at low radiation doses and in patients with larger body habitus, or with arms by the side of the body when performing chest or abdominal CT. There are generally two reasons for this. The first one is that spectral CT has fewer photon counts in each energy bin than the total number of photons exiting the patient, compared to conventional CT scanners. The second reason is that the material decomposition process itself is an ill-conditioned problem that will significantly magnify the noise. Many algorithms have been put forward to tackle this problem. These material decomposition algorithms can be generally divided into three categories: projection domain-based, image domain-based and one-step inversion methods. Projection-domain methods firstly decompose the projection data into a basis-specific sinogram and then perform a tomographic reconstruction to get the decomposed images (Alvarez and Macovski 1976, Roessl and Proksa 2007, Schlomka *et al* 2008). Image domain methods reconstruct the images of each energy bin first and then perform the decomposition process in the image domain (Maaß *et al* 2009, Le and Molloy 2011, Mendonça *et al* 2013, Niu *et al* 2014, Harms *et al* 2016, Zhao *et al* 2016, Xue *et al* 2017). Direct reconstruction methods are mathematically most elegant as they incorporate the tomographic reconstruction and decomposition process into one model and estimate decomposition images directly from the projection data (Elbakri and Fessler 2002, Cai *et al* 2013, Zhang *et al* 2013, Long and Fessler 2014, Zhao *et al* 2014, Barber *et al* 2016, Mory *et al* 2018). Most of the existing material decomposition algorithms deal with the noise using hand-crafted regularization terms based on explicit prior knowledge, such as continuity and sparsity, which cannot thoroughly express the knowledge of clean images. A more adaptive, implicit and delicate way of prior information designing using a data-driven method should be applied to better depict the underlying knowledge of noise-free images.

Recently, deep learning has shown its great success in several image-based tasks (Krizhevsky *et al* 2012). Many deep-learning-based algorithms have been proposed for CT image reconstruction and denoising (Chen *et al* 2017, Jin *et al* 2017, Kang *et al* 2017, Wu *et al* 2017, Adler and Öktem 2018, Fang *et al* 2020). For material decomposition, neural networks have also been used to approximate the nonlinear mapping function from projection data to a decomposed material-specific sinogram in the projection domain (Zimmerman *et al* 2020) and the linear mapping function from reconstruction to material images in the image domain (Clark *et al* 2018, Xu *et al* 2018, Chen and Li 2019, Zhang *et al* 2019). However, most of these existing algorithms are supervised learning methods that need to train the neural network in advance with training pairs, which may not apply well when a domain mismatch exists. Recently, an unsupervised denoising method called Noise2Noise has been proposed, where the network is trained by mapping one noisy realization to another noisy realization without the need for any clean images (Lehtinen *et al* 2018). Noise2Noise was first proposed for image restoration and then applied to denoising in many medical imaging modalities, such as positron emission tomography (PET) (Wu *et al* 2020a) and dynamic CT (Wu *et al* 2020b). Wu *et al* applied Noise2Noise in low-dose CT denoising by splitting measurements in the angular dimension as odd and even projections to generate different noisy realizations for training (Wu *et al* 2019a). In addition, Yuan *et al* used Noise2Noise for low-dose CT denoising respectively in the sinogram domain and the image domain by splitting count data in each energy bin as half dose to generate paired noisy realizations for training (Yuan *et al* 2020).

In contrast to previous work that utilizes Noise2Noise denoising as a post-processing method, in this paper, we propose to use a self-supervised Noise2Noise network as a prior for material image denoising in a framework of the one-step inversion approach. In our method, two sets of material images were firstly generated respectively from two sets of angularly split sinogram data by performing filtered backprojection and direct matrix inversion. An encoder-decoder network was trained on the two sets of material images by mapping them to each other, which outputs denoised material images. A cost function was built to jointly optimize the reconstructed material images and the encoder-decoder neural network. The final reconstructed material images were both constrained by the data fidelity term and the distance to the network denoised output, indicating better image detail reservation and noise suppression. The proposed method was compared with the total variation penalized one-step inversion material decomposition method and post-processing matrix inversion method on a simulated spectral CT dataset, preclinical photon-counting CT dataset and clinical dual-energy CT dataset, showing a promising performance regarding noise suppression and structure detail recovery.

2. Methods

2.1. Spectral CT reconstruction

For spectral CT, the mean value of the attenuated photon counts for the i th ray in the m th energy bin \bar{Y}_i^m after passing through an object can be illustrated as:

$$\bar{Y}_i^m = \int_{bin(E^m)} I_i(E) \exp\left(-\int_{L_i} \mu(x, y, z, E) dl\right) dE \approx I_i^m \exp\left(-\int_{L_i} \mu^m(x, y, z, E^m) dl\right), \quad (1)$$

where $I_i(E)$ is the incident photon counts at energy E , $\mu(x, y, z, E)$ is the spatially and energy-dependent linear attenuation coefficient and L_i is the i th x-ray path. As displayed in the second part of the equation, the attenuated photon counts can be approximated using a monoenergetic-like expression, where I_i^m is the number of incident photon counts of the m th energy bin and $\mu^m(x, y, z, E^m)$ represents the effective linear attenuation coefficient of energy bin E^m . In the theory of basis material decomposition, the effective linear attenuation coefficient for each pixel in the CT image can be approximated as the combination of linear attenuation coefficients of several basis materials, which can be represented as:

$$\mu^m = \sum_{k=1}^K \mu_k^m \mathbf{b}_k, \quad (2)$$

where μ^m is the effective linear attenuation coefficient of an unknown material in the m th energy bin, μ_k^m is the effective linear attenuation coefficient in the m th energy bin for the k th basis material, \mathbf{b}_k represents the normalized density of the k th basis material and is unitless. Assuming the measured photon counts \mathbf{Y}_i^m following Poisson distribution, we can depict it as:

$$\mathbf{Y}_i^m \sim \text{Poisson}\left\{I_i^m e^{-\left[\mathbf{A}\left(\sum_{k=1}^K \mu_k^m \mathbf{b}_k\right)\right]_i}\right\}, \quad i = 1, \dots, N_{ray} \quad (3)$$

where I_i^m is the total number of incident photon counts of the m th energy bin and N_{ray} is the number of measured rays. The notation \mathbf{A} represents the system matrix which accounts for the geometric factor and $[\mathbf{A}(\sum_{k=1}^K \mu_k^m \mathbf{b}_k)]_i$ denotes the attenuation integral of the i th line ray. The negative log-likelihood function can be given by:

$$-L(\mathbf{b}) = \sum_{m=1}^M \sum_{i=1}^N -Y_i^m \log\left(I_i^m e^{-\left[\mathbf{A}\left(\sum_{k=1}^K \mu_k^m \mathbf{b}_k\right)\right]_i}\right) + I_i^m e^{-\left[\mathbf{A}\left(\sum_{k=1}^K \mu_k^m \mathbf{b}_k\right)\right]_i}, \quad (4)$$

where Y_i^m is the actual measured photon counts. For convenience, we can rewrite equation (4) as:

$$-L(\mathbf{b}) = \sum_{m=1}^M \sum_{i=1}^N g_i^m(l_i^m), \quad (5)$$

where $g_i^m(l_i^m) = -Y_i^m \log(I_i^m e^{-l_i^m}) + I_i^m e^{-l_i^m}$ is the negative-log-likelihood for each line ray and l_i^m equals $[\mathbf{A}(\sum_{k=1}^K \mu_k^m \mathbf{b}_k)]_i$. Applying a second-order Taylor's expansion to function g_i^m at an estimate \hat{l}_i^m we achieve

$$g_i^m(l_i^m) \approx g_i^m(\hat{l}_i^m) + \dot{g}_i^m(\hat{l}_i^m)(l_i^m - \hat{l}_i^m) + \frac{\ddot{g}_i^m(\hat{l}_i^m)}{2}(l_i^m - \hat{l}_i^m)^2, \quad (6)$$

where \dot{g}_i^m and \ddot{g}_i^m are respectively the first and second order derivatives of g_i^m . Substituting the derivatives and setting the estimate \hat{l}_i^m with $-\log(Y_i^m/I_i^m)$ gives the following approximation for g_i^m :

$$g_i^m(l_i^m) \approx Y_i^m - Y_i^m \log Y_i^m + \frac{Y_i^m}{2}(l_i^m - \hat{l}_i^m)^2. \quad (7)$$

Omitting the first two terms which are not related to l_i^m , we can get the cost function for spectral CT material image reconstruction:

$$\Phi_1(\mathbf{b}) = \sum_{m=1}^M \sum_{i=1}^N \frac{Y_i^m}{2} \left(\left[\mathbf{A}\left(\sum_{k=1}^K \mu_k^m \mathbf{b}_k\right) \right]_i - \hat{l}_i^m \right)^2. \quad (8)$$

2.2. Noise2Noise denoising

The traditional deep learning method for denoising based on Noise2Clean mapping aims to minimize such a loss:

$$\Theta^* = \arg \min_{\Theta} \frac{1}{N} \sum_i \|f(\mathbf{z}_i + \mathbf{n}_i; \Theta) - \mathbf{z}_i\|_2^2, \quad (9)$$

where \mathbf{z}_i is the ground truth of the i th image and \mathbf{n}_i denotes the corresponding noise. The mapping function $f(\mathbf{z}; \Theta): \mathbb{R} \rightarrow \mathbb{R}$ represents the neural network and Θ is its trainable parameters. N denotes the number of training samples. In contrast to Noise2Clean learning, the Noise2Noise framework uses another independent

noisy realization as the training label, which can be illustrated as:

$$\Theta^* = \arg \min_{\Theta} \frac{1}{N} \sum_i \|f(\mathbf{z}_i + \mathbf{n}_{i1}; \Theta) - (\mathbf{z}_i + \mathbf{n}_{i2})\|_2^2, \quad (10)$$

where \mathbf{n}_{i1} and \mathbf{n}_{i2} are two independent noise realizations. It has been demonstrated that under certain mild conditions, Noise2Noise training is equivalent to Noise2Clean training, which can be summarized as the following theorem (Wu *et al* 2019a).

Theorem 1. The following equation holds:

$$\frac{1}{N} \sum_i \|f(\mathbf{z}_i + \mathbf{n}_{i1}; \Theta) - (\mathbf{z}_i + \mathbf{n}_{i2})\|_2^2 = \frac{1}{N} \sum_i \|f(\mathbf{z}_i + \mathbf{n}_i; \Theta) - \mathbf{z}_i\|_2^2 + C, \quad (11)$$

where C is irrelevant to Θ , if the following conditions are satisfied:

1. $N \rightarrow \infty$;
2. Conditional expectation $E\{\mathbf{n}_{i2}|\mathbf{z}_i\} = 0$;
3. \mathbf{n}_{i1} and \mathbf{n}_{i2} are independent;
4. $\forall i, |f(\mathbf{z}_i + \mathbf{n}_{i1}; \Theta)| < \infty$.

Theorem 1 states that Noise2Noise loss can be understood as a surrogate for Noise2Clean loss when optimizing regarding Θ . One can refer to this publication (Wu *et al* 2019a) for detailed proof. As long as the network has meaningful parameters, condition 4 can be easily satisfied. In this paper, the process of obtaining two independent and zero-mean noise realizations for the material images was as follows. Firstly, we performed an angular separation to the sinogram data to get odd and even projections. It should be noted that the odd-even splitting was not related to kVp switching. It does not mean that odd projections were low-energy data and even projections were high-energy data and vice versa. We split odd and even projections from the projection data that were under the same spectrum. After that, two sets of material images were acquired by respectively conducting FBP reconstruction and direct matrix inversion from the two projection datasets. The two sets of material images were related but the noise in the two sets of material images can be approximately regarded as independent and zero-mean. Thus conditions 2 and 3 are also satisfied. Considering the filter in convolutional neural networks is shift-invariant, different parts of the image actually serve as multiple training samples. Therefore, the number of training samples N is large enough even when only one image is used for training.

To efficiently utilize both noisy realizations, we can write the Noise2Noise loss function for CT material image denoising as follows:

$$\begin{aligned} \Theta^* = \arg \min_{\Theta} \frac{1}{N} \sum_i \frac{1}{2} \|f(\mathbf{z}_{i1}; \Theta) - \mathbf{z}_{i2}\|_2^2 + \frac{1}{2} \|f(\mathbf{z}_{i2}; \Theta) - \mathbf{z}_{i1}\|_2^2 \\ + \gamma_d \left\| \frac{f(\mathbf{z}_{i1}; \Theta) + f(\mathbf{z}_{i2}; \Theta)}{2} - \frac{\mathbf{z}_{i1} + \mathbf{z}_{i2}}{2} \right\|_2^2, \end{aligned} \quad (12)$$

And with the denoised material image given by:

$$\mathbf{y}_i = \frac{f(\mathbf{z}_{i1}; \Theta^*) + f(\mathbf{z}_{i2}; \Theta^*)}{2}, \quad (13)$$

where \mathbf{z}_{i1} and \mathbf{z}_{i2} are respectively two noisy realizations of material images, which are calculated from two separated projections. γ_d is a hyperparameter for the regularization term. The regularization term will help to prevent the over-smoothing of network output images by constraining its distance to the material images that are calculated by using FBP reconstruction and a subsequent matrix inversion.

2.3. The whole loss function and its optimization

Instead of simply using Noise2Noise as a post-processing tool, the material images were also constrained by the projection data in our method. By incorporating the Noise2Noise prior into the iterative material reconstruction framework, the whole loss function can be written as:

$$\begin{aligned}
\mathbf{b}^*, \Theta^* = \arg \min_{\mathbf{b}, \Theta} & \frac{1}{\|\mathbf{A}\|_{\mathbf{w}}^2} \sum_{m=1}^M \sum_{i=1}^N \frac{Y_i^m}{2} \left(\left[\mathbf{A} \left(\sum_{k=1}^K \mu_k^m \mathbf{b}_k \right) \right]_i - \hat{l}_i^m \right)^2 \\
& + \beta \gamma \left\| \mathbf{b} - \frac{f(\mathbf{z}_1; \Theta) + f(\mathbf{z}_2; \Theta)}{2} \right\|_2^2 \\
& + \frac{\beta}{2} (\|f(\mathbf{z}_1; \Theta) - \mathbf{z}_2\|_2^2 + \|f(\mathbf{z}_2; \Theta) - \mathbf{z}_1\|_2^2),
\end{aligned} \quad (14)$$

where \mathbf{A} is the system matrix, Y_i^m is the measured photon counts for the i th ray in the m th energy bin, μ_k^m is the effective linear attenuation coefficient of the k th basis material in the m th energy bin, \mathbf{b} is the material image to be calculated, \hat{l}_i^m is the projection data and \mathbf{w} is the noise weighting matrix. $\|\mathbf{A}\|_{\mathbf{w}}^2$ is the square of the w -norm of the system matrix, equalling $\|\mathbf{A}^T \mathbf{w} \mathbf{A}\|_2$, which can be calculated by using the power method. $\|\mathbf{A}\|_{\mathbf{w}}^2$ is used for the data fidelity term normalization to make hyperparameter tuning easier. The second term is a cross term, containing both the material image \mathbf{b} and network parameters Θ . The third term is the Noise2Noise loss term, where \mathbf{z}_1 and \mathbf{z}_2 are two noisy implementations of material images.

The above loss function can be optimized by alternatively solving two subproblems. If the two subproblems can be optimized monotonically, the entire optimization will be monotonic. The first one is a regularized material image reconstruction problem, which is to minimize the first two terms of the whole loss function with respect to material image \mathbf{b} . We use the monotonic separable quadratic surrogate (SQS) (Erdogan and Fessler 1999) method for its optimization. The loss function is:

$$\begin{aligned}
g_1(\mathbf{b}; \Theta^{(n)}) = & \frac{1}{\|\mathbf{A}\|_{\mathbf{w}}^2} \sum_{m=1}^M \sum_{i=1}^N \frac{Y_i^m}{2} \left(\left[\mathbf{A} \left(\sum_{k=1}^K \mu_k^m \mathbf{b}_k \right) \right]_i - \hat{l}_i^m \right)^2 \\
& + \beta \gamma \left\| \mathbf{b} - \frac{f(\mathbf{z}_1; \Theta^{(n)}) + f(\mathbf{z}_2; \Theta^{(n)})}{2} \right\|_2^2.
\end{aligned} \quad (15)$$

In the data fidelity term, the material image pixel values are coupled in the quadratic term. In order to separate the pixel values for simultaneous updating, the optimization transfer principle and an additive convexity trick (De Pierro 1993) are used here, as in the publication (Elbakri and Fessler 2002). The surrogate for the data fidelity term can be derived as:

$$\begin{aligned}
Q(\mathbf{b}) = & \frac{1}{\|\mathbf{A}\|_{\mathbf{w}}^2} \sum_{m=1}^M \sum_{i=1}^N \frac{Y_i^m}{2} \left(\sum_{j=1}^P \sum_{k=1}^K \gamma_{ijk}^m \left\{ \frac{a_{ij} \mu_k^m}{\gamma_{ijk}^m} ([\mathbf{b}_k]_j - [\mathbf{b}_k^{(n)}]_j) \right. \right. \\
& \left. \left. + \sum_{j=1}^P \sum_{k=1}^K a_{ij} \mu_k^m [\mathbf{b}_k^{(n)}]_j - \hat{l}_i^m \right\}^2 \right),
\end{aligned} \quad (16)$$

where a_{ij} is the element of system matrix \mathbf{A} , representing the attenuation contribution of j th pixel to the i th ray line, $\gamma_{ijk}^m = \frac{a_{ij} \mu_k^m}{\sum_{j=1}^P \sum_{k=1}^K a_{ij} \mu_k^m}$ is an intermediate variable for performing the additive convexity trick, $[\mathbf{b}_k^{(n)}]_j$ is the j th pixel of the k th basis material image in the n th iteration. By applying the optimization transfer principle, the updating formula for minimizing (15) can be derived as:

$$\begin{aligned}
[\mathbf{b}_k^{(n+1)}]_j = & [\mathbf{b}_k^{(n)}]_j \\
& - \frac{\sum_{m=1}^M \sum_{i=1}^N a_{ij} Y_i^m \mu_k^m \left\{ \sum_{j=1}^P a_{ij} \left(\sum_{k=1}^K \mu_k^m [\mathbf{b}_k^{(n)}]_j \right) - \hat{l}_i^m \right\} + \|\mathbf{A}\|_{\mathbf{w}}^2 \beta \gamma ([\mathbf{b}_k^{(n)}]_j - [f(\mathbf{z}_1; \Theta^{(n)}) + f(\mathbf{z}_2; \Theta^{(n)})]_{k,j} / 2)}{\sum_{m=1}^M \sum_{i=1}^N a_{ij} Y_i^m \mu_k^m \left(\sum_{k=1}^K \mu_k^m \right) \sum_{j=1}^P a_{ij} + \|\mathbf{A}\|_{\mathbf{w}}^2 \beta \gamma}.
\end{aligned} \quad (17)$$

The second subproblem is a network training problem, which is to minimize the last two terms of the whole loss function with respect to network parameters Θ . The loss function can be illustrated as:

$$g_2(\Theta; \mathbf{b}^{(n)}) = \gamma \left\| \mathbf{b}^{(n)} - \frac{f(\mathbf{z}_1; \Theta) + f(\mathbf{z}_2; \Theta)}{2} \right\|_2^2 + \frac{1}{2} (\|f(\mathbf{z}_1; \Theta) - \mathbf{z}_2\|_2^2 + \|f(\mathbf{z}_2; \Theta) - \mathbf{z}_1\|_2^2). \quad (18)$$

The network training subproblem can be monotonically optimized by using the limited memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm (Gong *et al* 2018). However, in practice, acceleration techniques and more computationally effective methods such as ADAM (Kingma and Ba 2014) can be applied, although they may not work in a rigorous monotonic way. In this work, to improve computational efficiency, we use ADAM for network training. The loss reduction is established empirically. Figure 1 shows the diagram of our

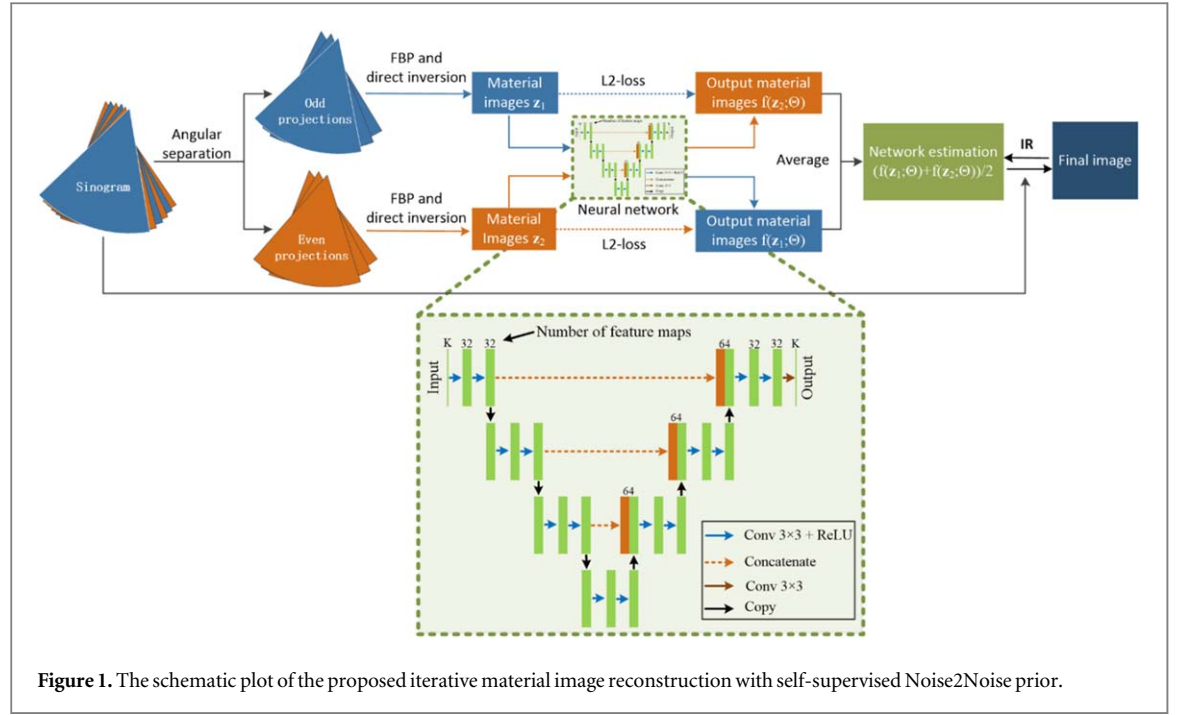


Table 1. The pseudo-code of our proposed material image reconstruction algorithm.

Algorithm 1. The Noise2Noise-prior-based iterative material image reconstruction algorithm

INPUT: Hyperparameters β, γ , total number of iterations N_{iter} ; system matrix \mathbf{A} , sinogram $\hat{\mathbf{I}}$, noise weighting matrix \mathbf{w} ; network structures $f(z; \Theta)$, network training iterations N_{sub} .

OUTPUT: Reconstructed material image \mathbf{b}^* .

1 Angularly split sinogram $\hat{\mathbf{I}}$, generate \mathbf{z}_1 and \mathbf{z}_2 by conducting FBP reconstruction and direct inversion;

2 Initialize $\mathbf{b}^{(0)}, \Theta^{(0)}$;

3 For n from 0 to $N_{iter} - 1$:

4 For material index k from 1 to K :

5 For pixel index j from 1 to P :

6 SQS updating: $[\mathbf{b}_k^{(n+1)}]_j = [\mathbf{b}_k^{(n)}]_j - \frac{\sum_{m=1}^M \sum_{i=1}^N a_{ij} \gamma_i^m \mu_k^m (\sum_{j=1}^P a_{ij} (\sum_{k=1}^K \mu_k^m [\mathbf{b}_k^{(n)}]_j) - \hat{I}_i^m) + \|\mathbf{A}\|_{\mathbf{w}}^2 \beta \gamma ([\mathbf{b}_k^{(n)}]_j - [f(\mathbf{z}_1; \Theta^{(n)}) + f(\mathbf{z}_2; \Theta^{(n)})]_{k,j} / 2)}{\sum_{m=1}^M \sum_{i=1}^N a_{ij} \gamma_i^m \mu_k^m (\sum_{k=1}^K \mu_k^m) \sum_{j=1}^P a_{ij} + \|\mathbf{A}\|_{\mathbf{w}}^2 \beta \gamma}$;

7 end for

8 end for

9 For network iteration index t from 1 to N_{sub} :

10 Network training via one iteration of Adam: $\Theta \leftarrow \text{Adam}(\Theta, \nabla g_2(\Theta; \mathbf{b}^{(n+1)}))$. $g_2(\cdot; \cdot)$ is given in (18);

11 end for

12 end for

13 $\mathbf{b}^* \leftarrow \mathbf{b}^{N_{iter}}$.

Noise2Noise-prior-based iterative material image reconstruction method. Table 1 shows the pseudo code of our proposed algorithm.

2.4. Experimental setting

We used a four-stage encoder-decoder network for Noise2Noise mapping. It had similar structures to U-net (Ronneberger et al 2015) but without resampling and doubled channels when moving to different stages. The input of the network is the material images with K channels, where K is the number of basis materials. The network was pretrained under the classic Noise2Noise framework, which was described by formula (10). In the pretraining process, no projection information was introduced and only the two sets of independently reconstructed noisy material images were utilized for mutual mapping. The number of network pretraining iterations was set as 600 and the learning rate was 10^{-3} . The network structure is shown in figure 1. The pretraining will help to provide a better initialization of prior images for the regularized reconstruction. For the reconstruction part, the total number of outer iterations N_{iter} was set as 200. The number of network training iterations N_{sub} was set to five. Parameter sweep experiments were conducted to optimize the two hyperparameters β and γ in (14) by using structural similarity (SSIM) as the evaluation indicator. The two hyperparameters β and γ were respectively

chosen as 1×10^{-5} and 1 for the simulation data after the parameter sweep experiment. The result of the parameter sweep experiment on the simulation data was shown in figure 11 in the discussion section. The hyperparameters β and γ were chosen as 7.5×10^{-6} and 5 for the preclinical photon-counting CT data and the clinical dual energy CT data. The network was optimized using the Adam optimizer with a learning rate of 10^{-3} . Tensorflow 1.1 was used for neural network implementation (Abadi *et al* 2016). A distance-driven projector and backprojectors implemented by CUDA on graphic processing units (GPU) Nvidia GTX 1080Ti were used for all the reconstructions. It should be noted that the proposed framework is an iterative reconstruction with network optimization during reconstruction, which is different from the existing unrolled framework, hence there is no end-to-end training. We transformed the whole problem into two subproblems. The first one was a regularized material image reconstruction problem solved using the separable quadratic surrogate (SQS) framework and the second one was a network training problem solved using Tensorflow. Thus, the CUDA-based reconstruction and the network training were accomplished separately. There is also no need to backpropagate the gradients through the projectors so we did not use Tensorflow-based projectors in this work. And the CUDA-based forward and backward projectors were not compiled together with Tensorflow.

2.5. Comparing methods

The basic method for comparison was FBP reconstruction with a subsequent direct matrix inversion. The second method for comparison was iterative material reconstruction with total variation (TV) prior (Sidky and Pan 2008, Xu and Noo 2020). It should be noted that the total variation was calculated in two dimensions on material images, not on the attenuation images. The loss function can be written as:

$$\begin{aligned} \Phi_{TV}(\mathbf{b}) = & \sum_{m=1}^M \sum_{i=1}^N \frac{Y_i^m}{2} \left(\left[\mathbf{A} \left(\sum_{k=1}^K \mu_k^m \mathbf{b}_k \right) \right]_i - \hat{l}_i^m \right)^2 \\ & + \beta_{TV} \sum_{k=1}^K \sum_{j=1}^N \sqrt{([\mathbf{b}_k]_{j+s} - [\mathbf{b}_k]_j)^2 + ([\mathbf{b}_k]_{j+1} - [\mathbf{b}_k]_j)^2 + \varepsilon}, \end{aligned} \quad (19)$$

where s represents the height of the image and ε is a small constant set as 10^{-8} . The separable quadratic surrogate (SQS) method was used to optimize the above loss function. SQS is a general framework to optimize the object function by constructing quadratic surrogates at each step (Elbakri and Fessler 2002). When constructing the surrogates for the TV-term, we used smoothed TV by adding a small constant ε to the original TV term to make it quadratic near the origin (Hou *et al* 2018). The number of iterations was set as 10000. Parameter sweep experiments were conducted to choose the weighting factor β_{TV} by using SSIM as the evaluation indicator. The result of the parameter sweep experiment on simulation data was shown in figure 11 in the discussion section. The hyperparameter β_{TV} was set to 4×10^{-4} for the simulation data and 2×10^{-5} for the preclinical photon-counting CT data and clinical dual-energy CT data.

3. Experiments

3.1. Numeric simulation study

To evaluate the performance of the proposed Noise2Noise-prior-based iterative material decomposition algorithm, we first used simulated phantom data for validation. The simulated spectral CT data is based on a thoracic phantom, which includes materials like adipose, bone, muscle, iodized blood, lung and air. Figure 2 shows different subregions defined in the phantom. Table 2 gives the densities of these submaterials.

The polychromatic projection generation process was conducted using a homemade toolkit. The spectrum we used for simulation was generated using SpekCalc software (Poludniowski *et al* 2009) with 0.8 mm beryllium, 1.5 mm aluminium and 0.5 mm copper filtration. A 2D fan-beam CT geometry was utilized in the simulation. The distance from the source to the centre of the rotation is 500 mm and the distance from the source to the detector is 1400 mm. A total of 900 views were acquired in an angular range of 360 degrees. The detector array has a length of 450 mm with 256 elements arranged equally distanced. The phantom is discretized as a 256×256 matrix in a $200 \text{ mm} \times 200 \text{ mm}$ area at the center of rotation. There are four energy bins in the simulation, with the threshold respectively set at 33, 58, 67, 81 and 120 keV. The basis materials we chose for decomposition are bone, adipose and iodine. To get realistic projection data, Poisson noise was introduced in the simulation process (Nuyts *et al* 2013).

3.2. Practical photon-counting CT study

Our second dataset for validation is a public photon counting CT dataset from MARS company (Aamir *et al* 2014), which can be found at <http://hdl.handle.net/10092/8531>. The dataset was a scan of fresh lamb chop using a small animal targeted spectral CT system. The system used a 2 mm thick CdTe sensor (128×128 ,

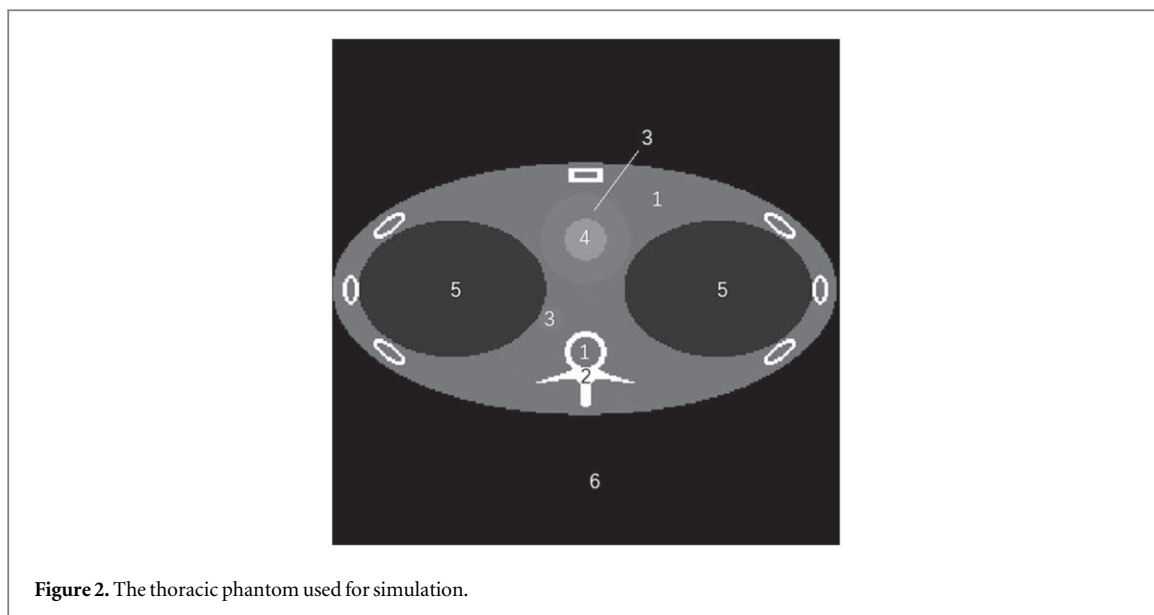


Figure 2. The thoracic phantom used for simulation.

Table 2. Material types of the phantom for simulation.

Number	Material types	Density
1	Adipose	1.0012 g cm^{-3}
2	Bone	1.5040 g cm^{-3}
3	Muscle	1.0644 g cm^{-3}
4	Iodized blood (95%Blood + 5% Iodine)	1.0968 g cm^{-3}
5	Lung tissue	0.2608 g cm^{-3}
6	Air	0.0013 g cm^{-3}

110 μm detector pitch), which was bonded to a Medipix3RX readout chip. The system used a Source-Ray SB-80-1K x-ray tube (Source-Ray Inc., Ronkonkoma, NY, USA), which had a tungsten anode and was equipped with a 1.8 mm Al equivalent intrinsic filter. The focal spot size was about 33 μm . The distance between source and detector was 131.8 mm and the distance between object and detector was 48 mm. Before the experiment, the energy calibration and pixel response variation correction were performed for the detector. The x-ray tube was operated at 50 kVp with a current of 120 μA , with four low energy thresholds set at 15, 20, 25, and 30 keV. The exposure time of each acquisition was 40 ms. Ten vertical translational positions of the detector were enough to create a virtual detector that covers a 23 mm field of view (FOV). After the oversampling correction, the raw data was packed into a tensor of 651 (views) \times 374 (detectors) \times 4 (energy bins). The basis materials we chose for material decomposition are bone and adipose.

3.3. Clinical dual-energy CT study

The third dataset is a clinical DECT dataset, which was acquired from patients who underwent CT on a dual-source, 64-detector-row multidetector CT (Siemens Somatom Definition Flash). The dual energy x-ray source was operated at 80 kVp/140 kVp. The scanner has 64 detector rows and two types of detector channels in which the numbers of channels are 736 (140 kVp) and 480 (80 kVp). The number of views is 1152 and the size of the reconstructed image is 512×512 . The basis materials we used for decomposition are soft tissue and iodine. The study was approved by the Institutional Review Board at Massachusetts General Hospital under IRB number 2019P000802.

4. Results

4.1. Numerical simulation results

Figure 3 shows the reconstruction results of different methods on a simulated spectral CT dataset and figure 4 shows the corresponding zoom-ins. According to the definition of the material decomposition coefficients in (2), the material coefficient image represents the normalized density of the corresponding basis material. So the display window is unitless. We can figure out that the direct inversion method gives very noisy output, especially when looking at the adipose image. The TV prior method is advantageous in numerical simulation study since

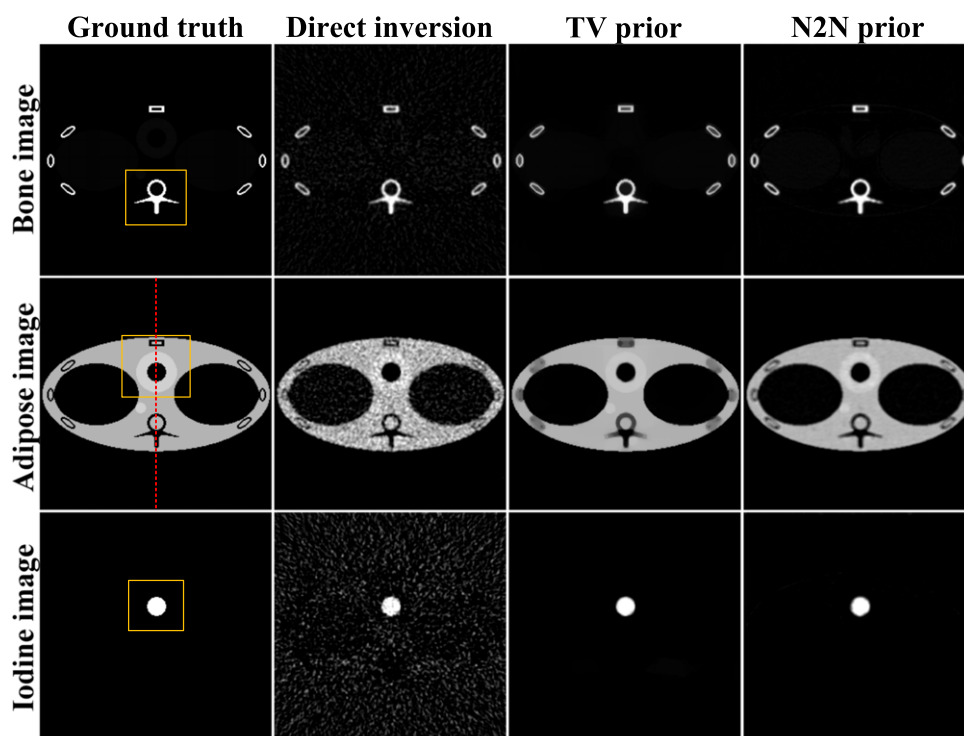


Figure 3. Simulation results. The window for bone and iodine image is $[0, 1]$. The window for adipose is $[0.3, 1.3]$.

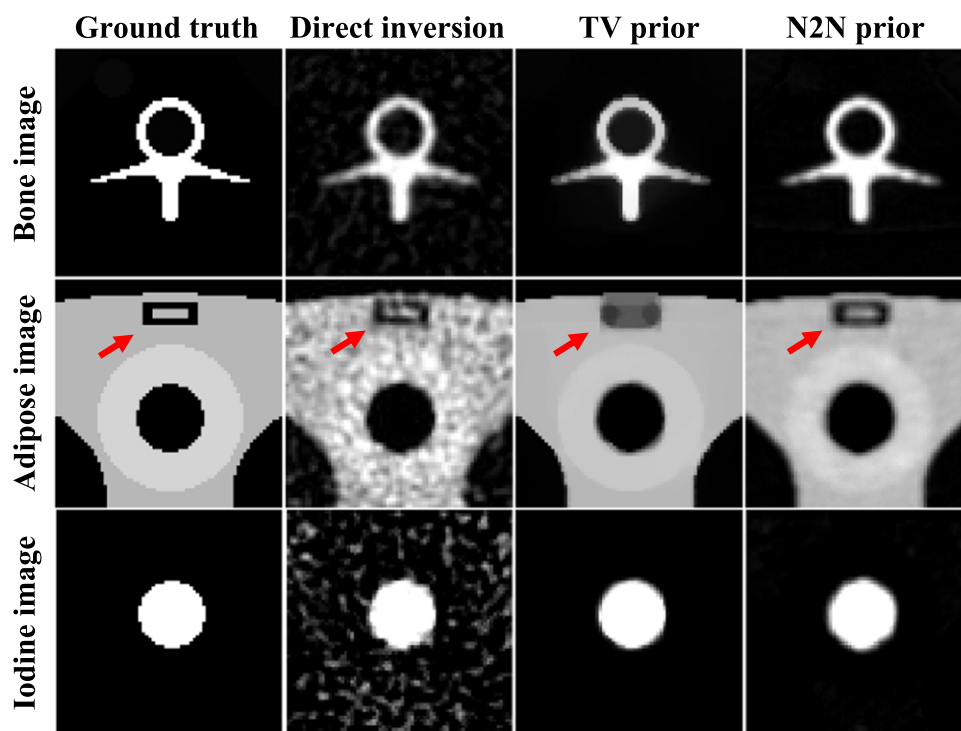


Figure 4. Zoom-in images of simulation results (marked with yellow rectangle in figure 3).

there is no texture in the phantom. For the adipose image in the TV prior results, the noise level is significantly reduced in regions where pixel values are uniform. But for small structures where high contrast exists, the TV prior method tends to over-smooth the image and incurs blur (red arrow in figure 4). The results show that the Noise2Noise (N2N) prior method can effectively suppress the noise and at the same time preserve the textural details in the image, indicating a well-balanced trade-off.

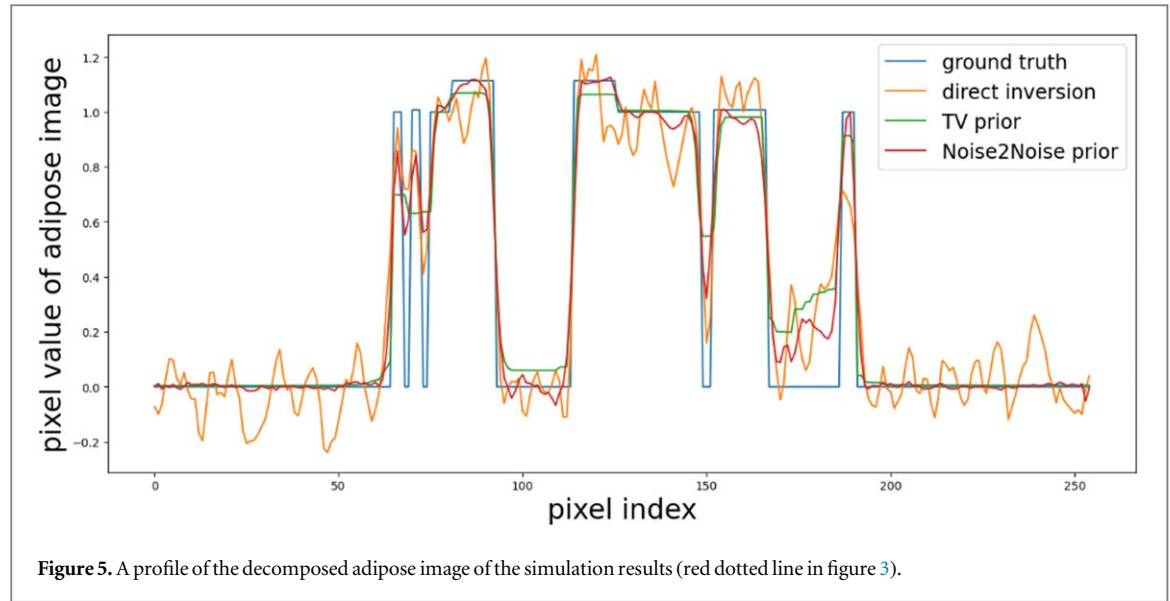


Figure 5. A profile of the decomposed adipose image of the simulation results (red dotted line in figure 3).

As illustrated in figure 3, a profile is drawn vertically in the middle of the adipose image. We can see in figure 5 that the profile of the direct inversion result has the largest fluctuation, indicating strong noise in the image. The profile of the TV prior-based method has the smallest fluctuation of the three methods, suggesting lowest noise. However, the TV prior method does not recover the original signal well in areas where pixel values change very intensely. Besides, the TV prior method tends to have bias in some platform area when compared to the ground truth. Judging from the profile, the Noise2Noise prior method achieves a good balance between noise suppression and bias control.

Quantitative indicators were calculated for result evaluation, including bias, standard deviation (STD), root of mean square error (RMSE) and structural similarity (SSIM). The bias of each result was calculated by comparing with the ground truth image per material. RMSE measures the L2-norm error between the estimated image and the ground truth image, which is calculated as:

$$\text{RMSE}(\hat{\mathbf{b}}_k, \mathbf{b}_k) = \|\hat{\mathbf{b}}_k - \mathbf{b}_k\|_2, \quad (20)$$

where $\hat{\mathbf{b}}_k$ is the estimated k th material image and \mathbf{b}_k is the corresponding ground truth image. The SSIM index measures the structural similarity by comparing both the mean value and distribution relevance between estimated and reference images. The SSIM was also calculated per material. The calculation of SSIM can be illustrated as:

$$\text{SSIM}(\hat{\mathbf{b}}_k, \mathbf{b}_k) = \frac{(2\overline{\hat{\mathbf{b}}_k} \overline{\mathbf{b}_k} + C_1)(2\sigma_{\hat{\mathbf{b}}_k \mathbf{b}_k} + C_2)}{(\overline{\hat{\mathbf{b}}_k}^2 + \overline{\mathbf{b}_k}^2 + C_1)(\sigma_{\hat{\mathbf{b}}_k}^2 + \sigma_{\mathbf{b}_k}^2 + C_2)}, \quad (21)$$

where $\hat{\mathbf{b}}_k$ is the estimated material image and \mathbf{b}_k is the reference image. $\overline{\hat{\mathbf{b}}_k}$ and $\overline{\mathbf{b}_k}$ are respectively mean values of $\hat{\mathbf{b}}_k$ and \mathbf{b}_k . $\sigma_{\hat{\mathbf{b}}_k \mathbf{b}_k}$ is the covariance of $\hat{\mathbf{b}}_k$ and \mathbf{b}_k . $\sigma_{\hat{\mathbf{b}}_k}^2$ is the variance of $\hat{\mathbf{b}}_k$ and $\sigma_{\mathbf{b}_k}^2$ is the variance of \mathbf{b}_k . C_1 and C_2 are two constants to avoid a zero denominator.

Figure 6 shows the decay curve of the Noise2Noise loss in the pretraining and the whole loss in the iterative material reconstruction. It implies that with the Adam optimizer the convergence and the monotonic decrease of the cost function could be empirically achieved. Table 3 shows the quantitative evaluation indicators we calculated in the numerical phantom study. We can conclude that the proposed Noise2Noise (N2N) prior method outperforms other methods on most indicators. The Noise2Noise prior method can significantly suppress the noise compared to the direct inversion method and had better detail and texture preservation compared to the TV prior method. The result shows that the TV prior method had better performance than the Noise2Noise prior method on the iodine image. It is understandable since TV prior is good at processing this kind of standard uniform structure with no textures.

4.2. Practical photon-counting CT results

Figure 7 shows the reconstruction results of different methods on the practical photon-counting CT dataset. We can tell that the direct inversion method still gave very noisy results. Specifically, in the adipose image, we can see that the small structures in the image (the red rectangle in figure 7) were heavily contaminated by the strong noise. The region marked with a red rectangle was also enlarged at the bottom of each material image. It is worth noting that the results given by the TV prior method tended to have checkboard artefacts when it comes to

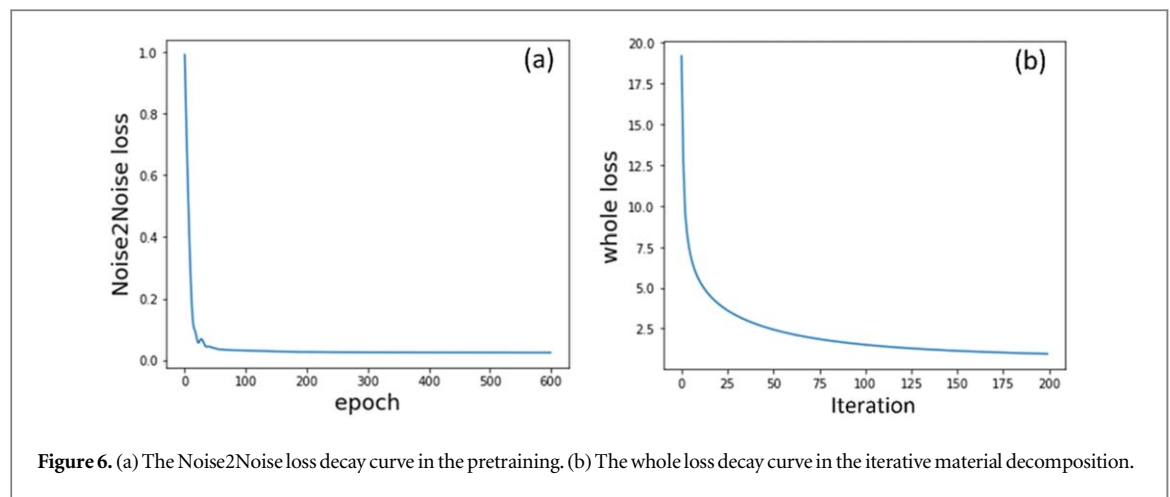


Figure 6. (a) The Noise2Noise loss decay curve in the pretraining. (b) The whole loss decay curve in the iterative material decomposition.

Table 3. Quantitative analysis of simulation results.

		Direct inversion	TV prior	N2N prior
Bone	BIAS	0.0428	0.1332	0.0103
	STD	0.0928	0.1055	0.0620
	RMSE	0.1022	0.1699	0.0628
	SSIM	0.5754	0.6305	0.7835
Adipose	BIAS	0.0031	0.0131	0.0069
	STD	0.0989	0.0200	0.0159
	RMSE	0.0990	0.0242	0.0175
	SSIM	0.4195	0.9445	0.9584
Iodine	BIAS	0.0049	0.0817	0.0116
	STD	0.1628	0.0101	0.0303
	RMSE	0.1629	0.0824	0.0325
	SSIM	0.1664	0.9764	0.9556

practical scan images. The proposed Noise2Noise prior method can effectively suppress the noise and recover the small image structure. Table 4 is the quantitative analysis of the practical photon-counting CT results. Since we did not have the ground truth image for this practical photon counting CT dataset, we assumed the result of the direct inversion method was unbiased. The calculation of bias was done by comparing to the mean value of the direct inversion result per material. The region of interest (ROI) we chose for quantitative analysis is the green rectangle in figure 7. It is understandable that TV prior can obtain a slightly lower standard deviation than the Noise2Noise prior method in the soft tissue result since the TV prior method aims to minimize the variation between pixels. For other indicators, the Noise2Noise prior method performed better than the TV prior method, indicating good performance on bias controlling and noise suppression. Figure 8 displays the profile which is drawn with a blue dashed line in figure 7. We can figure out that, compared to the TV prior method, the proposed Noise2Noise prior method can restore the signals in the edge area well and at the same time effectively suppress the noise in the flat area.

4.3. Clinical dual energy CT results

Figure 9 illustrates the decomposed results of the clinical dual energy CT dataset. We can see that the decomposed adipose image and iodine image, by using the direct inversion method, were extremely noisy, which impaired the visualization of small intrahepatic structures. Although the TV prior method suppressed the noise, it over-smoothed the image, generated blocky artifacts, and impaired the visibility of internal structures (such as the blood vessels in the liver). The Noise2Noise prior method suppressed image noise while improving the visibility of intrahepatic structures. Figure 10 displays a profile which was denoted with a green dashed line in figure 9. It shows that the TV prior method failed to suppress the noise at the far-left part of the profile while the Noise2Noise prior method handled it well. Compared to the profile of the TV prior method, the profile of the Noise2Noise prior method was less flat, which indicates less over-smoothness, especially in the liver area (the middle flat area). The red rectangle marks the region of interest (ROI) we enlarged at the bottom. Table 5 shows the bias and standard deviation calculated from the region. Similar to the bias calculation in practical photon

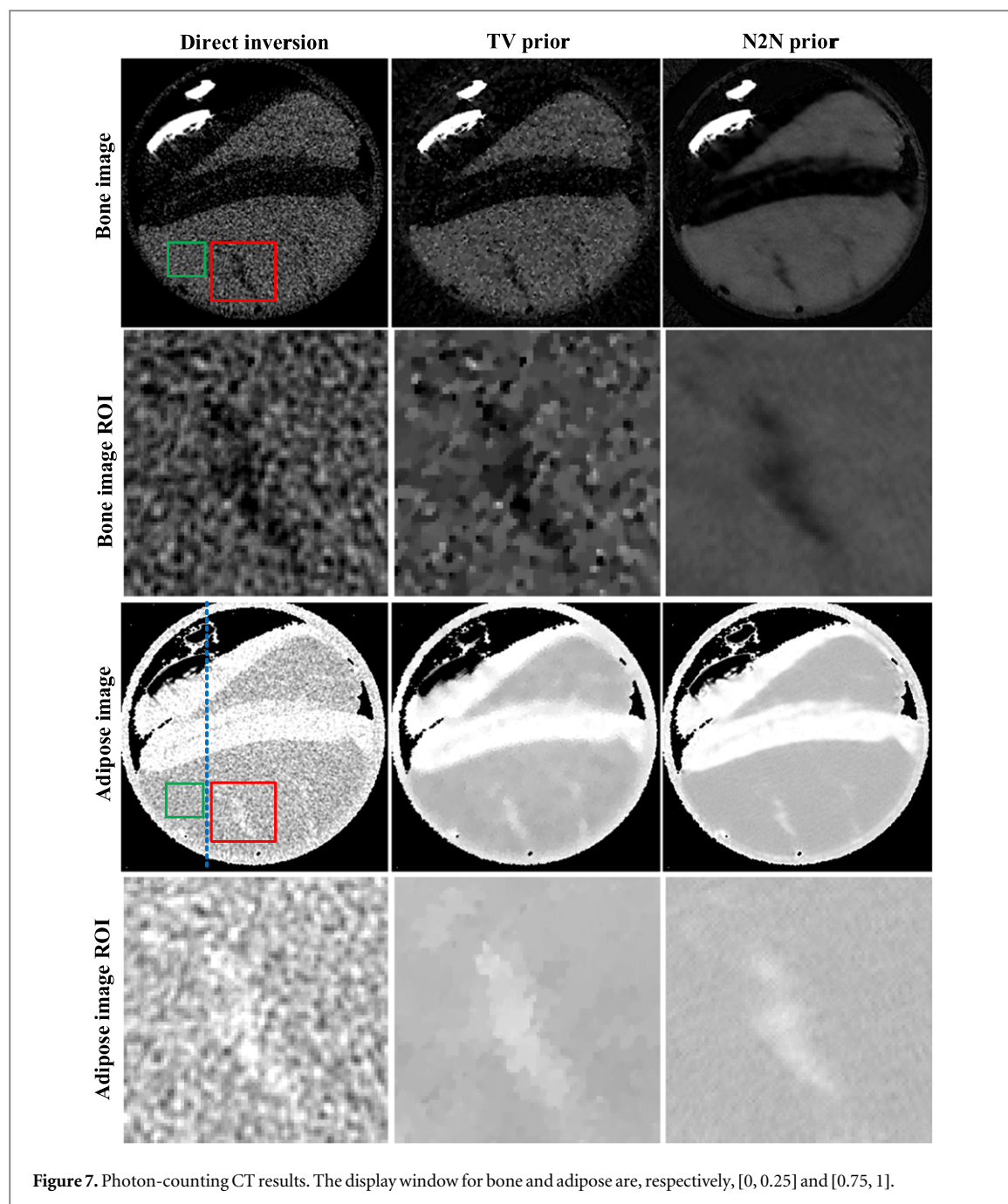


Table 4. Quantitative analysis of practical photon-counting CT results.

		Direct inversion	TV prior	N2N prior
Bone	BIAS	0	0.0049	0.0026
	STD	0.02592	0.0168	0.0033
Soft tissue	BIAS	0	0.0024	0.0011
	STD	0.02591	0.0034	0.0036

counting CT results. We assumed the result of direct inversion had no bias. The bias was calculated by comparing to the mean of the direct inversion result. It shows that the Noise2Noise prior based method can perform better than the TV prior method on bias controlling. It is understandable that for iodine image, the standard deviation achieved by Noise2Noise prior method is slightly larger than the TV prior method since the target of the TV prior method is to minimize the variation between pixels.

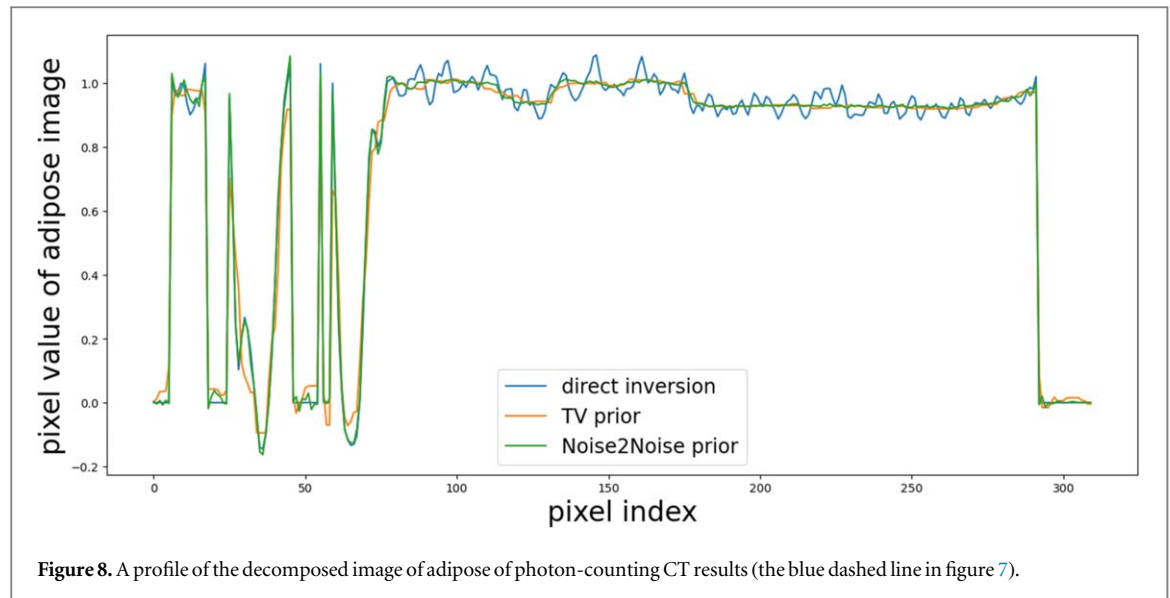


Figure 8. A profile of the decomposed image of adipose of photon-counting CT results (the blue dashed line in figure 7).

5. Discussion

Figure 11 shows the hyperparameter sweep experiment results of the Noise2Noise prior method and the TV prior method on numerical simulation data. The hyperparameters were chosen by using structural similarity (SSIM) as the evaluation indicator. The two hyperparameters β and γ for the Noise2Noise prior method on numerical simulation data were respectively set to 1×10^{-5} and 1. The regularization term balance factor β_{TV} for the numerical simulation data was set as 4×10^{-4} . The hyperparameters for the practical photon-counting CT dataset and dual energy clinical CT dataset were chosen in the same way.

The network size is critical for Noise2Noise denoising performance. A small network will give blurry outputs because the network fails to express the structure detail of the image. While a large network will restore the noise at the same time due to its strong representation power. In our work, we found that a four-depth encoder-decoder network with 32 filter channels was suitable for our material image denoising task. Besides, the computation time consumed on in-iteration network training in reconstruction part was negligible since only one pair of reconstruction image was used for network training. So the computation time of the proposed method was just typical iterative reconstruction time.

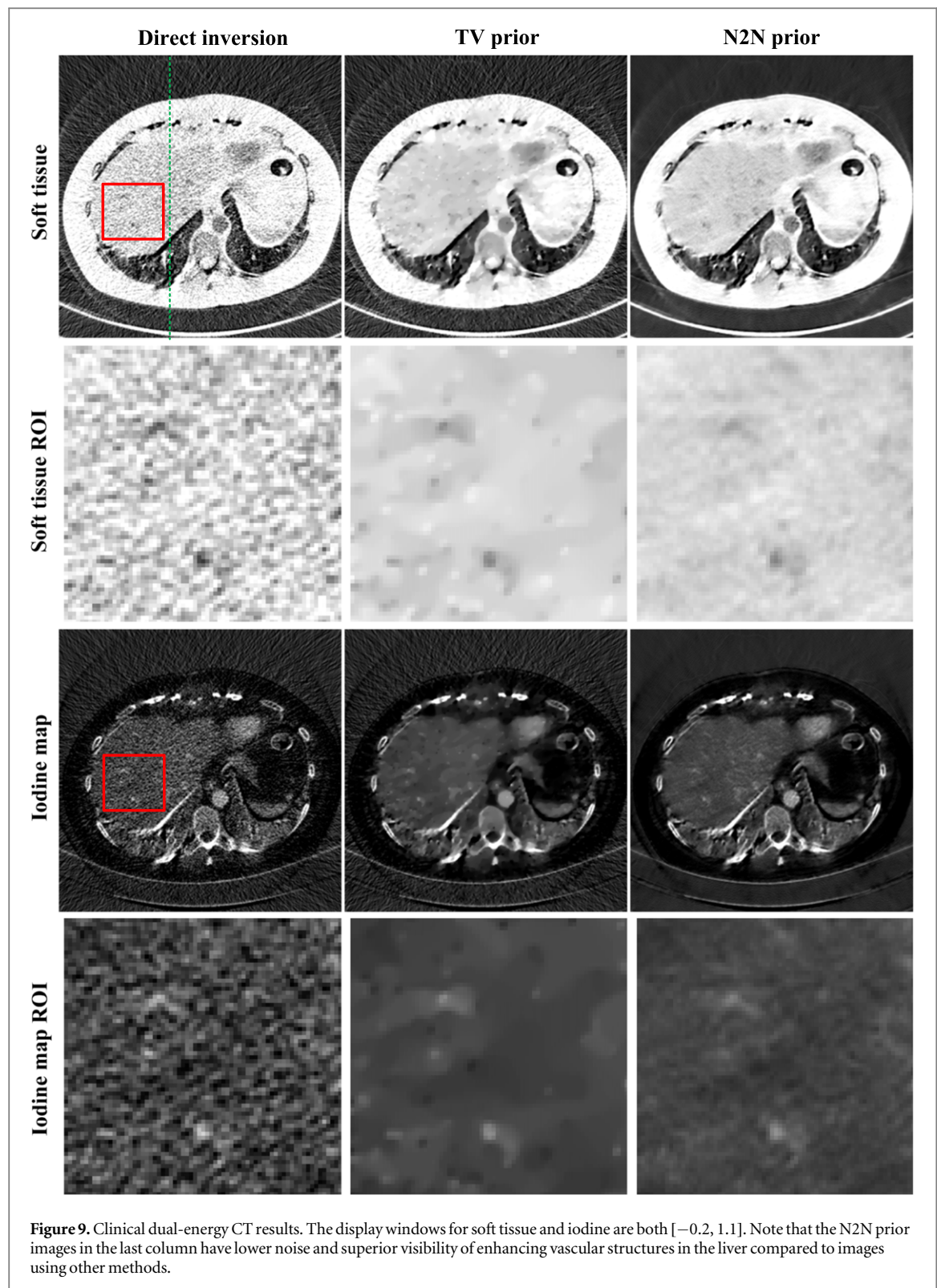
We admit that 3D reconstruction is a common problem for DL-prior-based image reconstruction. To reduce the time consumption in 3D reconstruction, possible solutions may be: (1) use a pretrained network to reduce the overall iterations needed for reconstruction; (2) when optimizing the network in each iteration, instead of using the whole image to perform back-propagation, it is possible to use only patches, which could reduce both computational time and memory consumption (Wu et al 2019b).

Compared to supervised learning, the proposed Noise2Noise reconstruction mainly has two benefits. The first one is that it will adaptively adjust the network during the reconstruction for individual images. This is mainly thanks to the existence of the supervised target, which are the two separately reconstructed material images using FBP and a subsequent direct inversion. Such adaptive network optimization is hard to achieve using a trained network without the existence of a self-supervising target. The second one is that the proposed method does not need any external data for the network training. This would be extremely helpful when handling experimental data such as the photon-counting CT data used in our work. For supervised methods, each application needs the corresponding training dataset and noise levels. Certainly, it is very meaningful if we can combine a supervised learning with the proposed Noise2Noise approach, like initializing our network with a pretrained supervised network. This may lead to a faster convergence and better results.

The data fidelity term can supply some degree of self-supervision to the images. However, it should be noted that the self-supervision that the projection data could provide is limited. If one only relies on the projection data without any other prior knowledge, the image would converge to the noisy reconstruction. Hence, almost all the iterative reconstruction algorithms need both the data fidelity and the prior term. Assuming that we only use a data fitting term as the learning objective, a possible object function is a deep image prior (Gong et al 2018):

$$\Theta^* = \arg \min_{\Theta} \|Af(\mathbf{z}; \Theta) - \mathbf{p}\|_{\mathbf{w}}^2, \quad (22)$$

where $f(\mathbf{z}; \Theta)$ is the network to be trained and \mathbf{z} is a prior image. \mathbf{p} represents the projection data and \mathbf{w} is the weighting matrix. The final reconstructed PET image is the output of the network $f(\mathbf{z}; \Theta^*)$. However, in the



work by Gong *et al*, high quality CT images were used as the prior image \mathbf{z} and $f(\mathbf{z}; \Theta)$ is the PET image. In spectral CT, we do not have the prior image that has a much higher quality than the image to be reconstructed.

Furthermore, equation (22) suffers from overfitting problem. If one iterates too much, $f(\mathbf{z}; \Theta)$ would converge to the noisy solution that $\mathbf{x} = \arg \min \|\mathbf{Ax} - \mathbf{p}\|_{\mathbf{w}}^2$. The overfitting problem is less significant in Gong's work since \mathbf{z} and $f(\mathbf{z}; \Theta)$ are in different domains (CT and PET), and overfitting is harder. But it is also needed to control the number of iterations to avoid the overfitting. In the spectral CT case, overfitting will be much easier since we can only take CT images as \mathbf{z} . The Noise2Noise prior can largely prevent the aforementioned overfitting issue since it is much harder for the independent noises in the two realizations to overfit to each other.

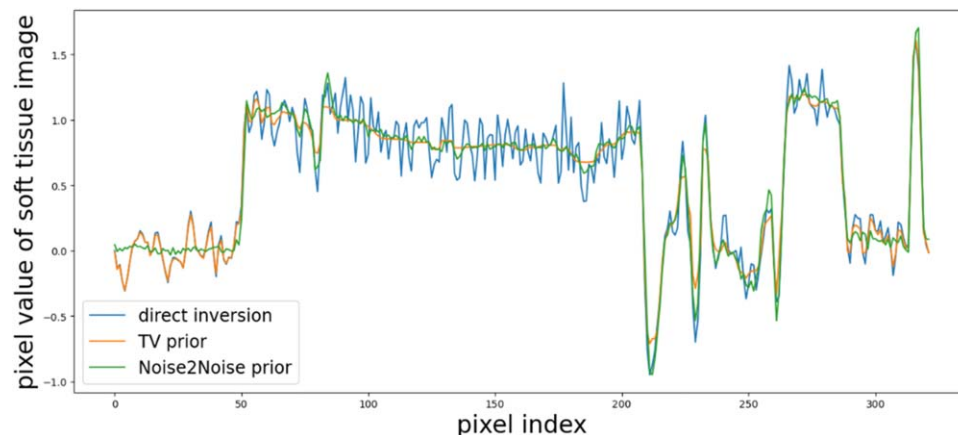


Figure 10. A profile of the decomposed image of soft tissue from clinical CT results (the green dashed line in figure 9).

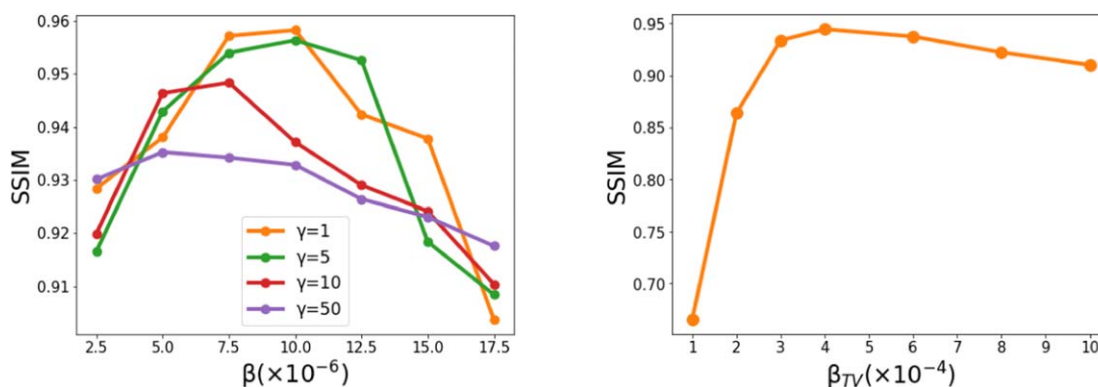


Figure 11. The hyperparameters sweep experiment result of (a) the Noise2Noise prior method and (b) the TV prior method.

Table 5. Quantitative analysis of clinical dual-energy CT results.

Material image	Indicators	Direct inverse	TV prior	N2N prior
Soft tissue	BIAS	0	0.0012	0.0005
	STD	0.1894	0.0578	0.0568
Iodine	BIAS	0	0.0062	0.0031
	STD	0.1588	0.0468	0.0511

In our work, the information from the projection data and Noise2Noise prior are both needed to achieve high quality reconstructed material images because they each play different roles. The supervision of projection data helps control the bias and avoid the over-smoothing of the reconstructed material images, while the Noise2Noise prior plays a denoising role.

Comprehensive studies on DECT images processed with the Noise2Noise prior method are needed to assess if the method can reduce the radiation dose for DECT and/or improve image quality in patients with larger body habitus, or with arms by the side, which tend to have higher image noise. Likewise, future studies should also assess the impact of our method on the reliability of CT attenuation numbers and the reduction of beam hardening artefacts at the shoulder or in larger patients.

6. Conclusion

In this paper, we proposed an iterative material image reconstruction algorithm based on self-supervised Noise2Noise penalty. Instead of using Noise2Noise as a post-processing tool, we used Noise2Noise as a prior

function for material image denoising in an iterative material image reconstruction framework. In the proposed algorithm, the data consistency term and the Noise2Noise network were alternatively optimized in the iterative framework, respectively, using separable the quadratic surrogate (SQS) method and the Adam algorithm. The advantage of our method is that the deep Noise2Noise prior was built in a self-supervised manner with no need for external training data. Besides, since the data fidelity term was incorporated in the loss function, the estimated image had small bias. The proposed method was compared with direct inversion and the TV-prior-based iterative reconstruction method on a simulated spectral CT dataset, a preclinical photon-counting CT dataset and a clinical dual-energy CT dataset. The comparative study showed the effectiveness of our proposed method on noise suppression and minor structure preservation.

Acknowledgments

This work was supported by the grants from NSFC 11775124 and National Key R&D Program of China (2018YFC0115502).

ORCID iDs

Dufan Wu  <https://orcid.org/0000-0002-3204-3502>

References

- Aamir R, Chernoglazov A, Bateman C J, Butler A P, Butler P H, Anderson N G, Bell S, Panta R, Healy J and Mohr J 2014 MARS spectral molecular imaging of lamb tissue: data collection and image analysis *J. Instrum.* **9** P02005
- Abadi M, Agarwal A, Barham P, Brevdo E, Chen Z, Citro C, Corrado G S, Davis A, Dean J and Devin M 2016 Tensorflow: large-scale machine learning on heterogeneous distributed systems arXiv:1603.04467
- Adler J and Öktem O 2018 Learned primal-dual reconstruction *IEEE Trans. Med. Imaging* **37** 1322–32
- Alvarez R E and Macovski A 1976 Energy-selective reconstructions in x-ray computerised tomography *Phys. Med. Biol.* **21** 733–44
- Barber R F, Sidky E Y, Schmidt T G and Pan X 2016 An algorithm for constrained one-step inversion of spectral CT data *Phys. Med. Biol.* **61** 3784–818
- Cai C, Rodet T, Legoupil S and Mohammad-Djafari A 2013 A full-spectral Bayesian reconstruction approach based on the material decomposition model applied in dual-energy computed tomography *Med. Phys.* **40** 111916
- Chandarana H, Megibow A J, Cohen B A, Srinivasan R, Kim D, Leidecker C and Macari M 2011 Iodine quantification with dual-energy CT: phantom study and preliminary experience with renal masses *Am. J. Roentgenol.* **196** W693–700
- Chen H, Zhang Y, Kalra M K, Lin F, Chen Y, Liao P, Zhou J and Wang G 2017 Low-dose CT with a residual encoder-decoder convolutional neural network *IEEE Trans. Med. Imaging* **36** 2524–35
- Chen Z and Li L 2019 Robust multimaterial decomposition of spectral CT using convolutional neural networks *Opt. Eng.* **58** 013104
- Clark D P, Holbrook M and Badea C T 2018 Multi-energy CT decomposition using convolutional neural networks *Proc. SPIE* **10573** 415–23
- De Pierro A R 1993 On the relation between the ISRA and the EM algorithm for positron emission tomography *IEEE Trans. Med. Imaging* **12** 328–33
- Digumarthy S R, Singh R, Rastogi S, Otrakji A, Homayounieh F, Zhang E W, McDermott S and Kalra M K 2020 Low contrast volume dual-energy CT of the chest: quantitative and qualitative assessment *Clin. Imaging* **69** 305–10
- Elbakri I A and Fessler J A 2002 Statistical image reconstruction for polyenergetic x-ray computed tomography *IEEE Trans. Med. Imaging* **21** 89–99
- Erdogan H and Fessler J A 1999 Monotonic algorithms for transmission tomography *IEEE Trans. Med. Imaging* **18** 801–14
- Fang W, Li L and Chen Z 2020 Removing ring artefacts for photon-counting detectors using neural networks in different domains *IEEE Access* **8** 42447–57
- Gong K, Catana C, Qi J and Li Q 2018 PET image reconstruction using deep image prior *IEEE Trans. Med. Imaging* **38** 1655–65
- Harms J, Wang T, Petrongolo M, Niu T and Zhu L 2016 Noise suppression for dual-energy CT via penalized weighted least-square optimization with similarity-based regularization *Med. Phys.* **43** 2676–86
- Hou X, Teng Y, Kang Y and Qi S 2018 A separable quadratic surrogate total variation minimization algorithm for accelerating accurate CT reconstruction from few-views and limited-angle data *Med. Phys.* **45** 535–48
- Jin K H, McCann M T, Froustey E and Unser M 2017 Deep convolutional neural network for inverse problems in imaging *IEEE Trans. Image Process.* **26** 4509–22
- Kang E, Min J and Ye J C 2017 A deep convolutional neural network using directional wavelets for low-dose x-ray CT reconstruction *Med. Phys.* **44** e360–75
- Kingma D P and Ba J 2014 Adam: a method for stochastic optimization arXiv:1412.6980
- Krizhevsky A, Sutskever I and Hinton G E 2012 Imagenet classification with deep convolutional neural networks *Adv. Neural Inf. Process. Syst.* pp 1097–105
- Le H Q and Molloy S 2011 Least squares parameter estimation methods for material decomposition with energy discriminating detectors *Med. Phys.* **38** 245–55
- Lehmann L, Alvarez R, Macovski A, Brody W, Pelc N, Riederer S J and Hall A 1981 Generalized image combinations in dual KVP digital radiography *Med. Phys.* **8** 659–67
- Lehtinen J, Munkberg J, Hasselgren J, Laine S, Karras T, Aittala M and Aila T 2018 Noise2noise: learning image restoration without clean data arXiv:1803.04189
- Li L, Chen Z, Cong W and Wang G 2014 Spectral CT modeling and reconstruction with hybrid detectors in dynamic-threshold-based counting and integrating modes *IEEE Trans. Med. Imaging* **34** 716–28

- Li L, Zhao T and Chen Z 2018 First Dual MeV energy x-ray CT for container inspection: design, algorithm, and preliminary experimental results *IEEE Access* **6** 45534–42
- Long Y and Fessler J A 2014 Multi-material decomposition using statistical image reconstruction for spectral CT *IEEE Trans. Med. Imaging* **33** 1614–26
- Maaß C, Baer M and Kachelrieß M 2009 Image-based dual energy CT using optimized precorrection functions: a practical new approach of material decomposition in image domain *Med. Phys.* **36** 3818–29
- McCollough C H, Leng S, Yu L and Fletcher J G 2015 Dual-and multi-energy CT: principles, technical approaches, and clinical applications *Radiology* **276** 637–53
- Mendonça P R, Lamb P and Sahani D V 2013 A flexible method for multi-material decomposition of dual-energy CT images *IEEE Trans. Med. Imaging* **33** 99–116
- Mory C, Sixou B, Si-Mohamed S, Boussel L and Rit S 2018 Comparison of five one-step reconstruction algorithms for spectral CT *Phys. Med. Biol.* **63** 235001
- Niu T, Dong X, Petrongolo M and Zhu L 2014 Iterative image-domain decomposition for dual-energy CT *Med. Phys.* **41** 041901
- Nuyts J, De Man B, Fessler J A, Zbijewski W and Beekman F J J 2013 Modelling the physics in the iterative reconstruction for transmission computed tomography *Phys. Med. Biol.* **58** R63–96
- Otrakji A, Digumarthy S R, Lo Gullo R, Flores E J, Shepard J-A O and Kalra M K 2016 Dual-energy CT: spectrum of thoracic abnormalities *Radiographics* **36** 38–52
- Poludniowski G, Landry G, Deblois F, Evans P and Verhaegen F 2009 SpekCalc: a program to calculate photon spectra from tungsten anode x-ray tubes *Phys. Med. Biol.* **54** N433–8
- Pourmorteza A, Symons R, Sandfort V, Mallek M, Fuld M K, Henderson G, Jones E C, Malayeri A A, Folio L R and Bluemke D A 2016 Abdominal imaging with contrast-enhanced photon-counting CT: first human experience *Radiology* **279** 239–45
- Primak A N, Fletcher J G, Vrtiska T J, Dzyubak O P, Lieske J C, Jackson M E, Williams J C Jr and McCollough C H 2007 Noninvasive differentiation of uric acid versus non-uric acid kidney stones using dual-energy CT *Acad. Radiol.* **14** 1441–7
- Rajendran K, Löbker C, Schon B S, Bateman C J, Younis R A, de Ruiter N J, Chernoglazov A I, Ramyar M, Hooper G J and Butler A P 2017 Quantitative imaging of excised osteoarthritic cartilage using spectral CT *Eur. Radiol.* **27** 384–92
- Roessl E and Proksa R 2007 K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors *Phys. Med. Biol.* **52** 4679–96
- Ronneberger O, Fischer P and Brox T 2015 *International Conference on Medical Image Computing and Computer-assisted Intervention* pp 234–41
- Schlomka J, Roessl E, Dorscheid R, Dill S, Martens G, Istel T, Bäumer C, Herrmann C, Steadman R and Zeitler G 2008 Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography *Phys. Med. Biol.* **53** 4031–47
- Sidky E Y and Pan X 2008 Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization *Phys. Med. Biol.* **53** 4777–807
- Thieme S F, Becker C R, Hacker M, Nikolaou K, Reiser M F and Johnson T R 2008 Dual energy CT for the assessment of lung perfusion-correlation to scintigraphy *Eur. J. Radiol.* **68** 369–74
- Wu D, Gong K, Kim K, Li X and Li Q 2019a *International Conference on Medical Image Computing and Computer-Assisted Intervention* pp 741–9
- Wu D, Kim K and Li Q 2019b Computationally efficient deep neural network for computed tomography image reconstruction *Med. Phys.* **46** 4763–76
- Wu D, Gong K, Kim K, Zhang X, Ouyang J and Li Q 2020a Deep denoising of O-15 water dynamic PET images without training data *J. Nucl. Med.* **61** 433
- Wu D, Ren H and Li Q 2020b Self-supervised dynamic CT perfusion image denoising with deep neural networks *IEEE Trans. Radiation Plasma Med. Sci.* **5** 350–61
- Wu D, Kim K, El Fakhri G and Li Q 2017 Iterative low-dose CT reconstruction with priors trained by artificial neural network *IEEE Trans. Med. Imaging* **36** 2479–86
- Xu J and Noo F 2020 A robust regularizer for multiphase CT *IEEE Trans. Med. Imaging* **39** 2327–38
- Xu Y, Yan B, Zhang J, Chen J, Zeng L and Wang L 2018 Image decomposition algorithm for dual-energy computed tomography via fully convolutional network *Comput. Math. Methods Med.* **2018** 2527516
- Xue Y, Ruan R, Hu X, Kuang Y, Wang J, Long Y and Niu T 2017 Statistical image-domain multimaterial decomposition for dual-energy CT *Med. Phys.* **44** 886–901
- Yao Y, Li L and Chen Z 2019 Dynamic-dual-energy spectral CT for improving multi-material decomposition in image-domain *Phys. Med. Biol.* **64** 135006
- Yu L, Leng S and McCollough C H 2012 Dual-energy CT–based monochromatic imaging *Am. J. Roentgenol.* **199** S9–15
- Yuan N, Zhou J and Qi J 2020 Half2Half: deep neural network based CT image denoising without independent reference data *Phys. Med. Biol.* **65** 215020
- Zhang L-J, Peng J, Wu S-Y, Wang Z J, Wu X-S, Zhou C-S, Ji X-M and Lu G-M 2010 Liver virtual non-enhanced CT with dual-source, dual-energy CT: a preliminary study *Euro. Radiol.* **20** 2257–64
- Zhang R, Thibault J-B, Bouman C A, Sauer K D and Hsieh J 2013 Model-based iterative reconstruction for dual-energy x-ray CT using a joint quadratic likelihood model *IEEE Trans. Med. Imaging* **33** 117–34
- Zhang W, Zhang H, Wang L, Wang X, Hu X, Cai A, Li L, Niu T and Yan B 2019 Image domain dual material decomposition for dual-energy CT using butterfly network *Med. Phys.* **46** 2037–51
- Zhao W, Niu T, Xing L, Xie Y, Xiong G, Elmore K, Zhu J, Wang L and Min J K 2016 Using edge-preserving algorithm with non-local mean for significantly improved image-domain material decomposition in dual-energy CT *Phys. Med. Biol.* **61** 1332–51
- Zhao Y, Zhao X and Zhang P 2014 An extended algebraic reconstruction technique (E-ART) for dual spectral CT *IEEE Trans. Med. Imaging* **34** 761–8
- Zimmerman K C, Sharma G, Parchur A K, Joshi A and Schmidt T G 2020 Experimental investigation of neural network estimator and transfer learning techniques for K-edge spectral CT imaging *Med. Phys.* **47** 541–51