Dual-domain joint learning reconstruction method (JLRM) combined with physical process for spectral computed tomography

Genwei Ma[1,3] and Xing Zhao[2,3*]

[1]Academy for multidisciplinary studies, Captial Normal University, West third ring road , Beijing, 100048, Beijing, China.
[2]School of Mathematical Sciences, Captial Normal University, West third ring road , Beijing, 100048, Beijing, China.
[3]Beijing Higher Institution Engineering Research Center of Testing and Imaging,West third ring road , Beijing, 100048, Beijing, China.

*Corresponding author(s). E-mail(s): zhaoxing_1999@126.com;
Contributing authors: GenweiMa@126.com;

Abstract

Spectral computed tomography (SCT) is an powerful imaging modality with broad applications and advantages such as contrast enhancement, artifact reduction, and material differentiation. The positive process or data collected process of SCT is a nonlinear physical process existing scatter and noise, which make it is an extremely ill-posed inverse problem in mathematics. In this paper, we propose a dual-domain iterative network combining a joint learning reconstruction method (JLRM) with a physical process. Specifically, a physical module network is constructed according to the SCT physical process to accurately describe this forward process, which makes the nonlinear use of the traditional mathematical iterative algorithm effective and stable. Additionally, we build a residual-to-residual strategy with an attention mechanism to overcome the slow speed of the traditional mathematical iterative algorithm. We have verified the feasibility of the method through our winning submission to the AAPM DL-spectral CT challenge, and demonstrated that high-accuracy also basis material decomposition results can be achieved with noisy data.

Keywords: Spectral computed tomography, basis material decomposition, deep learning, physical process

1 Introduction

Spectral computed tomography (SCT) is popular and promising in clinical and industrial nondestructive testing. Traditional CT aims to accurately recover a detailed internal image of an object by scanning it from different angles. However, it does not consider the polychromaticity of x-rays. SCT, as shown in Fig. 1, considers the beam spectrum is broad and the polychromatic nature of the x-ray beam makes quantitative imaging. Considering the x-ray spectral information, this leads to more complex reconstruction problems. Hence, in SCT, an object is scanned with different x-ray spectra or spectral separations from the broad spectrum to supplement the information. Spectral x-ray data collection methods include kVp-switching [1], dual-source scanning [2], dual-layer detection, simplistic two-pass scanning, and photon counting detectors [3, 4]. Generally, the collected data of SCT, also known as polychromatic projection data, are utilized to perform energy- and material-selective reconstructions [5–8]. In other words, SCT not only can distinguish materials better but allows quantification of the mass density of two or more materials in a tissue or a mixture with known elemental composition [9].

In mathematics, the SCT reconstruction problem can be seen as solving for $\mu(x, E)$ from K sets of polychromatic projection data under different spectra or spectral separation,

$$q_{k,L}(\vec{x}) = -ln \int S_k(E) e^{-\int_L \mu(\vec{x},E)dL} dE + \epsilon_k, k = 1, 2, \cdots K, \tag{1}$$

where $\mu(\vec{x}, E)$ is the linear attenuation coefficient of the scanned object at point $\vec{x} = (x_1, x_2, x_3)$ with energy E and $S_k(E), k = 1, 2$.is normalized effective spectrum, $\int S_k(E)dE = 1$ and ϵ_k represent scatter or noise of the kth projection. In SCT, the linear attenuation coefficient is generally decomposed as a linear combination of some predefined basis functions,

$$\mu(\vec{x}, E) = \sum_{i=1}^{M} \theta_i(E) f_i(\vec{x}), \tag{2}$$

where $\theta_i(E)$ are energy-dependent functions, and $f_i(\vec{x})$ are functions of spatial position. There are two physical explanations for this equation. The first one is the decomposition method based on the physical effect, where $M = 2$ in this case and the functions based on the photoelectric effect and Compton scattering.The second explanation is basis material decomposition. $\theta_i(E)$, represent

the mass attenuation coefficients of the selected basis materials. The reconstructed $f_i(\vec{x})$, represent the densities of the basis materials of the scanned object.

Substituting (2) into (1), we can obtain the discrete form of SCT as follow:

$$q_{k.L}(\vec{x}) = -ln \sum_{E=1}^{N_k} S_k(E) e^{-\sum_{i=1}^{M} \theta_{i,k}(E) R_{i,k} f_i(\vec{x})} + \epsilon_k, k = 1, 2 \cdots K, \tag{3}$$

where $R_{i,k}, k = 1, 2, \cdots, K$, represents the kth Radon transform.

Compared with traditional CT, the reconstruction problem of SCT is nonlinear and ill-posed. Moreover, real projection data may be inconsistent. Algorithms to solve this problem can be classified in two ways. According to the domain where the data are processed, the methods can be divided into image domain methods [8–14] and projection domain methods [15, 16]. They can also be categorized into direct mapping methods [5, 10–15, 17–19], iterative algorithms [16, 20–31], and deep-learning methods [32–40]. Several researchers combine these classifications.

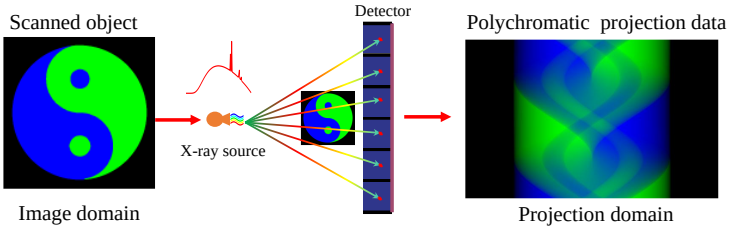


Fig. 1 Spectral computed tomography imaging process.

Image domain methods, also known as post-processing methods, directly process images that have been reconstructed from collected raw data. The reconstructed image is generally solved using a linear algorithm, such as the algebra reconstruction technique (ART) or filter back-projection (FBP). This is actually a first-order approximation of the original nonlinear data, although image domain methods use different basis functions [10, 11, 15, 18] to approximate it. Hence, the reconstructed images still suffer from artifacts such as beam-hardening. Projection domain methods process either raw or log-transformed data before image reconstruction. Theoretically, projection domain methods can obtain better images than image domain methods, because they can compensate for nonlinear effects on material decomposition and suppress artifacts in the decomposition process. However, polychromatic projections require geometric consistency, and it is notoriously difficult to practical SCT system. Direct mapping constructs a basis function and establishes a mapping between the SCT data and basis material. While direct mapping methods are quick and simple and do not consider the spectrum information and attenuation coefficient of materials, they are highly dependent on the

calibration phantom, which limits the accuracy of the reconstructed images. Iteration reconstruction algorithms are the most effective way to solve ill-posed nonlinear inverse problems. Unfortunately, they are time-consuming when the geometric parameters of different spectral scans are inconsistent. Many researchers combine the statistical assumptions of raw data and prior information on images, but this involves hyperparameters that can only be adjusted empirically and do not offer consistent image quality improvement [41]. source-dual-detector CT systems [42, 43]. In recent years, deep learning methods have obtained high-quality reconstructed images with complete training samples. However, much training data are required to train the network-trainable parameters, and it is difficult to obtain in many cases. Sidky *et al.* lead to the AAPM challenge named DL-spectral CT. This challenge provides an opportunity for investigators in SCT image reconstruction, using data-driven or iterative techniques, to compete with their colleagues on the accuracy of their methodology for solving the inverse problem associated with spectral CT acquisition from incomplete, noiseless measurements.[a]

This study aims to demonstrate whether deep neural networks can realize high-precision decomposition images of basis materials and establish a foundation for SCT reconstruction algorithms based on big data and deep learning [41]. We propose an iterative network method that fuses SCT physical processes and traditional mathematical algorithms. Although the AAMP challenge provides accurate physical processes (including spectrum and attenuation coefficient information), our method describes the physical process and relevant physical quantities as learnable network parameters, which is more flexible. As with DL-sparse-view CT, high reconstruction accuracy is only possible if the forward model is explicitly incorporated into the solution map, e.g., by an iterative promotion of data-consistency [44]. Here, we also build physical process into precise physical module, which can realize the accurate expression of the positive process in iterative steps, so as to realize the residual transmission correctly. For the noisy case, we add regularization terms, which is more practical. The main contributions of our work are as follows:

- The proposed method is based on the complementary advantages of physical processing, traditional mathematical iteration algorithms (TMIAs), and deep learning. Physical processing realizes the accurate expression of physical phenomena, TMIAs robustly achieve domain transformation, and deep learning can accurately materials decomposition.
- A network joint physical model governs SCT projection data generation. Hence, the network can more accurately express the SCT physical process. Only when the physical process can be accurately and adequately described by one model can the iterative process continuously improve reconstruction accuracy. The physical model of SCT is defined as a module of the network whose x-ray spectra $S_k(E)$ and attenuation coefficient $\theta(E)$ correspond to the learnable parameters.

[a] https://dl-sparse-view-ct-challenge.eastus.cloudapp.azure.com/competitions/3

- In the iteration process, the input/label data of networks use residual-to-residual to train the network model, rather than image-to-image or image-to-residual. Because places with large errors are observable in the residual image, it is difficult to distinguish them in the image. Moreover, residual-to-residual builds into the residual attention mechanism. Attention operations in residual images make network training easier.
- The network structure is combined with practical problems. The SCT projection data and reconstructed image are domain-related, but even in inconsistent projections, the associated region is relatively small. Therefore, CNN can apply SCT, but the receptive field should not be generous, so as to avoid phenomena such as overfitting and poor robustness.

2 Method

2.1 JLRM architecture

We formulate SCT reconstruction as a nonlinear optimization problem to solve basis material images $\vec{x}^\star$ from acquired polychromatic projections $q_k^\star$:

$$f(\vec{x}^\star) = arg \min_{f(\vec{x})} \|q(f(\vec{x})) - q_k^\star\|_2^2 + \lambda_1 \mathscr{R}_1(q(f(\vec{x}))) + \lambda_2 \mathscr{R}_2(f(\vec{x})), \quad (4)$$

where $q(\vec{x})$ corresponds to the nonlinear SCT physical process, and $\mathscr{R}_1$ and $\mathscr{R}_2$ are network regularization terms in the projection and image domains, respectively, which are used for the case of existing noise or scattering. In this way, $\mathscr{R}_1$ can more accurately describe the physical model for the noise or scattering case. $\mathscr{R}_2$ can effectively suppress noise or scattering and obtain better reconstructed images.

In our experiments, the traditional mathematical iteration algorithm (TMIA) uses the SART algorithm during iteration process and the other TMIA also support. The third step is worth noting. We first calculate the projection residuals for reconstruction, and the results are input to the CNN to obtain the residuals of the base material images. That is, we use the residual-to-residual strategy for network training, which can more effectively improve the image quality.

Generally, the initial value also has a pronounced influence on the convergence speed of the iterative algorithm. Therefore, we incorporate a pre-decomposition module before the iterative process.

Our SCI network includes a pre-decomposition module and an iteration process, as shown in Fig. 2. The pre-decomposition module includes TMIA and CNN steps. For TMIA, we use the E-ART method, which is not sensitive to spectral information, and the quality of reconstructed images is higher than with other analytic algorithms. For the iterative process, the three steps of the incremental method are consistent. As mentioned above, we use a residual-to-residual strategy for effective network training. An attention mechanism is

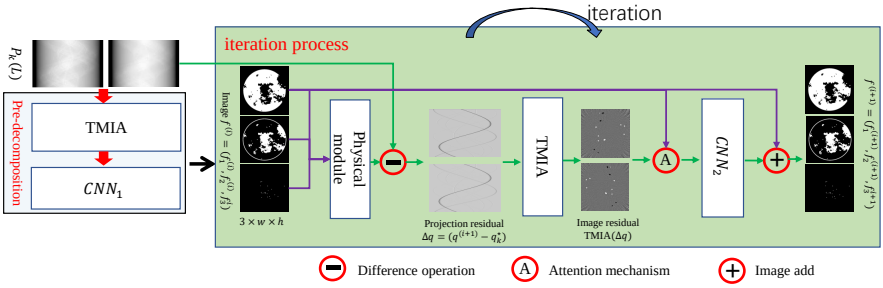


Fig. 2 Architecture of proposed method, including pre-decomposition module and iteration process.

incorporated before the residual image input CNN, for easy network training and quick convergence. Below we show pseudocode for the training stage (Algorithm 1) and test stage (Algorithm 2).

Algorithm 1 Training stage

Input: projection data $q = (q_1, q_2)$, ground truth data $f^\star = (f_1^\star(x), f_2^\star(x), f_3^\star(x))$, iteration number n, relaxation factor ω_1 of E-ART and ω_2 of ART, initial value $f^{(0)} = 0$.
Output: $f^{n+1} = (f_1^{n+1}, f_2^{n+1}, f_3^{n+1})$
step 1: Pre-decomposition (projection domain to image domain)
$f^{(0)} = (f_1^{(0)}, f_2^{(0)}, f_3^{(0)}) = CNN_1(\text{E-ART}(p;\ \omega_1, f^{(0)});\ \Theta_1^{(0)}, \Theta_2^{(0)}, \Theta_3^{(0)})$
$loss\_function = MSELoss(f^{(0)}, f^\star)$
update network parameters $\Theta_1^{(0)}, \Theta_2^{(0)}, \Theta_3^{(0)}$
# step 2: iteration for residual-to-residual assign
for $j = 0$ to n do
  # nonlinear physical module (image domain to projection domain)
  $q_{k.L}^{(j+1)}(\vec{x}) = -ln \sum_{E=1}^{N_k} S_k(E) e^{-\sum_{i=1}^{M} \theta_i(E) R_i f_i^{(j)}(\vec{x})} + \epsilon_k, k = 1, 2, \cdots, K$
  # residual-to-residual assign (projection domain to image domain)
  $\Delta f^{(j)} = CNN_2(\text{ART}(q^{(j+1)} - q^\star;\ \omega_2), f_1^{(j)}, f_2^{(j)}, f_3^{(j)};\ \Theta_1^{(j)}, \Theta_2^{(j)}, \Theta_3^{(j)})$
  $loss\_function = MSELoss(\Delta f^{(j)}, f^\star - f^{(j)})$
  update network parameter $\Theta_1^{(j)}, \Theta_2^{(j)}, \Theta_3^{(j)}$
  # update basis material images
  $f_1^{(j+1)} = f_1^{(j)} + \Delta f_1^{(j)}, f_2^{(j+1)} = f_2^{(j)} + \Delta f_2^{(j)}, f_3^{(j+1)} = f_3^{(j)} + \Delta f_3^{(j)}$
end for

The proposed method is based on the complementary advantages of physical processing, TMIAs, and deep learning. Physical processing realizes the accurate expression of physical phenomena, while TMIAs achieve domain transformation, and deep learning can accurately materials decomposition. The three parts are interdependent and complementary.

Algorithm 2 test stage

Input: projection data $q = (q_1, q_2)$ iteration number n, relaxation factor ω_1 of E-ART and ω_2 of ART.
Output: $f^{n+1} = (f_1^{n+1}, f_2^{n+1}, f_3^{n+1})$
step 1: Pre-decomposition
$f^{(0)} = (f_1^{(0)}, f_2^{(0)}, f_3^{(0)}) = CNN_1(\text{E-ART}(p;\ \omega_1);\ \Theta_1^{(0)}, \Theta_2^{(0)}, \Theta_3^{(0)})$
# step 2: Iteration for residual-to-residual assign
for $j = 0$ to n do
  # nonlinear physical module (image domain to projection domain)
  $q_{k.L}^{(j+1)}(\vec{x}) = -ln \sum_{E=1}^{N_k} S_k(E) e^{-\sum_{i=1}^{M} \theta_i(E) R_i f_i^{(j)}(\vec{x})} + \epsilon_k, k = 1, 2, \cdots, K$
  # residual-to-residual assign (projection domain to image domain)
  $\Delta f^{(j)} = CNN_2(\text{ART}(q^{(j+1)} - q^{\star};\ \omega_2), f_1^{(j)}, f_2^{(j)}, f_3^{(j)};\ \Theta_1^{(j)}, \Theta_2^{(j)}, \Theta_3^{(j)})$
  # update basis material images
  $f_1^{(j+1)} = f_1^{(j)} + \Delta f_1^{(j)}, f_2^{(j+1)} = f_2^{(j)} + \Delta f_2^{(j)}, f_3^{(j+1)} = f_3^{(j)} + \Delta f_3^{(j)}$
end for

2.1.1 Physical module for nonlinear physical process

Only when the physical process can be accurately described by the model can the iterative reconstruction process continuously improve reconstruction accuracy. Here, to accurately describe the nonlinear physical process of SCT, its positive process is equivalently mapped into a network module. As shown in Fig. 3, Firstly, do a Radon transform for three basis material. This can be achieved by a projection layer [45, 46]. Here, we have overridden the interface which projection and back-projection operators are provide by official. Next, the module can be constructed according to the discrete polychromatic projection formula, where the spectral information and attenuation coefficient can be set as corresponding learning parameters. In fact, both sets of learnable parameters can be achieved by a convolution operator. For the attenuation coefficient, this layer can set K convolution kernels, of size $N_k \times M \times 1 \times 1$ ($out\_channels \times in\_channels \times height \times width$). For spectral information, the size of K convolution kernels is $1 \times N_k \times 1 \times 1$. For the noise case, a regularization term network should be added in the physical module to more accurately express the physical process.

2.1.2 CNN network structure design

The network structure is also critical to the reconstruction result, because our network is trained block-to-block. Once overfitting occurs, the results of subsequent iterations are disastrous. The CNNs of pre-decomposition and iterative processes are basically similar, with structures as displayed in Fig. 4. The biggest difference between them is the channel setting of input data, in which the channel of CNN_1 is set to 2, while that of CNN_2 is 6. The CNN contains eight blocks, each with three convolution layers, and the kernel sizes

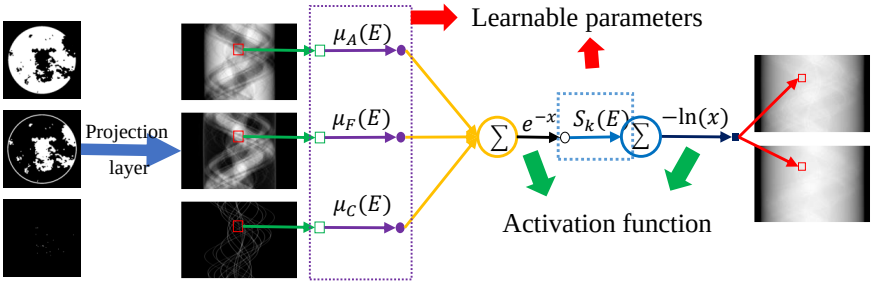


Fig. 3 Physical process equivalent mapping physical module.

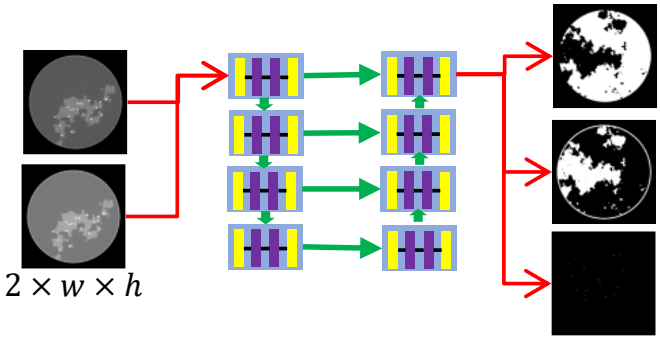


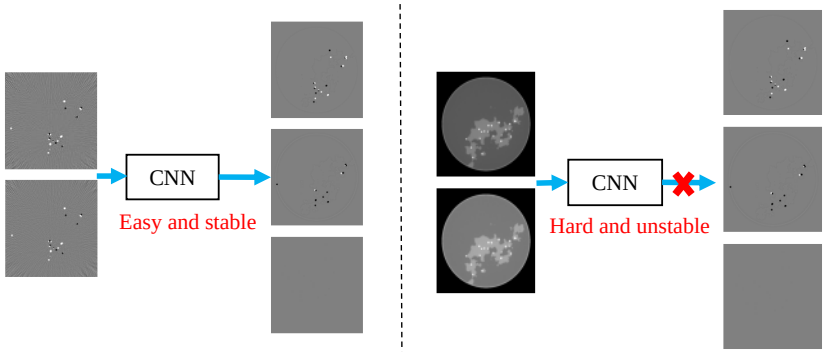
Fig. 4 CNN network structure with eight blocks, each with three convolution layers.

are $3 \times 3, 1 \times 1$, and 3×3. It is worth noting that network operations such as downsampling, upsampling, and transpose convolution, which make the receptive field large, cannot be used. Although the projection is inconsistent and the pixels of basis material images are domain-related, the regional correlation caused by the inconsistency is very small, and increasing the receptive field will aggravate overfitting, which is fatal for later iterations to obtain more accurate decomposition results.

2.1.3 Residual-to-residual strategy

The residual-to-residual strategy proposed based on experiments observations which can effectively improve the accuracy of reconstruction. As shown in Fig. 5, regions with large errors can be directly located in the residual image, where it is difficult to find in reconstructed images. Hence, the image-to-residual strategy may bring new errors, with many uncertainties to improve the quality of reconstructed images. In other words, image-to-residual is unstable for improving image quality.

The ReLU activation function is often used in the network, which makes the images obtained by the network nonnegative. The residual image does not possess this property. To use the nonlinear ReLU activation function, we



(a) Residual-to-residual mapping (b) Image-to-residual mapping

Fig. 5 Residual-to-residual vs. image-to-residual strategy. Residual-to-residual strategy makes it easier to locate where there are large errors.

decompose the residual image dx into the form

$$dx = dx^{+} - dx^{-}, \tag{5}$$

where dx^{+} and dx^{-} are, respectively, the positive and negative parts of the residual image. Thus, the network output becomes an image with two nonnegative channels.

2.1.4 Residual attention mechanism

Generally speaking, in reconstructed residual images, even if the decomposition is correct, there are still some errors, and this phenomenon occurs at the edges of basis material images. This can be overcome by incorporating an attention mechanism [47–49] to emphasize the edges, which can make the network training more simple and efficient. Attention includes both channel and spatial attention. In our case, we need to emphasize edges that are spatial features, so spatial attention is employed. Given a tensor of intermediate feature map $\eta \in R^{c\times h\times w}$, spatial attention generates a spatial attention map $A \in R^{1\times h\times w}$ by utilizing the inter-spatial relationships of features. Then the output of the attention mechanism, $\widehat{\eta} \in R^{c\times h\times w}$, can be expressed as

$$\widehat{\eta} = A \odot \eta, \tag{6}$$

where $\odot$ denotes element-wise multiplication and c is the channels of input data. During multiplication, the attention values are broadcast (copied) to all feature maps along the channel dimension.

The spatial attention focusing on "where" is the most informative part. Considering that there are three materials, we implement the attention mechanism in three directions.

As shown in Fig. 6, our attention mechanism consists of three pipelines to construct three spatial attention maps. Given a two-dimensional feature map

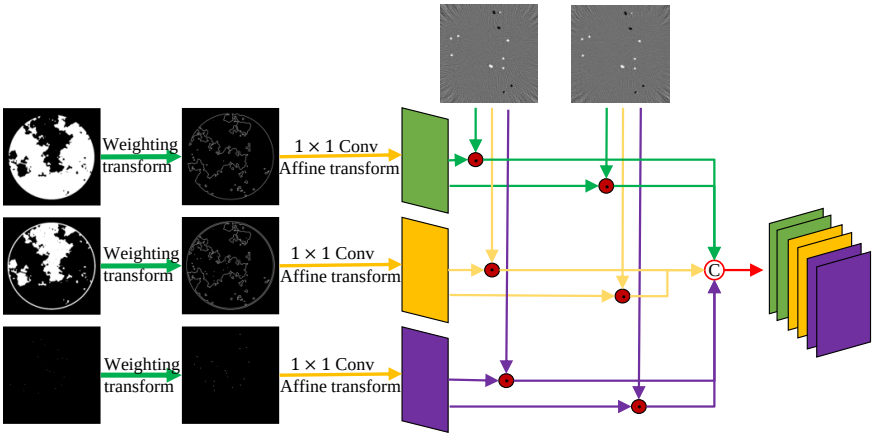


Fig. 6 Residual attention mechanism.

$f \in R^{h \times w}$, which is obtained by applying the weighting transform on three basis material images, let us define

$$\begin{cases} A_1 = a_1 \cdot sigmoid(2(0.5 - \|f_1 - 0.5\|)) + b_1 \\ A_2 = a_2 \cdot sigmoid(2(0.5 - \|f_2 - 0.5\|)) + b_2 \\ A_3 = a_3 \cdot sigmoid(2(0.5 - \|f_3 - 0.5\|)) + b_3 \end{cases}, \tag{7}$$

where a_1, b_1 and a_2, b_2 represent the weight parameters and bias in the 1×1 convolutional layer, and *sigmoid* denotes the sigmoid activation function. Here, we set the initial values as $a_i = 1, i = 1, 2, 3$, and $b_i = 0, i = 1, 2, 3$. To apply the attention mechanism, each residual image element-wise multiply spatial attention maps and obtained feature maps would have been re-weighted, which is helpful to accelerate the convergence of the network.

2.2 Network training

The proposed method was implemented using Python under the PyTorch framework. Each module was trained module-by-module, and the network was optimized using the Adam optimization method with a min-batch of nine image patches for each iteration. The objective function for each module used an MSE loss function,

$$\min_{\Theta} L(\Theta) = \sum_{i=1}^{N} \|CNN(x, \Theta) - y^*\|, \tag{8}$$

where x, y^* are the input/label data, and Θ denotes the network learnable parameters. The complete training samples were randomly shuffled, and as an empirical stopping criterion of each module, the number of training epochs was

Table 1 System and geometric scanning parameters of AAPM DL-spectral CT challenge dataset

X-ray source to rotation center (mm)	500
X-ray source to detector distance (mm)	1000
Scanning angle (deg)	$[0°, 360°]$
Scanning angular interval (deg)	1.40625
Number of detector units	1024
Detector unit size (mm)	0.3574
Reconstructed image size (pixel)	512×512

set to 1000. Comparative experiments were performed on a PC (AMD EPYC-Rome processor, 128 GB RAM). Training was executed on a single graphic processing Nvidia RTX A6000 with 48 GB of video memory.

3 Results

Experiments were carried out to demonstrate the behavior of the proposed method, including architecture validation, for example, residual analysis and the performance of JLRM in noise suppression, reconstruction method, CNN network structure design and residual-to-residual strategy.

3.1 DL-spectral CT setup of AAPM challenge

This challenge focuses on spectral CT, where the object is scanned by X-rays with two spectra, i.e., $K = 2$, without noise. This challenge model an ideal form of fast kVp-switching where the X-ray tube voltage alternates between two settings for consecutive projections. For each kVp setting, the beam spectrum is broad and the polychromatic nature of the X-ray beam makes quantitative imaging in CT quite challenging when only one kVp setting is used. Scanned objects are simulated for a breast model [50] that contains three tissue types ($M = 3$): adipose, fibroglandular, and calcification. Truth images and simulated data are provided for 1000 cases so that participants can decide to use either a data-driven or optimization-based technique. The SCT imaging process can be seen in Fig. 7, which shows the spectral information used in the imaging process and the linear attenuation coefficient of the material. It can be observed that the linear attenuation coefficients of the two materials are very close. This is a tricky point, as discussed in the experimental results. To be fair, the test dataset adopts 10 additional cases, which are provided in a "starting kit" with ground-truth images. The fan-beam geometric parameters of the SCT system are listed in Table 1. The DL-spectral-CT challenge seeks the image reconstruction algorithm that provides the most accurate reconstruction of the adipose, fibroglandular, and calcification spatial maps from SCT projection data. Details about the challenge setup and results can be found on the official website.[b]

[b] https://dl-sparse-view-ct-challenge.eastus.cloudapp.azure.com/competitions/3#learn_the_details

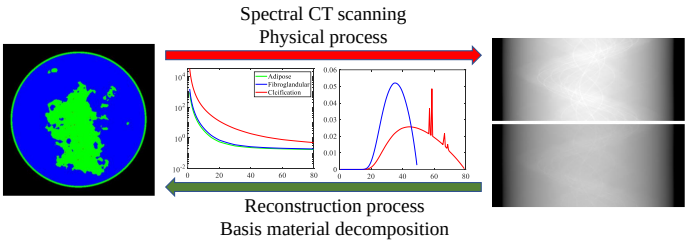


Fig. 7 SCT imaging, where forward process is data collection, which is also a physical process. Inverse process is reconstruction, i.e., basis material decomposition.

Table 2 Average RMSE for quantitative evaluation.

		Pre-decomposition	1 iteration	2 iterations
adipose	training set	2.53×10^{-3}	9.24×10^{-5}	6.79×10^{-6}
	test set	2.87×10^{-3}	7.09×10^{-5}	7.54×10^{-7}
fibroglandular	training set	2.68×10^{-3}	1.20×10^{-4}	5.54×10^{-6}
	test set	3.06×10^{-3}	9.74×10^{-5}	6.67×10^{-7}
calcification	training set	3.30×10^{-7}	6.60×10^{-7}	3.00×10^{-7}
	test set	6.40×10^{-7}	6.30×10^{-7}	6.60×10^{-7}

3.2 Performance and analysis of JLRM

We demonstrate that JLRM can be trained to accurately reconstruct basis material images from collected projection data obtained from the DL-spectral CT of the AAPM challenge. As described in section 2, the JLRM method includes a pre-decomposition module and two iteration of the iteration process. Fig. 8 shows the basis material images of these three modules. It is difficult to visually distinguish the image quality reconstructed by each module. The last row indicates the tissue label map of the three materials, from which we can clearly see the proportion of each material in each pixel. Fig. 9 shows the different images. The accuracy of the proposed method can be further improved, and the sequence module can continuously improve the accuracy of reconstruction based on comparison with the differential images of each module. Even in the display window with almost mechanical accuracy, the difference images of the final result (2 iterations) are difficult to indicate a large error. This can verify that our method can achieve near exact reconstruction. Table 2 shows the average RMSE for quantitative evaluation, which further indicates that our method can obtain accurate reconstruction. The accuracy of the results finally submitted in the AAPM challenge is 6.8×10^{-7}. Report to the second place accuracy of 6.21×10^{-6}, our reconstruction accuracy is higher one order of magnitude.

To better understand the behavior of each module of the proposed method, Fig. 10 shows the loss function update curve of each material of each module. In the pre-decomposition module, adipose and fibroglandular material cannot reach the mechanical accuracy, but more cases of calcification material can

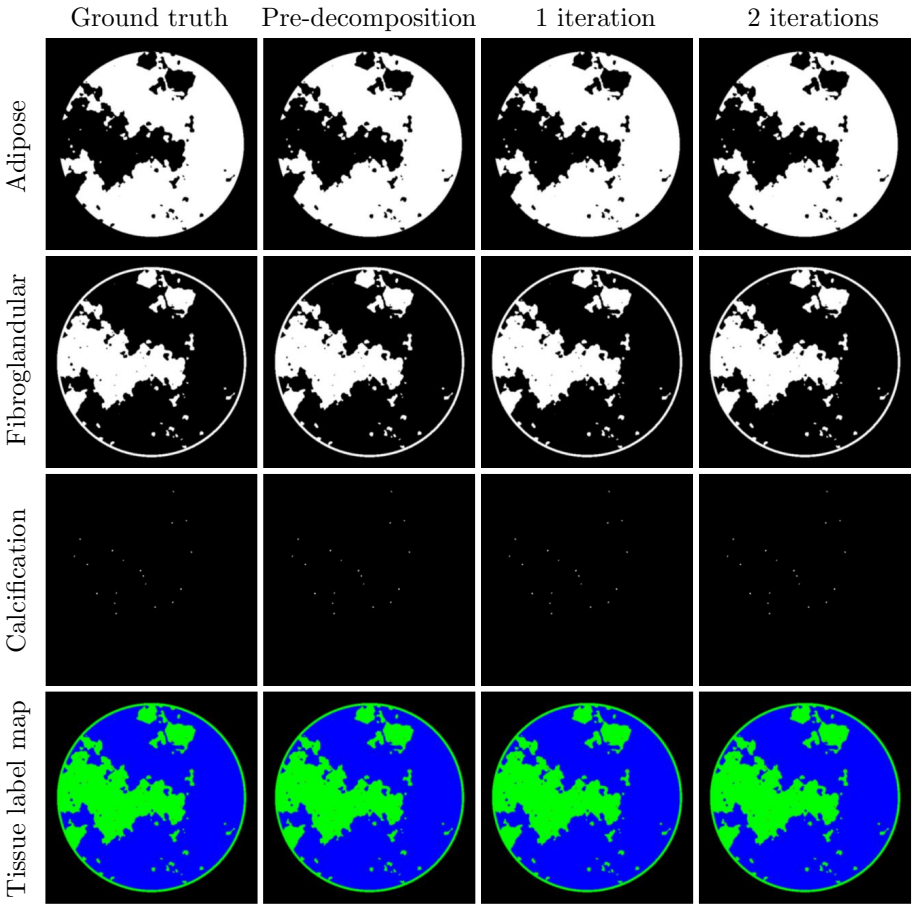


Fig. 8 Left to right: ground truth, results of pre-decomposition block, iteration process with 1 and 2 iterations. Display windows: $[0.15, 0.35]$ for adipose material, $[0.3, 0.7]$ for fibroglandular material, $[0.1, 0.5]$ for calcification material. Last row: tissue label map with adipose (blue), fibroglandular (green), calcification (red).

reach the mechanical accuracy. In the iteration process, the accuracy can be well improved using the residual-to-residual strategy.

3.3 Residual analysis

Because the projection data of the AAPM challenge are noise-free, a crucial way of a proper solver for this DL spectral-CT problem is evaluating this difference images $q_k^n - q_k^\star$, which should be as low as possible. We analyze this aspect in Fig. 11. For better analysis and observation, we reconstructed the residual images from the residual of projection. Fig. 12 displays the reconstructed residual images. We note that the residual images in the image domain also showed a decreasing trend. The results of the final iteration showed that each pixel of the residual was all within the mechanical accuracy.

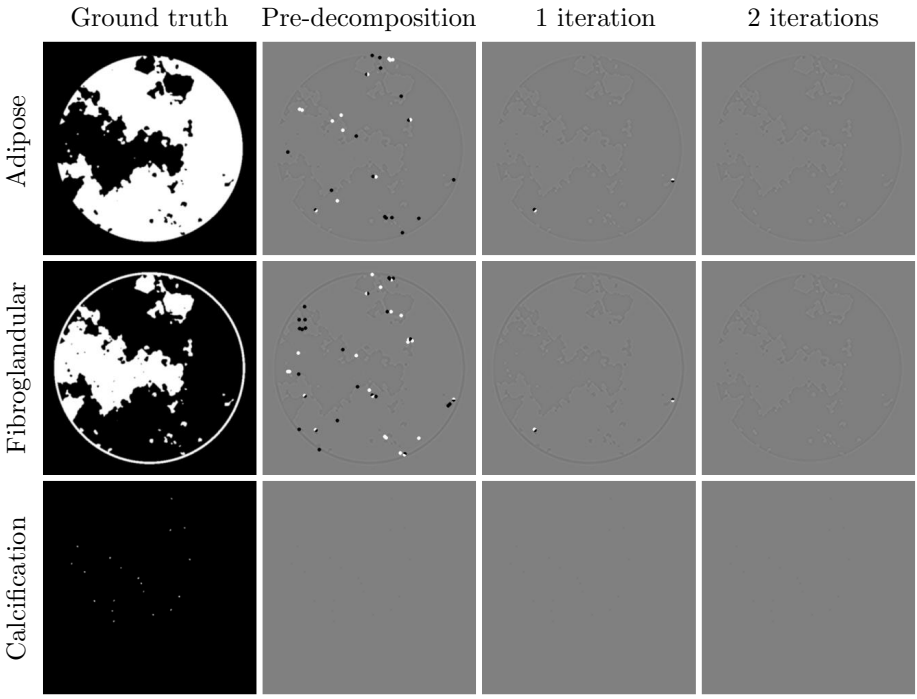


Fig. 9 Left to right: ground truth, difference images of pre-decomposition block, iteration process with 1 and 2 iterations. Display windows of difference images are [-0.00005, 0.00005].

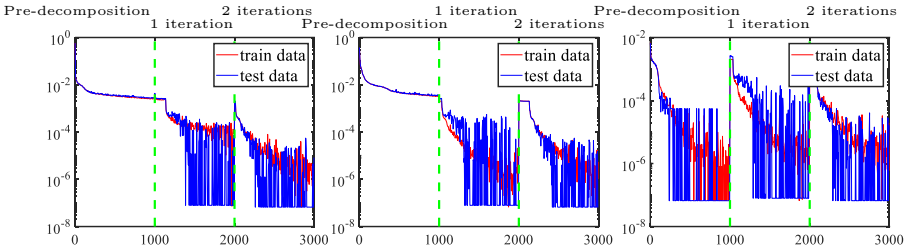


Fig. 10 Update curves of three material loss functions. Left to read: adipose, fibroglandular, and calcification material.

Some pixels with small residuals, marked by red boxes in Fig. 12, cannot be updated effectively in the first iteration, and are shown to be more obvious in the residual image after one iteration, which facilitates further performance improvement. Some regions with large errors can continuously shrink during the iteration process, although updates still have larger errors, as marked by blue boxes.

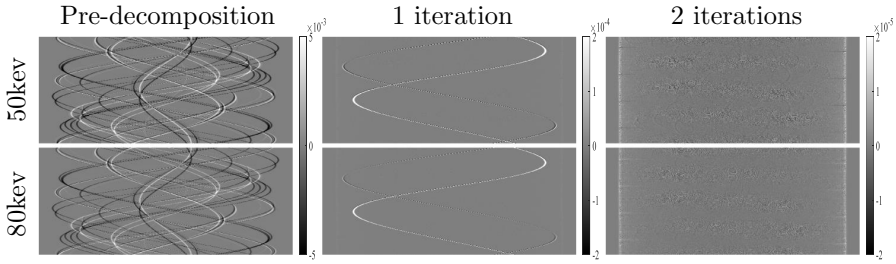


Fig. 11 Errors of projection domain for each module.

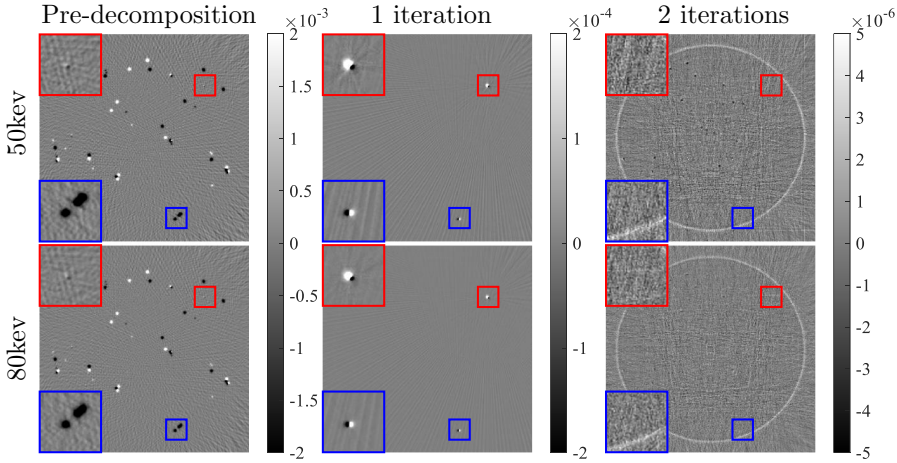


Fig. 12 Errors of image domain for each module. Colored boxes mark trend of updates during iterative process; corresponding zoomed-in views are displayed in left corner.

3.4 Performance of JLRM in noise suppression

To ensure universality, we study the noise data. The statistics of x-ray measurements are often described using a Poisson distribution to model the incident intensity,

$$I_{noisy,k} \sim Poisson(I_{0,k} \times e^{-q_k}), k = 1, \cdots, K, \tag{9}$$

where $I_{noisy,k}$ and q_k denote received intensity of detectors and noise-free projection data, respectively. The noisy projection data $q_{noisy,k} = -\ln(\frac{I_{noisy,k}}{I_{0,k}})$.

Table 3 lists the accuracy of the proposed method under different noise levels. As shown in previous experimental results, the phenomenon of adipose and fibroglandular materials is also consistent, mainly because the attenuation coefficients of the two materials are very similar. Hence, the two materials are easily decomposed incorrectly, and their performance evaluation indices are basically consistent. The attenuation coefficient of calcification material is large, and its size is very small; hence, calcification material can still have mechanical accuracy with high noise levels ($I_0 = 1 \times 10^5$) and can do so quickly when the noise level is relatively low ($I_0 = 1 \times 10^6$).

Table 3 Average RMSE of dataset with different noise levels for quantitative evaluation.

	Noise level	1×10^5	5×10^5	1×10^6
adipose	training set	7.84×10^{-2}	2.87×10^{-2}	6.58×10^{-3}
	test set	9.97×10^{-2}	4.45×10^{-2}	7.54×10^{-3}
fibroglandular	training set	7.23×10^{-2}	3.24×10^{-2}	6.69×10^{-3}
	test set	9.17×10^{-2}	4.68×10^{-2}	8.20×10^{-3}
calcification	training set	7.07×10^{-5}	3.54×10^{-6}	2.72×10^{-6}
	test set	8.92×10^{-5}	5.47×10^{-6}	1.03×10^{-6}

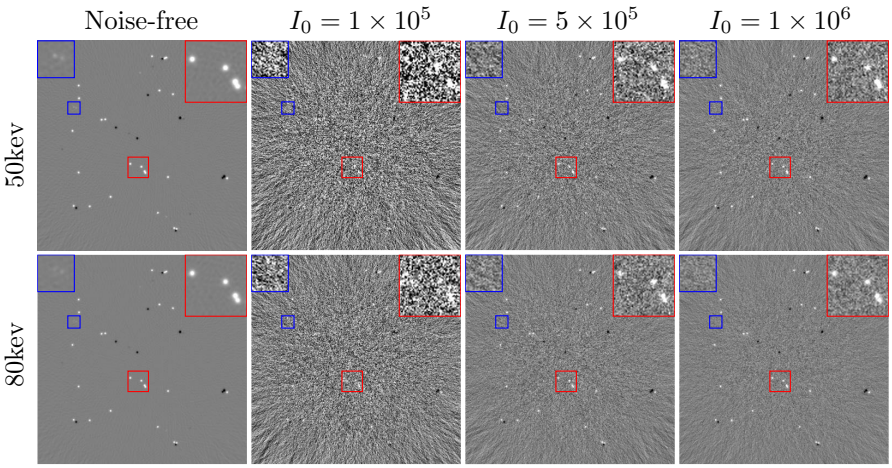


Fig. 13 Residual images with different noise levels. Colored boxes mark details of images under different noise levels; corresponding zoomed-in views are displayed in upper corners.

To further understand the effect of noise on the decomposition results of adipose and fibroglandular materials, we analyze a set of results obtained from noise-free pre-decomposition. As in the residual analysis of section 3.3, the obtained projection domain residuals are reconstructed to obtain the results shown in Fig. 13, where the projection residuals are obtained by performing the physical process with the pre-decomposition results minus the noisy projections, i.e., $q_k^n - q_{noisy,k}$. Fig. 13 shows that it is too difficult to reflect some decomposition errors in the image when the noise is relatively large, such as those marked in the red box (zoomed-in view is shown in upper-right corner). However, in some locations where the error is smaller, the noise is small but also difficult to decompose, while the noise-free case can be obviously found, such as in the blue box (zoomed-in view shown in upper-left corner).

3.5 Receptive field: small or large?

As argued in CNN network structure design of section 2 increasing the receptive field will aggravate the overfitting phenomenon, which is fatal for the later iteration to obtain more accurate decomposition results. Fig. 14 shows the loss function curve update of the pre-decomposition module. During the training

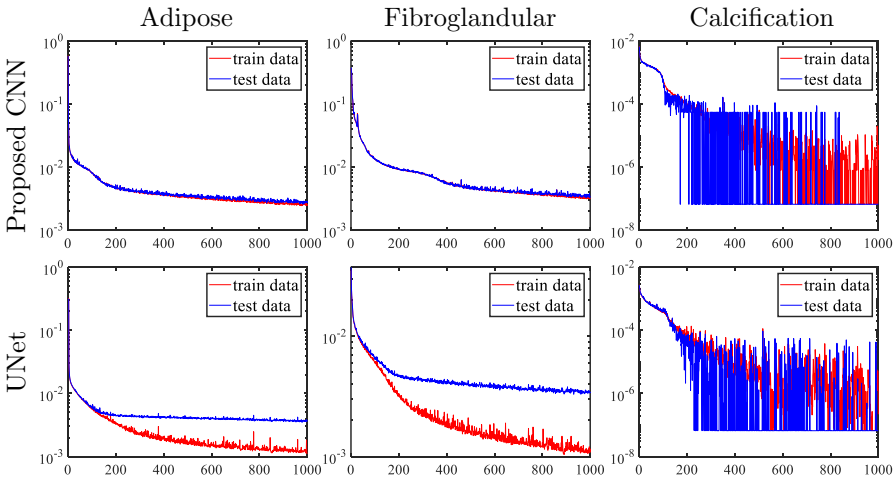


Fig. 14 Update curves of three material loss functions for pre-decomposition module with networks with different receptive fields.

process of the proposed CNN, the changes of the training and test sets are essentially the same with the iterative update. For UNet, using pooling and upsampling layers causes the network to have a large receptive field. However, the experimental results show that adipose and fibroglandular materials display an obvious overfitting phenomenon after a period of training.

Fig. 15 shows the performance of overfitting in the reconstructed image of adipose material, where fibroglandular material has the same phenomenon. The training set has higher accuracy for UNet, and the error in the difference image is quite small, while the test set has a significantly larger error.

3.6 Residual-to-residual vs. image-to-residual

We research the mapping strategy. Fig. 16 displays the partial results of two iterations of the image-to-residual and residual-to-residual mapping strategies for adipose and fibroglandular materials. The image-to-residual mapping strategy may bring new reconstruction errors, so that the accuracy of the final reconstruction cannot be really improved, and there exists the risk of instability. Due to the places with large errors are difficult to distinguish them in the reconstruction images. The residual-to-residual strategy can effectively improve the accuracy compared with the image-to-residual strategy. The high-precision reconstruction results can be attributed to the following: 1) It can calculate the residual in combination with the physical module; 2) The attention mechanism is used to reduce the complexity of network decomposition; 3) The residual-to-residual strategy avoids large numbers eating decimals.

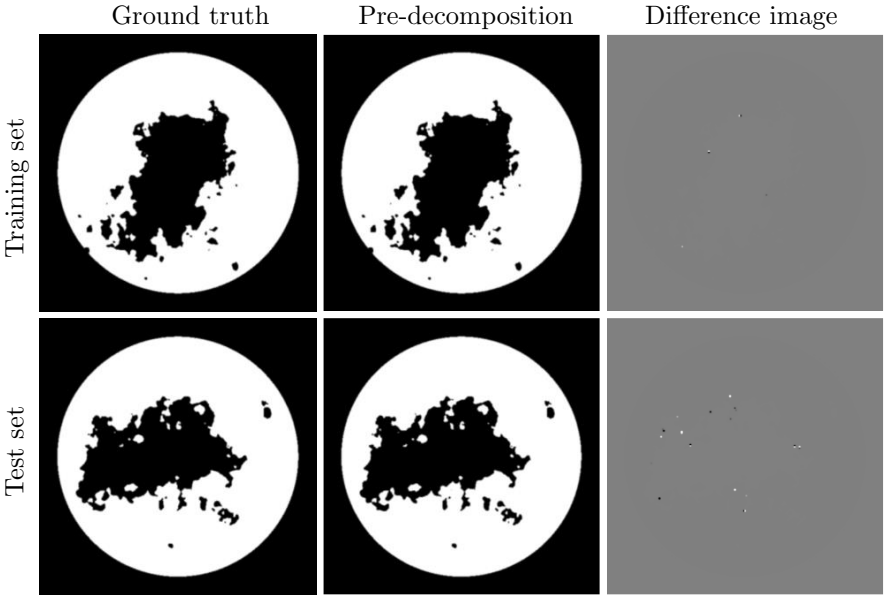


Fig. 15 Results of reconstruction of training and test sets using a larger receptive field. Display windows: [0,1] for material images, and [-0.05,0.05] for difference images.

4 Discussion

To solve the problem of high-precision base material decomposition in spectral CT, this paper proposes an algorithm fused on a physical model, traditional mathematical iterative algorithm, and deep learning. Each part has clear functions and is built based on given problems and actual situations. The physical module, which is embedded and expressed as a network structure, can accurately describe the physical process of SCT imaging. The mathematical iterative algorithm realizes information transfer between domains. These three modules complement each other. We introduce some network skills or strategies in some modules, which are observations related to SCT problems, and can improve the reconstruction results. Examples are a residual attention mechanism and the study of the receptive field.

To further illustrate the effectiveness of our method, we study it when the data are noisy.

The JLRM algorithm is based on deep learning, which mixes mathematics and physics. It can obtain satisfactory decomposition results due to a large number of training datasets. In practice, it is difficult to obtain a large volume of SCT projection data, and it requires matching label data (basis material images). The current energy spectrum is based on the data generated under a single specific energy spectrum. However, the spectral curves of the actual scanning system under the same energy are not necessarily consistent.

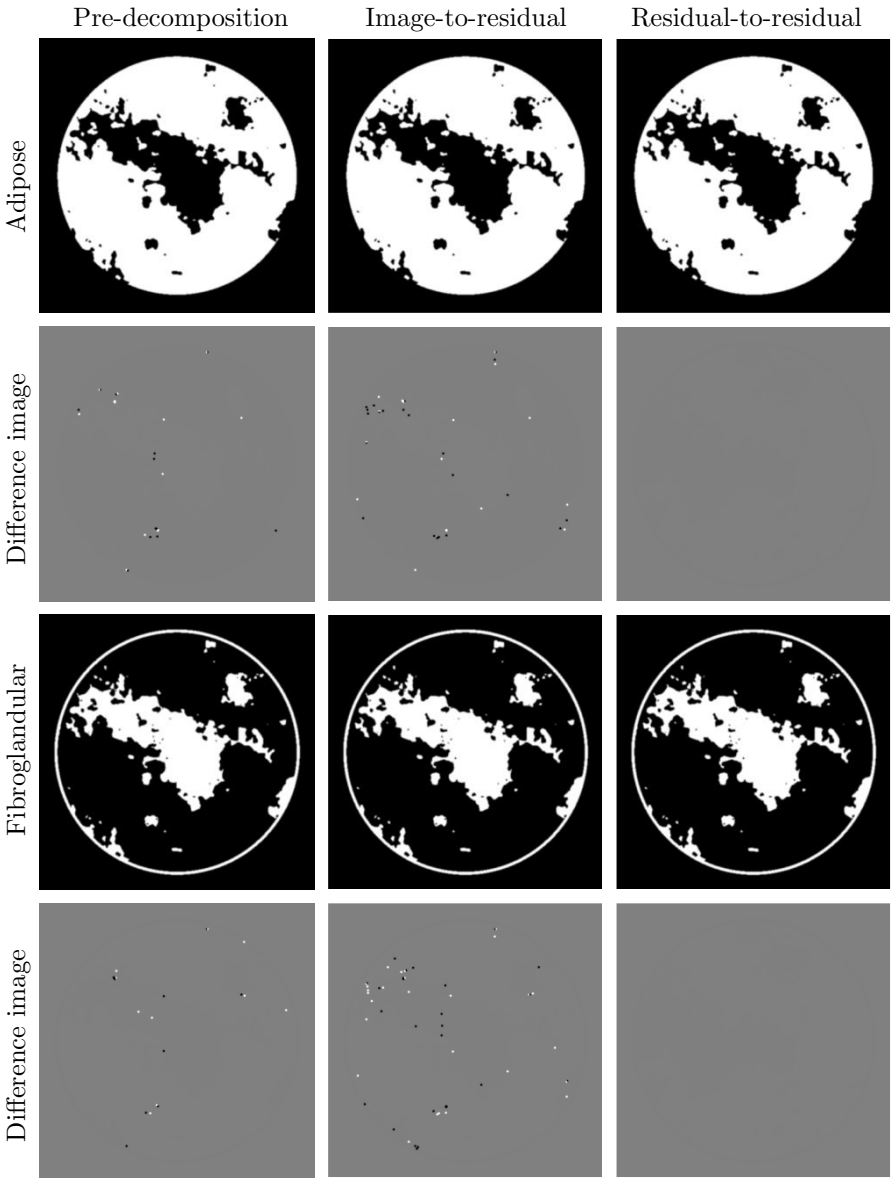


Fig. 16 Results of residual-to-residual and image-to-residual mapping strategies. Display windows: [0,1] for material images and [-0.05,0.05] for different images.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 61827809) and the National Key Research and Development Program of China (No. 2020YFA0712200). The authors are grateful to the Beijing Higher

Institution Engineering Research Center of Testing and Imaging and the Beijing Advanced Innovation Center for Imaging Technology for funding this research work.

References

[1] Xu, D., Langan, D.A., Wu, X., Pack, J.D., Benson, T.M., Tkaczky, J.E., Schmitz, A.M.: Dual energy ct via fast kvp switching spectrum estimation. In: Medical Imaging 2009: Physics of Medical Imaging, vol. 7258, pp. 1198–1207 (2009). SPIE

[2] Petersilka, M., Bruder, H., Krauss, B., Stierstorfer, K., Flohr, T.G.: Technical principles of dual source ct. European journal of radiology **68**(3), 362–368 (2008)

[3] Leng, S., Bruesewitz, M., Tao, S., Rajendran, K., Halaweish, A.F., Campeau, N.G., Fletcher, J.G., McCollough, C.H.: Photon-counting detector ct: System design and clinical applications of an emerging technology. Radiographics **39**(3), 729–743 (2019). https://doi.org/10.1148/rg.2019180115

[4] Willemink, M.J., Persson, M., Pourmorteza, A., Pelc, N.J., Fleischmann, D.: Photon-counting ct: Technical principles and clinical prospects. Radiology **289**(2), 293–312 (2018). https://doi.org/10.1148/radiol.2018172656

[5] Alvarez, R.E., Macovski, A.: Energy-selective reconstructions in x-ray computerised tomography. Physics in Medicine and Biology **21**(5), 733 (1976)

[6] Alvarez, R., Seppi, E.: A comparison of noise and dose in conventional and energy selective computed tomography. IEEE Transactions on Nuclear Science **26**(2), 2853–2856 (1979)

[7] Alvarez, R.E., Seibert, J.A., Thompson, S.K.: Comparison of dual energy detector system performance. Medical physics **31**(3), 556–565 (2004)

[8] Kalender, W.A., Perman, W., Vetter, J., Klotz, E.: Evaluation of a prototype dual-energy computed tomographic apparatus. i. phantom studies. Medical physics **13**(3), 334–339 (1986)

[9] Yu, L., Leng, S., McCollough, C.H.: Dual-energy ct–based monochromatic imaging. American journal of Roentgenology **199**(5_supplement), 9–15 (2012)

[10] Brooks, R.A., Di Chiro, G.: Beam hardening in x-ray reconstructive tomography. Physics in medicine & biology **21**(3), 390 (1976)

[11] Coleman, A., Sinclair, M.: A beam-hardening correction using dual-energy computed tomography. Physics in medicine & biology **30**(11), 1251 (1985)

[12] Cardinal, H.N., Fenster, A.: An accurate method for direct dual-energy calibration and decomposition. Medical physics **17**(3), 327–341 (1990)

[13] Dong, X., Niu, T., Zhu, L.: Combined iterative reconstruction and image-domain decomposition for dual energy ct using total-variation regularization. Medical physics **41**(5), 051909 (2014)

[14] Zeng, D., Huang, J., Zhang, H., Bian, Z., Niu, S., Zhang, Z., Feng, Q., Chen, W., Ma, J.: Spectral ct image restoration via an average image-induced nonlocal means filter. IEEE Transactions on Biomedical Engineering **63**(5), 1044–1057 (2015)

[15] Stenner, P., Berkus, T., Kachelriess, M.: Empirical dual energy calibration (edec) for cone-beam computed tomography. Medical physics **34**(9), 3630–3641 (2007)

[16] Maaß, N., Sawall, S., Knaup, M., Kachelrieß, M.: Empirical multiple energy calibration (emec) for material-selective ct. In: 2011 IEEE Nuclear Science Symposium Conference Record, pp. 4222–4229 (2011). IEEE

[17] Ying, Z., Naidu, R., Crawford, C.R.: Dual energy computed tomography for explosive detection. Journal of X-ray Science and Technology **14**(4), 235–256 (2006)

[18] Maaß, C., Meyer, E., Kachelrieß, M.: Exact dual energy material decomposition from inconsistent rays (mdir). Medical physics **38**(2), 691–700 (2011)

[19] Feng, C., Shen, Q., Kang, K., Xing, Y.: An empirical material decomposition method (emdm) for spectral ct. In: 2016 IEEE Nuclear Science Symposium, Medical Imaging Conference and Room-Temperature Semiconductor Detector Workshop (NSS/MIC/RTSD), pp. 1–5 (2016). IEEE

[20] Sukovic, P., Clinthorne, N.H.: Penalized weighted least-squares image reconstruction for dual energy x-ray transmission tomography. IEEE transactions on medical imaging **19**(11), 1075–1081 (2000)

[21] Elbakri, I.A., Fessler, J.A.: Statistical image reconstruction for polyenergetic x-ray computed tomography. IEEE transactions on medical imaging **21**(2), 89–99 (2002)

[22] Xu, Q., Mou, X., Tang, S., Hong, W., Zhang, Y., Luo, T.: Implementation of penalized-likelihood statistical reconstruction for polychromatic dual-energy ct. In: Medical Imaging 2009: Physics of Medical Imaging, vol.

7258, pp. 1666–1674 (2009). SPIE

[23] Long, Y., Fessler, J.A.: Multi-material decomposition using statistical image reconstruction for spectral ct. IEEE transactions on medical imaging **33**(8), 1614–1626 (2014)

[24] Niu, T., Dong, X., Petrongolo, M., Zhu, L.: Iterative image-domain decomposition for dual-energy ct. Medical physics **41**(4), 041901 (2014)

[25] Zhang, R., Thibault, J.-B., Bouman, C.A., Sauer, K.D., Hsieh, J.: Model-based iterative reconstruction for dual-energy x-ray ct using a joint quadratic likelihood model. IEEE transactions on medical imaging **33**(1), 117–134 (2013)

[26] Zhao, Y., Zhao, X., Zhang, P.: An extended algebraic reconstruction technique (e-art) for dual spectral ct. IEEE transactions on medical imaging **34**(3), 761–768 (2014)

[27] Hu, J., Zhao, X., Wang, F.: An extended simultaneous algebraic reconstruction technique (e-sart) for x-ray dual spectral computed tomography. Scanning **38**(6), 599–611 (2016)

[28] Chen, B., Zhang, Z., Sidky, E.Y., Xia, D., Pan, X.: Image reconstruction and scan configurations enabled by optimization-based algorithms in multispectral ct. Physics in Medicine & Biology **62**(22), 8763 (2017)

[29] Li, M., Zhao, Y., Zhang, P.: Accurate iterative fbp reconstruction method for material decomposition of dual energy ct. IEEE transactions on medical imaging **38**(3), 802–812 (2018)

[30] Zhao, S., Pan, H., Zhang, W., Xia, D., Zhao, X.: An oblique projection modification technique (opmt) for fast multispectral ct reconstruction. Physics in Medicine & Biology **66**(6), 065003 (2021)

[31] Zhang, W., Zhao, S., Pan, H., Zhao, Y., Zhao, X.: An iterative reconstruction method based on monochromatic images for dual energy ct. Medical Physics **48**(10), 6437–6452 (2021)

[32] Touch, M., Clark, D.P., Barber, W., Badea, C.T.: A neural network-based method for spectral distortion correction in photon counting x-ray ct. Physics in Medicine & Biology **61**(16), 6132 (2016)

[33] Liao, Y., Wang, Y., Li, S., He, J., Zeng, D., Bian, Z., Ma, J.: Pseudo dual energy ct imaging using deep learning-based framework: basic material estimation. In: Medical Imaging 2018: Physics of Medical Imaging, vol. 10573, pp. 1190–1194 (2018). SPIE

[34] Zhao, W., Vernekohl, D., Han, F., Han, B., Peng, H., Yang, Y., Xing, L., Min, J.K.: A unified material decomposition framework for quantitative dual-and triple-energy ct imaging. Medical physics **45**(7), 2964–2977 (2018)

[35] Xu, Y., Yan, B., Chen, J., Zeng, L., Li, L.: Projection decomposition algorithm for dual-energy computed tomography via deep neural network. Journal of X-ray science and technology **26**(3), 361–377 (2018)

[36] Xu, Y., Yan, B., Zhang, J., Chen, J., Zeng, L., Wang, L.: Image decomposition algorithm for dual-energy computed tomography via fully convolutional network. Computational and mathematical methods in medicine **2018** (2018)

[37] Hu, D., Wu, W., Xu, M., Zhang, Y., Liu, J., Ge, R., Chen, Y., Luo, L., Coatrieux, G.: Sister: Spectral-image similarity-based tensor with enhanced-sparsity reconstruction for sparse-view multi-energy ct. IEEE Transactions on Computational Imaging **6**, 477–490 (2019)

[38] Wu, X., He, P., Long, Z., Guo, X., Chen, M., Ren, X., Chen, P., Deng, L., An, K., Li, P., *et al.*: Multi-material decomposition of spectral ct images via fully convolutional densenets. Journal of X-Ray Science and Technology **27**(3), 461–471 (2019)

[39] Li, D., Li, S., Zhu, M., Gao, Q., Bian, Z., Huang, H., Zhang, S., Huang, J., Zeng, D., Ma, J.: Unsupervised data fidelity enhancement network for spectral ct reconstruction. In: Medical Imaging 2020: Physics of Medical Imaging, vol. 11312, pp. 1081–1088 (2020). SPIE

[40] Yao, L., Li, S., Li, D., Zhu, M., Gao, Q., Zhang, S., Bian, Z., Huang, J., Zeng, D., Ma, J.: Leveraging deep generative model for direct energy-resolving ct imaging via existing energy-integrating ct images. In: Medical Imaging 2020: Physics of Medical Imaging, vol. 11312, pp. 1190–1195 (2020). SPIE

[41] Shan, H., Padole, A., Homayounieh, F., Kruger, U., Khera, R.D., Nitiwarangkul, C., Kalra, M.K., Wang, G.: Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose ct image reconstruction. Nature Machine Intelligence **1**(6), 269–276 (2019)

[42] Flohr, T.G., McCollough, C.H., Bruder, H., Petersilka, M., Gruber, K., Süβ, C., Grasruck, M., Stierstorfer, K., Krauss, B., Raupach, R., *et al.*: First performance evaluation of a dual-source ct (dsct) system. European radiology **16**(2), 256–268 (2006)

[43] Primak, A., Ramirez Giraldo, J., Liu, X., Yu, L., McCollough, C.H.:

Improved dual-energy material discrimination for dual-source ct by means of additional spectral filtration. Medical physics **36**(4), 1359–1369 (2009)

[44] Genzel, M., Gühring, I., Macdonald, J., März, M.: Near-exact recovery for tomographic inverse problems via deep learning. In: International Conference on Machine Learning, pp. 7368–7381 (2022). PMLR

[45] Würfl, T., Hoffmann, M., Christlein, V., Breininger, K., Huang, Y., Unberath, M., Maier, A.K.: Deep learning computed tomography: Learning projection-domain weights from image domain in limited angle problems. IEEE transactions on medical imaging **37**(6), 1454–1463 (2018)

[46] Syben, C., Michen, M., Stimpel, B., Seitz, S., Ploner, S., Maier, A.K.: Pyro-nn: Python reconstruction operators in neural networks. Medical physics **46**(11), 5110–5115 (2019)

[47] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018)

[48] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

[49] Ma, G., Zhang, Y., Zhao, X., Wang, T., Li, H.: A neural network with encoded visible edge prior for limited-angle computed tomography reconstruction. Medical Physics **48**(10), 6464–6481 (2021)

[50] Reiser, I., Nishikawa, R.: Task-based assessment of breast tomosynthesis: effect of acquisition parameters and quantum noise a. Medical physics **37**(4), 1591–1600 (2010)