



HAL
open science

Estimating dual-energy CT imaging from single-energy CT data with material decomposition convolutional neural network

Tianling Lyu, Wei Zhao, Yinsu Zhu, Zhan Wu, Yikun Zhang, Yang Chen,
Limin Luo, Shuo Li, Lei Xing

► To cite this version:

Tianling Lyu, Wei Zhao, Yinsu Zhu, Zhan Wu, Yikun Zhang, et al.. Estimating dual-energy CT imaging from single-energy CT data with material decomposition convolutional neural network. *Medical Image Analysis*, 2021, 70, pp.102001. 10.1016/j.media.2021.102001 . hal-03190848

HAL Id: hal-03190848

<https://hal.science/hal-03190848>

Submitted on 31 May 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Estimating dual-energy CT imaging from single-energy CT data with material decomposition convolutional neural network

Tianling Lyu^{a,b,1}, Wei Zhao^{b,1}, Yinsu Zhu^c, Zhan Wu^a, Yikun Zhang^a, Yang Chen^{a,d,e,f,*}, Limin Luo^{a,d,e,f}, Shuo Li^g, Lei Xing^b

^aLaboratory of Image Science and Technology, Southeast University, Nanjing, Jiangsu, China

^bStanford Cancer Center, 875 Blake Wilbur Dr, Palo Alto, CA, US

^cNanjing Medical University, Nanjing, Jiangsu, China

^dSchool of Cyber Science and Engineering, Southeast University, Nanjing, Jiangsu, China

^eKey Laboratory of Computer Network and Information, Southeast University, Nanjing, China

^fCentre de Recherche en Information Biomedicale Sino-Francais (LIA CRIBs), Rennes, France

^gDepartment of Medical Imaging, Western University, London, ON N6A 3K7, Canada

Abstract

Dual-energy computed tomography (DECT) is of great significance for clinical practice due to its huge potential to provide material-specific information. However, DECT scanners are usually more expensive than standard single-energy CT (SECT) scanners and thus are less accessible to undeveloped regions. In this paper, we show that the energy-domain correlation and anatomical consistency between standard DECT images can be harnessed by a deep learning model to provide high-performance DECT imaging from fully-sampled low-energy data together with single-view high-energy data. We demonstrate the feasibility of the approach with two independent cohorts (the first cohort including contrast-enhanced DECT scans of 5,753 image slices from 22 patients and the second cohort including spectral CT scans without contrast injection of 2463 image slices from other 22 patients) and show its superior performance on DECT applications. The deep-learning-based approach could be useful to further significantly reduce the radiation dose of current premium DECT scanners and has the potential to simplify the hardware of DECT imaging systems and to enable DECT imaging using standard SECT scanners.

Keywords: Dual-energy CT, deep learning, material decomposition, convolutional neural network

1. Introduction

Material differentiation and quantification using a standard single-energy computed tomography (SECT) are extremely challenging because different materials may have the same CT value (McCollough et al., 2015). To tackle this challenge, dual-energy CT (DECT) takes full advantage of the energy dependence of the linear attenuation coefficient by scanning the patients using two different energy spectra (Alvarez and Macovski, 1976a, Kalender et al., 1986, Flohr et al., 2006, Johnson et al., 2007, Boll et al., 2008, Lee et al., 2017, Petrongolo and Zhu, 2018, Xue et al., 2019). This enables DECT imaging providing

*This work was supported by the State's Key Project of Research and Development Plan [grant numbers 2017YFA0104302, 2017YFC0109202 and 2017YFC0107900], by the National Natural Science Foundation [grant 81530060 and 61871117], by Science and Technology Program of Guangdong [grant number 2018B030333001], by NIH grant numbers 1R01CA223667 and R01CA227713] and a Faculty Research Award from Google Inc.

*Corresponding author

Email address: chenyang.list@seu.edu.cn (Yang Chen)

¹The authors contribute equally to this work.

energy- and material-selective images, and having been very widely used in clinical practice for many applications, such as virtual monochromatic imaging (Yu et al., 2012, Pomerantz et al., 2013), differentiating intracerebral hemorrhage from iodinated contrast (Phan et al., 2012), automated bone removal in CT angiography (Sommer et al., 2009, Buerke et al., 2009, Morhard et al., 2009, Schulz et al., 2012), virtual non-contrast-enhanced imaging (Takahashi et al., 2008, Ferda et al., 2009, Graser et al., 2009, Ho et al., 2012, Mangold et al., 2012, Toepker et al., 2012) and urinary stone characterization (Primak et al., 2007, Boll et al., 2009, Ascenti et al., 2010, Leng et al., 2015b). However, it is still an open and challenging task for clinical DECT imaging due to complex practical implementations, proprietary patents for major CT vendors, and lower popularity for DECT scanners compared to the standard SECT scanners.

Since the low- and high-energy CT images acquired from the DECT scanners have the same anatomical structures, there is substantial redundant anatomical information between the DECT images. For the scanned patients using the same DECT imaging protocols, the low- and high-energy CT images are also correlated in the energy-domain, resulting in information redundancies in the energy-domain (Leng et al., 2015a, Zhao et al., 2016). Meanwhile, both DECT images are reconstructed using fully-sampled projection data which have to meet the classical Shannon-Nyquist theorem in angular-data sampling to reconstruct artifacts-free images. By fully exploiting the anatomical consistency and energy-domain correlation between the DECT images, it is possible to provide high-quality artifacts-free DECT images using conventional SECT images together with sparse sampling projection data at different energy levels.

CT imaging with the full use of as low as reasonably achievable (ALARA) principle has been commonly accepted in routine practice and further reducing radiation dose from CT scanning is clinically favorable and has been extensively studied for almost two decades. Deep learning (DL) has recently been proved to be a powerful tool for mapping complex relationships and incorporating existing knowledge into an inference model through feature extraction and representation learning (Gulshan et al., 2016, Esteva et al., 2017, Ting et al., 2017, Liu et al., 2018, Xing et al., 2018, Zhao et al., 2019a, Maier et al., 2019, Shen et al., 2019, Lee et al., 2019, Zhao et al.,

2019b). It has also been applied in low-dose CT (Chen et al., 2017, Yang et al., 2018, Wolterink et al., 2017, Kang et al., 2018) and DECT imaging (Liao et al., 2018, Zhang et al., 2019, Poirot et al., 2019, Feng et al., 2018, Zhao et al., 2020).

To reduce the radiation dose of DECT imaging, in this study, we synergically exploit the energy-domain correlation and anatomical consistency between DECT images by leveraging the deep learning approach and the seamless integration of the correlation and consistency in a data-driven DECT imaging process, and eventually push the sparse sampling to the limit of a single projection view and demonstrate the feasibility of high-performance DECT imaging using a deep learning approach termed fully low-energy and single high-energy DECT imaging (FLESH-DECT).

2. Material and methods

The flowchart of the proposed FLESH-DECT strategy is shown in Fig. 1. The input to the model is a single-view high-energy projection together with the low-energy image I_{low} which is reconstructed using fully-sampled low-energy projection data. For the low-energy image, it is firstly denoised with a denoising network to mitigate the impact of image noise. Instead of training a network directly mapping high-energy images from the low-energy images, we use a convolutional neural network (CNN) to perform material-decomposition-type operations and it is termed material decomposition CNN (MD-CNN). The input of the MD-CNN is the denoised low-energy image I_{low}^{de} while the output is a "material component" matrix A . The matrix A has the same image size as I_{low}^{de} but with multiple channels each of which corresponds to a pseudo-material-specific image. The values are the percentages of corresponding "basis material" on these pixels. We ensure that the sum of the percentages equals one (mass conversation) for each unique pixel. Furthermore, another CNN is used to pre-process the differences between the given high-energy projections and their corresponding low-energy projections. This projection-domain network is used to fill the gap between the denoised images and the non-denoised projection. Since CT forward projection can be regarded as linear summations of pixel values, we can use the least-squares method to solve the

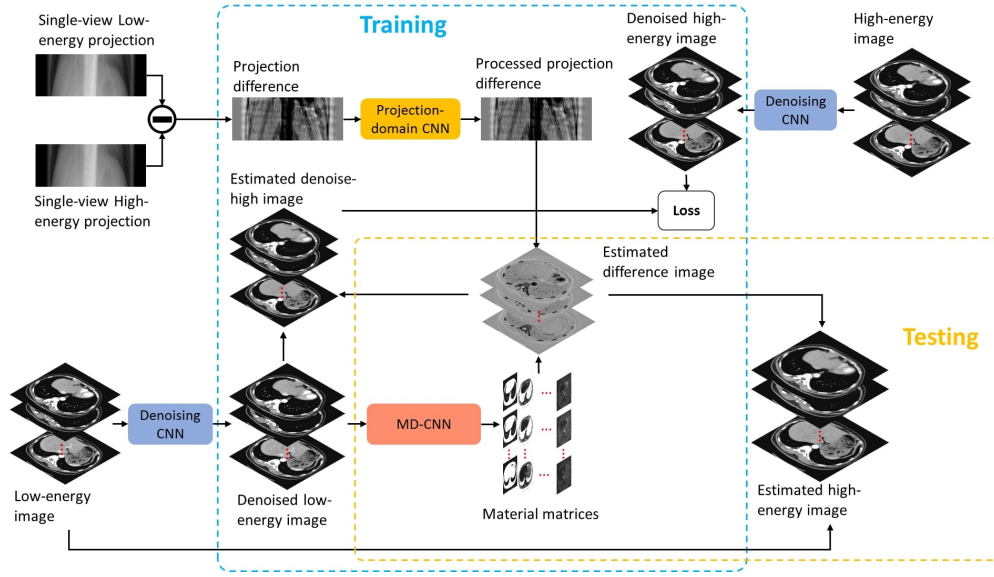


Figure 1: The workflow of the proposed fully-sampled low-energy and single-view high-energy DECT imaging approach. During the training phase, the denoised DECT images together with the single-view dual-energy projections are used to train the projection domain convolutional neural network (CNN) and the MD-CNN. In the testing phase, the trained networks use the input low-energy images and the single-view dual-energy projections to infer the corresponding high-energy images.

corresponding CT values of each "material" b_{dif} according to the matrix A and the pre-processed projection difference. The estimated high-energy image is calculated as the summation of low-energy images, and the inner product of matrix A and vector b_{dif} . Detailed formula derivation is described in the following subsections.

2.1. DECT Imaging

In CT imaging, the attenuation coefficient at each position can be represented as a linear combination of basis materials' attenuation coefficient (Alvarez and Macovski, 1976b).

$$\mu = \alpha_1\mu_1 + \alpha_2\mu_2 + \dots + \alpha_m\mu_m \quad (1)$$

where m is the number of basis materials, α_i is the percentage of the i -th basis material and μ_i is the attenuation coefficient of the i -th basis material. Since CT values in Hounsfield Unit (HU) can be represented as the linear transformation of attenuation coefficient μ with the following equation:

$$HU = 1000 \times \frac{\mu - \mu_{water}}{\mu_{water}} \quad (2)$$

Eq.(1) can also be written as:

$$HU = \alpha_1HU_1 + \alpha_2HU_2 + \dots + \alpha_mHU_m \quad (3)$$

where HU_i stands for the Hounsfield Unit CT value for the i -th basis material. Considering there are $n_{pix} = W \times H$ pixels in an image slice, we can write Eq.(3) into the following matrix multiplication form

$$I = A \cdot b^T \quad (4)$$

where I is the image vector sized $n_{pix} \times 1$ containing CT values at each pixel, $A = [\alpha_{ij}]_{n_{pix} \times m}$ is the material component matrix, α_{ij} stands for the percentage of material j at pixel i , $b = [HU_i]_{1 \times m}$ consists of HU values for each material.

In DECT, there are two different images I_{low} and I_{high} . The material component matrix A remains the same for both images because pixel compositions do not change between low- and high-energy scans. Therefore, we have the following equations

$$\begin{cases} I_{low} = A \cdot b_{low}^T \\ I_{high} = A \cdot b_{high}^T \end{cases} \quad (5)$$

By subtracting the high-energy equation from the low-energy equation, we get

$$I_{dif} = A \cdot b_{dif}^T \quad (6)$$

where I_{dif} is the difference image between I_{low} and I_{high} , b_{dif} is the difference between b_{low} and b_{high} . Let P_{high} and P_{low} be the given high- and low- energy projection measurements, and R be the projection matrix sized $n_{ray} \times n_{pix}$ corresponding to the high-energy view. We have the following equation

$$\begin{aligned} R \cdot A \cdot b_{dif}^T &= R \cdot I_{dif} \\ &= R \cdot I_{high} - R \cdot I_{low} \\ &= P_{high} - P_{low} \end{aligned} \quad (7)$$

For the CT imaging task with fully-sampled low-energy and single-view high-energy projection, the unknowns in Eq.(7) are the material component matrix A and the corresponding difference values b_{dif} . Assuming the material component matrix A is known, let $M = R \cdot A \in \mathfrak{R}^{n_{ray} \times m}$, $P_{dif} = P_{high} - P_{low} \in \mathfrak{R}^{n_{ray} \times 1}$, the difference values b_{dif} can be calculated by solving the equation $M \cdot b^T = P_{dif}$. In regular CT imaging, we have $n_{ray} \gg m$, the best b_{dif} can therefore be found using least-squares method which can be computed with Cholesky decomposition, i.e.,

$$\begin{aligned} b_{dif} &= \underset{\tilde{b}}{\operatorname{argmin}} \|M \cdot \tilde{b}^T - P_{dif}\|_2^2 \\ &= [(M^T M)^{-1} M^T P_{dif}]^T \end{aligned} \quad (8)$$

The only task now is to estimate the material component matrix A from the low energy image I_{low} .

2.2. Material decomposition-based dual-energy CT mapping

Due to its ability to learn complex relationships and incorporate existing knowledge into a nonlinear mapping model, a dedicated CNN model (termed MD-CNN) is used to estimate the material component matrix A . When designing the MD-CNN model, a major challenge is the lack of training labels due to unknown materials in the images and their percentages. To tackle this challenge, we train the MD-CNN indirectly. We firstly denoise the dual-energy image pairs, and the denoised low-energy images I_{low}^{de} are inputted into the MD-CNN to acquire material

component matrices A_{DL} . Meanwhile, we put the projection differences into another 1-D projection domain CNN to preprocess the projection data. Then, we compute b_{dif} using Eq.(8). The estimated denoised high-energy images I_{high}^{dl} can therefore be calculated as

$$I_{high}^{dl} = I_{low}^{de} + A_{DL} \cdot b_{dif}^T \quad (9)$$

An image similarity loss is calculated between the denoised high-energy image I_{high}^{de} and the DL-estimated image I_{DL}^{de} , and the mean squared error (MSE) loss is used for the task:

$$\mathcal{L}_{high} = \frac{1}{n} \|I_{high}^{de} - I_{DL}^{de}\|_2^2 \quad (10)$$

Instead of getting material component labels, we focus on the target high-energy image and it is not necessary to specify each channel in A_{DL} to represent real material or a linear combination of different materials. Meanwhile, since matrix A_{DL} is supposed to be the material component matrix, it should be able to recover the input denoised low-energy image as well, resulting in the following loss function:

$$\mathcal{L}_{low} = \frac{1}{n} \|I_{low}^{de} - A_{DL} \cdot b_{low}^T\|_2^2 \quad (11)$$

The same strategy in Eq.(8) is used to calculate the HU values b_{low} for each "material" under low energy,

$$b_{low} = \underset{\tilde{b}}{\operatorname{argmin}} \|A_{DL} \cdot \tilde{b}^T - I_{low}^{de}\|_2^2 \quad (12)$$

The final loss function is computed as the summation of \mathcal{L}_{low} and \mathcal{L}_{high} , i.e.,

$$\mathcal{L} = \mathcal{L}_{low} + \mathcal{L}_{high} \quad (13)$$

During the inference phase, the low-energy images are also firstly denoised and then inputted into the trained MD-CNN for A_{DL} . The projection domain CNN preprocesses the projection difference P_{dif} . The estimated difference images are calculated according to Eq.(8). The difference between inference and training is that the estimated high-energy image is calculated as the summation of the original low-energy image I_{low} and the difference image I_{dif} in the inference phase.

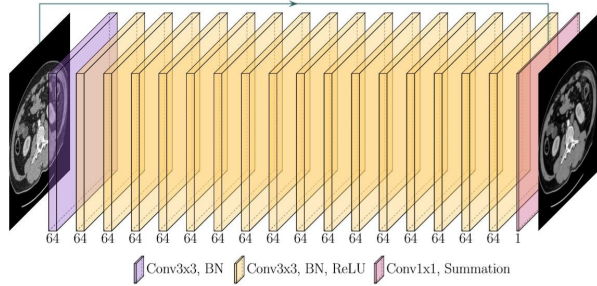


Figure 2: The architecture of the fully convolutional network for image denoising. A plain structure that encompasses 13 convolution layers is applied to learn the residual between the input image and the denoised image.

2.3. Network details

2.3.1. Denoising CNN

We employed the denoising network in our previous work (Zhao et al., 2020) to reduce the DECT image noise. The network uses a plain structure that encompasses 13 convolution layers to learn the residual between the input image and the denoised image. The first 12 layers are convolution layers with kernel size 3×3 . Each convolution layer is followed by a batch normalization layer (BN) and a rectified linear unit (ReLU) activation. The last layer is a convolution layer with kernel size 1×1 fusing the result. Fig. 2 shows the detailed structure of the denoising CNN. The denoised image is computed as the summation of the input image and the output from the last layer.

2.3.2. MD-CNN

For the material decomposition network, we employ a U-Net-type structure (Ronneberger et al., 2015) which has a large receptive field and is quite suitable for many medical image processing tasks. There are 10 normal 3×3 convolution layers in the proposed network (Fig. 3). Each convolution layer is followed by a BN layer and a ReLU activation layer. There are 3 resolution levels in total. For down-sampling, we use convolution layers with kernel size of 2×2 and stride equal to 2. Each strided convolution layer is also followed by a BN layer and a ReLU activation layer. For up-sampling, we use bilinear interpolation to double both image width and height. At the end of the network, a convolution layer with kernel size of 1×1 is added to fuse the channels. Since the values on

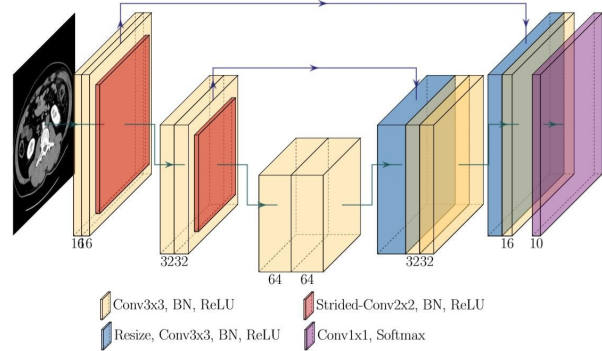


Figure 3: Structure of the MD-CNN. A much simplified UNet-like structure with 14 convolution layers is used here to estimate the percentages of each corresponding "basis material" from the denoised low-energy image. Numbers under each block show the channel number of the multichannel feature maps.

each output channel are supposed to be the percentages of corresponding "basis material", we apply softmax after the convolution layer with kernel size of 1×1 to make sure that the sum of materials' proportion equals to 1 at each pixel.

2.3.3. Projection-domain CNN

To enhance the robustness of the least-squares problem in Eq.(8), the projection-domain network (Fig. 4) is applied to slightly refine the inputting projection difference and to make its noise level to be consistent with that of the denoised DECT difference image, and we employ a concise 4-layer network for the task. A residual learning structure is used here for an easier startup at the beginning iterations. All basic convolution layers have a kernel with a size of 1×5 except for the last one which has a kernel with a size of 1×1 . The first three convolution layers are followed by a BN layer and a ReLU activation layer.

2.3.4. Network training

The denoising CNN was implemented in MATLAB with MatConvNet framework (Vedaldi and Lenc, 2015). Denoising was performed as a data preprocessing step for all DECT images. The material decomposition network and projection-domain network are implemented using Python with Tensorflow framework (Abadi et al., 2016). Those two networks were trained together in an end-to-

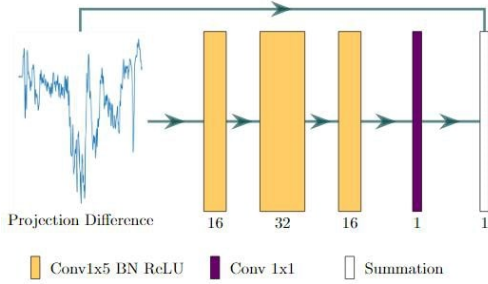


Figure 4: Structure of the projection-domain CNN. A concise 4-layer network is employed to slightly refine the inputting projection difference. Numbers under each block show the number of channels of the multichannel feature maps.

end fashion. The parameters in the networks were optimized using ADAM algorithm (Kingma and Ba, 2014) with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate was set to 10^{-3} during the training. The training set was randomly split into small batches in each epoch with a batch size of 8. The proposed network was trained for 200 epochs in total. We perform validation after each training epoch and the model with the best validation loss was selected as the final model for testing. The number of materials m was set to 10 in our experiments according to the loss on the validation set. We trained models with different m settings from 2 to 20 and the best validation loss came with $m = 10$. It has to note that the model was able to achieve reasonable results even with $m = 3$. As shown in Fig.5, the validation losses are close when $m > 2$. The networks were trained and tested on a workstation with configurations as follows: CPU is Intel(R) Xeon(R) Gold 6130 CPU @ 2.10GHz; GPU is NVIDIA RTX 2080 Ti with 12 GB memory.

2.4. Projectors for different CT geometries

To calculate matrix M in Eq.(8), a projector corresponding to the CT geometry is indispensable in our algorithm. There are mainly two types of 2D CT geometries, fan-beam and parallel-beam. The fan-beam geometry can be further divided into two sub-types, equiangular and equispaced. For each above-mentioned geometry, we developed a projector and trained a new model for evaluation.

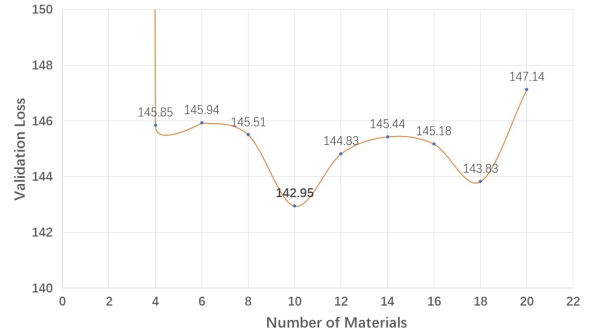


Figure 5: Validation loss under different m settings. Best validation loss comes with $m = 10$.

2.4.1. Equiangular fan-beam

Equiangular geometry is mainly implemented with arc detectors which keep the angles between two adjacent detector pixels and the source-detector-distance the same. We tested our model in the anterior-posterior (AP) direction, but it can be easily extended to any other projection view. To acquire projection data using the 2D equiangular fan-beam geometry, we first calculate the intersection of each projection ray with each image row. The values at each intersection point are computed using linear interpolation. Suppose the image size is $W \times H$, and a matrix I' with a size of $n_{ray} \times H$ can be obtained using the following rebinning equation:

$$I'(\phi, y) = I((D - y) \tan(\phi), y) \quad (14)$$

where ϕ is the angle between the projection ray and the central ray (as shown in Fig. 6), and y is the image pixel position along the y-axis in the Cartesian coordinate system centered at the image center O . D is the distance between the projection source and the rotation center which is overlapped with the image center O . After the linear interpolation, we calculate the summation of each column in matrix I' to obtain ray sum S which is weighted by the distance to yield the final projection P :

$$P(\phi) = \frac{dy}{\cos \phi} S(\phi) \quad (15)$$

where dy is the image spacing in y-axis. We set $D = 600mm$, $dy = 0.5mm$, the number of detector channels $n_{ray} = 800$ and the angle between adjacent channels $ds = \frac{7}{11000} rad$ for all images.

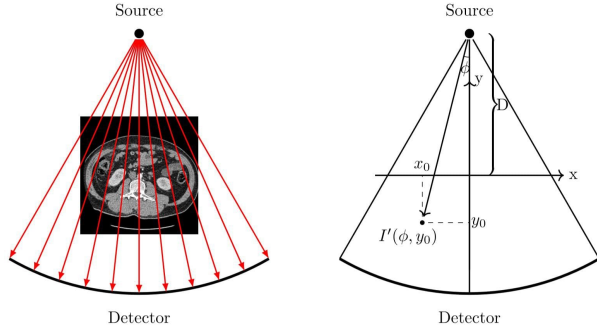


Figure 6: Illustration of the equiangular fan-beam geometry.

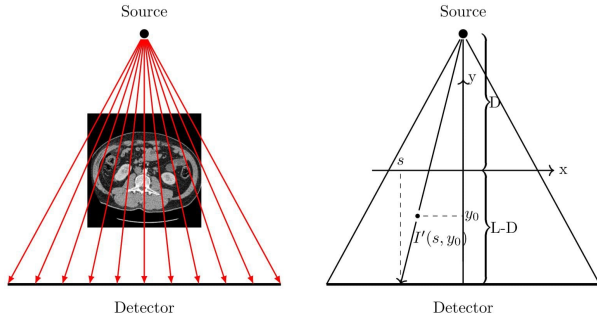


Figure 7: Illustration of the equispaced fan-beam geometry.

2.4.2. equispaced fan-beam

Flat detectors with equal spacing between the adjacent channels are commonly used to implement the equispaced geometry. We used a similar strategy to implement the equispaced fan-beam projection in the AP direction. In this case, the rebinning procedure is computed with the following equation:

$$I'(s, y) = I\left(\frac{s(D-y)}{L}, y\right) \quad (16)$$

where s is the distance between the current channel and the detector center (direction included), and L is the source to detector distance, as shown in Fig. 7. The rest part of the implementation is the same as the equiangular projection. For the parameters, we have $D = 600\text{mm}$, $L = 1100\text{mm}$, $dy = 0.5\text{mm}$, $n_{ray} = 800$ and the spacing between nearby channels $ds = 0.78\text{mm}$.

2.4.3. Parallel-beam

For the parallel-beam scenario, We assumed that the detector channel has the same spacing as the image pixels, and each X-ray projects exactly through an image column in the AP direction. Therefore, the projection in the AP direction can be computed as the summation of each image column.

2.5. Dataset

2.5.1. Training data for the denoising network

The AAPM Low-Dose CT Grand Challenge data was used to train the denoising network. This dataset consists of routine dose CT and the corresponding simulated low-dose CT data from 10 patients. The routine dose scanning voltage is 100 kV or 120 kV and the X-ray tube current varies from 200mA to 500mA. The detector has 736×64 elements, and each element has a size of $1.2856 \times 1.0947\text{mm}^2$. The source-to-axial distance is 59.5cm and the source-to-detector distance was 108.56cm . All the images were reconstructed to slice thickness of 1.0mm and 512×512 pixel. The pixel size varies from $0.66 \times 0.66\text{mm}^2$ to $0.78 \times 0.78\text{mm}^2$. To simulate the low-dose CT data, Poisson noise was introduced into the routine dose to mimic a noise level that corresponded to 25% of the routine dose, and noisy projection was reconstructed to yield the low-dose image.

2.5.2. Siemens DECT image dataset

Clinical DECT images of 22 patients who underwent iodine contrast-enhanced DECT exams were collected for the study. All the exams were performed in Nanjing General PLA Hospital, China, with the approval of the institutional review board and patient consent forms. The DECT images (5753 slices in total) were acquired using a SOMATOM Definition Flash DECT scanner (Siemens Healthineers, Forchheim, Germany) after administering iodine contrast agent. The low- and high-energy of the DECT scans were 100 kV and 140 kV, respectively. All CT images were reconstructed using the filtered back-projection (FBP) algorithm provided by the commercial CT vendor. The dataset was split into a training set, a validation set and a testing set randomly with 16, 3, and 3 patients included respectively.

2.5.3. Philips spectral CT image dataset

We collected clinical spectral CT images of 22 patients with Philips IQon Spectral CT devices. All exams were performed in Nanjing Gulou Hospital, China, with the approval of the institutional review board and patient consent forms. Philips IQon Spectral CT applies spectral detector technology to implement dual-energy scans. Virtual mono-energetic images at different X-ray energy levels can be reconstructed from the data collected. In our experiments, images at 70 keV, 110 keV, and 150 keV were used to evaluate the proposed method. Non-contrast-enhanced brain virtual mono-energetic images (2463 slices in total) were collected for each energy level and all images were reconstructed using a build-in algorithm provided by the CT vendor. Similar to the Siemens dataset, the Philips dataset was also split into a training set, a validation set and a testing set randomly with 16, 3 and 3 patients included respectively.

3. Results

We firstly focus on the results of equiangular geometry on the Siemens Dataset, and comparison results using different geometries are presented in the middle of this section. Results on the Philips dataset are described at last.

Fig. 8 shows the original DECT images and the 140 kV images predicted using the proposed method for a testing patient. The first, second and third columns show the original 100 kV images, the original 140 kV images, and the predicted results, respectively. The first, third, and fifth rows show CT images in transverse, sagittal, and coronal planes, respectively. The second, fourth, and sixth rows show difference images with respect to the corresponding real high-energy CT images in transverse, sagittal, and coronal planes, respectively. As can be seen, the proposed DL-derived high-energy images are highly consistent with the original high-energy images. There are some differences at sharp boundaries which also appear in the difference images between original high- and low-images. Those differences may be motion introduced differences between original DECT images because there are approximately 90 degrees out of phase for the low- and high-energy data acquired using a dual-source DECT scanner. When inputting the low-energy image into the model,

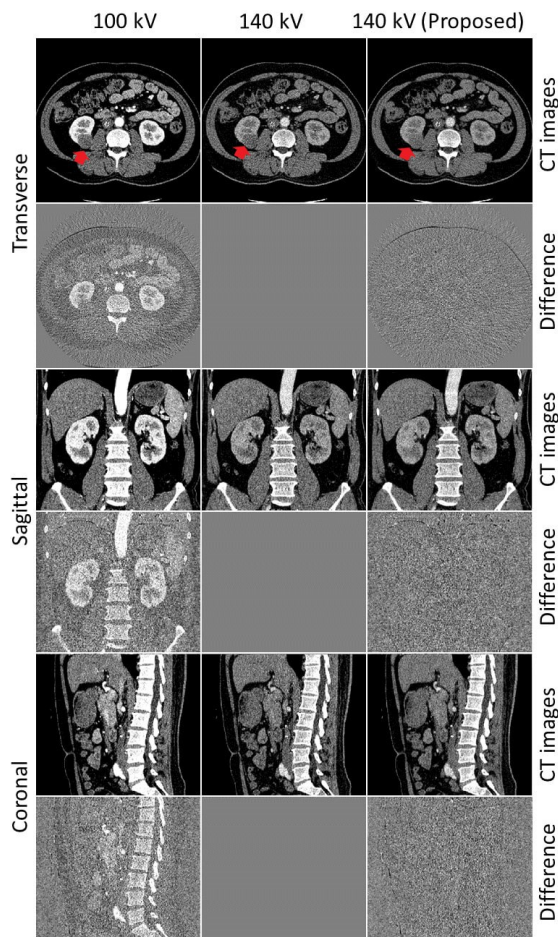


Figure 8: Example results on a testing slice and the difference with respect to the corresponding real high-energy image. The first, third and fifth rows display the images on axial, sagittal and coronal view. From left to right are real 100 kV image, real 140 kV image and result from the proposed method, respectively. A kidney lesion in transverse image is labeled using red arrow. The second, fourth and sixth rows are the corresponding differences with real 140 kV images. The CT images are displayed with a window width=300 HU and center=50 HU while the difference images are displayed under window width=300 HU and center=0 HU.

the model performed prediction based on the anatomical structure of the low image and can not reflect the change with respect to the original high-energy image.

Quantitative metrics were calculated to evaluate the accuracy of the predicted high-energy CT images. We used the well-established metrics MSE, PSNR and SSIM to as-

Table 1: Quantitative comparisons between the predicted and the real high-energy CT images for the testing patients.

	MSE	PSNR	SSIM	HU error	Time(s)
Patient1	875.06	36.80	0.8702	1.65±1.03	2.65
Patient2	887.60	36.99	0.8789	2.09±1.48	2.31
Patient3	927.47	36.89	0.8745	1.50±0.82	3.34

assess the image similarity between the real and the predicted high-energy images. Additionally, more than 100 region-of-interest (ROIs) were randomly selected for each testing volume on homogeneous areas (e.g. liver and stomach). We calculated the mean HU value differences on those ROIs and the results show that the averaged HU error between the predicted and the original high-energy image is smaller than 2.09 HU. For the computation time, it took around 2.5 seconds for the proposed method to process 300 slices on our server. All quantitative results are shown in Table 1. Since most of the differences come from noise, we also compared the denoised DL-predicted images with the denoised high-energy images. In this case, the DL-predicted images were calculated by adding the denoised low-energy image to the DL-estimated difference image. For those denoised images, the proposed method achieved an average MSE of 171.09, PSNR of 44.33 and SSIM of 0.9848.

In our previous work (Zhao et al., 2020), we proposed the DL-DECT method to estimate I_{dif} directly from I_{low} . In this work, with the introduction of the additional high-energy single-view projection, we can further enhance the accuracy of the predicted high-energy image and eventually the accuracy of the material- and energy-specific images. To show the benefit of the additional high-energy projection, we quantitatively compared the proposed FLESH-DECT method with the DL-DECT method. Virtual non-contrast (VNC) images and iodine maps were derived to demonstrate the clinical utility of the proposed method. Fig. 9 depicts the VNC images and iodine maps reconstructed using different methods on the transverse, sagittal and coronal planes. Both the DL-DECT and FLESH-DECT algorithms provided high-quality VNC and iodine images that were consistent with the images generated by original DECT images. Quan-

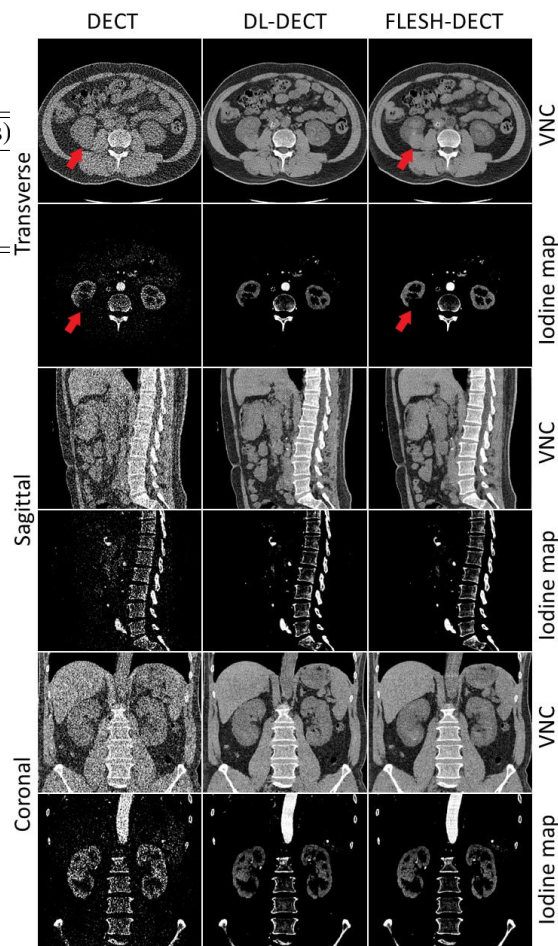


Figure 9: VNC images and Iodine maps reconstructed using original DECT, DL-DECT, and FLESH-DECT images. The first, third and fifth row are the VNC images on axial, sagittal and coronal view, respectively, while the second, fourth and sixth row are the corresponding Iodine maps. The kidney lesion in the transverse images is marked out with red arrows for better comparison.

titative metrics on VNC and iodine images are shown in Table 2. The results demonstrate FLESH-DECT can provide high-quality material-specific images and it outperforms the DL-DECT method.

From the VNC images and Iodine maps, we found that the DL-derived VNC images and iodine maps show a remarkably reduced noise level compared with those generated from original DECT images. Here we also compared the noise level by calculating the standard deviation

Table 2: Quantitative comparisons of the VNC image and the iodine maps reconstructed using original DECT and FLESH-DECT images.

		VNC		Iodine Map	
		MSE	PSNR	MSE	PSNR
Patient1	DL-DECT	3882.34	29.03	0.0313	23.65
	FLESH-DECT	3808.97	29.11	0.0308	23.73
Patient2	DL-DECT	4216.36	28.95	0.0342	23.78
	FLESH-DECT	4128.01	29.04	0.0335	23.87
Patient3	DL-DECT	3508.35	29.71	0.0313	24.85
	FLESH-DECT	3431.13	29.81	0.0306	24.95

Table 3: Quantitative Noise Level Comparison on VNCs and Iodine Maps.

Standard Deviation on ROIs		VNC	Iodine Map
Patient1	Real	81.58	0.3363
	DL-DECT	29.14	0.1618
	FLESH-DECT	28.09	0.1499
Patient2	Real	74.69	0.2837
	DL-DECT	24.81	0.1017
	FLESH-DECT	24.12	0.0998
Patient3	Real	76.95	0.2875
	DL-DECT	25.97	0.0946
	FLESH-DECT	25.67	0.0891

in ROIs. More than 500 ROIs were selected randomly within homogeneous areas. The mean standard deviations in ROIs on each testing patient are provided and compared in Table 3. The mean standard deviations of the images derived using FLESH-DECT are close to that of the DL-DECT method which is much lower than those from the original DECT images. This result shows that the FLESH-DECT method maintains the denoising feature.

We tested the proposed method with several 2D CT geometries. Example results on a testing volume using different geometries are displayed in Fig. 10. All CT images are displayed with window width=300HU and center=50HU while difference images are displayed with window width=300HU and center=0HU. As can be seen, all models can provide competitive results that are highly consistent with the original high-energy image. For better comparison, we also calculated quantitative metrics on

those results (Table 4). The differences between the results obtained using the proposed method with different geometries are marginal and all of them are superior to our previous DL-DECT method which does not utilize the additional single-view projection. Overall, the proposed method reduces mean-squared error (averaged for all testing cases) from 1858.32 to 898.35 while increasing PSNR from 33.83 to 36.89 and SSIM from 0.8641 to 0.8744.

Table 4: Quantitative Results for Different Geometries.

		MSE	PSNR	SSIM
Patient1	DL-DECT	2466.07	32.37	0.8558
	Parallel	866.48	36.84	0.8707
	Equiangular	875.06	36.80	0.8702
	equispaced	871.65	36.82	0.8703
Patient2	DL-DECT	1530.75	34.61	0.8723
	Parallel	876.68	37.03	0.8793
	Equiangular	887.60	36.99	0.8789
	equispaced	882.76	37.01	0.8789
Patient3	DL-DECT	1578.13	34.61	0.8660
	Parallel	927.47	36.89	0.8745
	Equiangular	932.40	36.88	0.8741
	equispaced	927.15	36.90	0.8747

The computation time for different methods are displayed and compared in Table 5. Note that computation time for the denoising CNN is not included in the results. The denoising network takes about 0.01 seconds for each slice and the denoising time is the same for all methods. Compared to the DL-DECT method, the proposed method speeds-up the computation time by 2-fold which can be

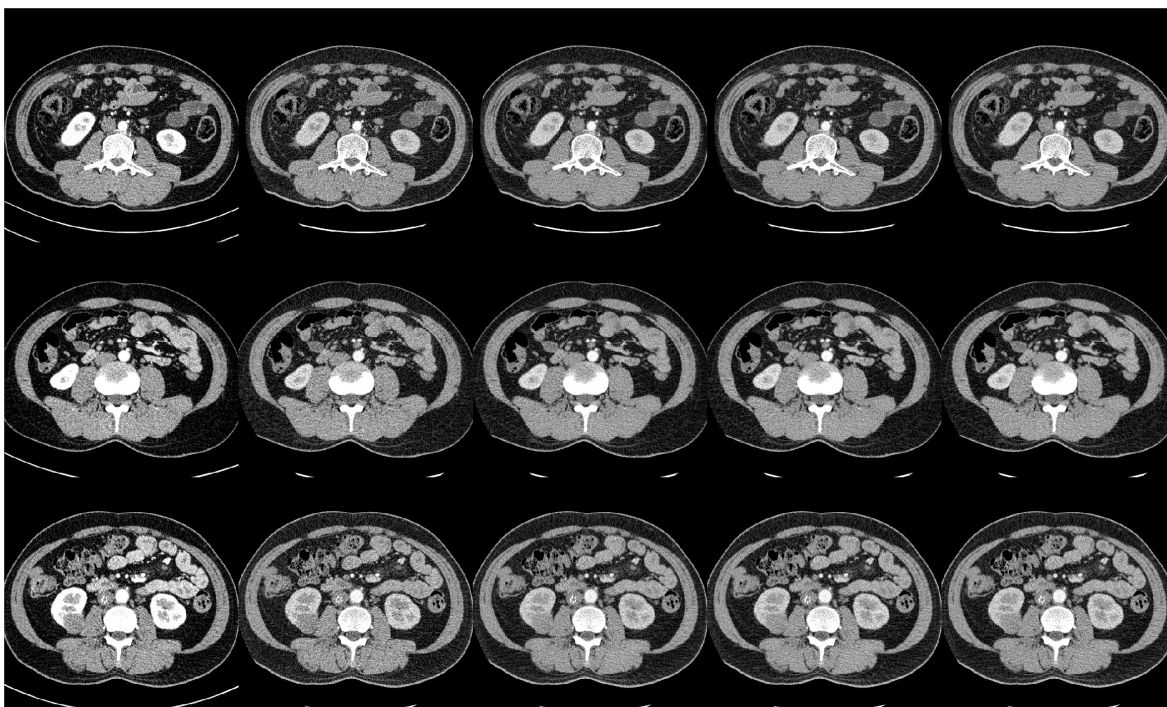


Figure 10: Results on three testing slices for different geometry models. From left to right are original 100 kV images, original 140 kV images, parallel-beam results, equispaced fan-beam results and equiangular fan-beam results, respectively.

Table 5: Computation Time for Different Methods in Seconds.

	Patient1	Patient2	Patient3
Slices	308	265	387
DL-DECT	5.43	4.68	6.86
Parallel	2.33	2.04	2.94
Equiangular	2.65	2.31	3.34
equispaced	2.53	2.18	3.23

attributed to the reduced number of weights and the simple network structure. There are some differences among the time using different geometries which means the computational cost of the proposed method depends on the projector.

To further examine the proposed method, cross-validation on the Siemens dataset was also derived. Seven-fold cross-validation was used with 3 testing cases

for each fold (except for the last one with 4 cases). The results are displayed in Table 6. In cross-validation, the average mean-squared-error value of all 22 cases achieves 823.63 while average PSNR achieves 37.36 and average SSIM achieves 0.8862. All those quantitative metrics are close to the results on our testing set which shows the robustness of our proposed method. MSE values in each cross-validation group are also analyzed and the results are depicted in Fig.11. For each group, the average MSE value lies between 700 and 1000.

Table 6: Quantitative Results During Cross-validation.

MSE	823.62±141.00
PSNR	37.36±0.7227
SSIM	0.8862±0.0184

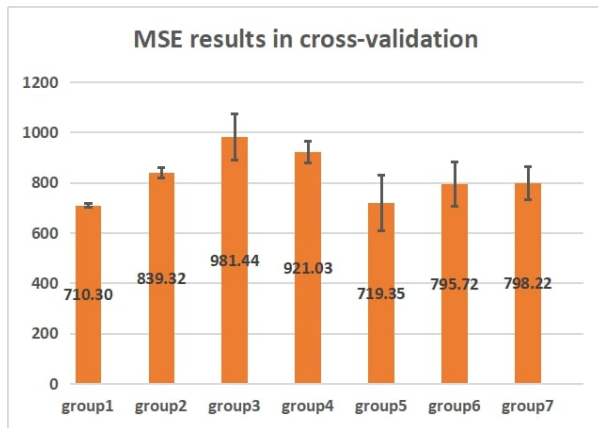


Figure 11: Mean MSE values (standard deviation) in each cross-validation group.

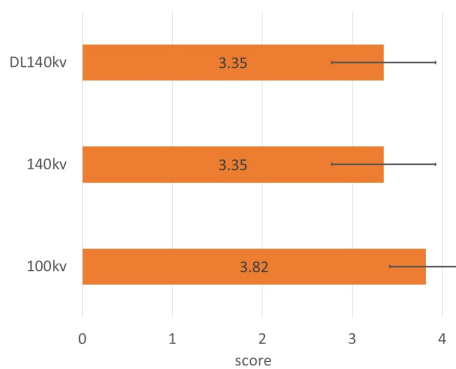


Figure 12: Clinical evaluation results.

Clinical doctors were invited to evaluate the artificially generated images. Three experienced radiologists from Jiangsu People's Hospital were involved in this process. The radiologists were supposed to provide a score from 1 to 5 (1 for useless, 5 for high clinical use) to real 100 kV, 140 kV images from practical CT machine as well as artificial 140 kV images generated using the proposed method. Images of all 22 patients (together with all generated results in cross-validation) were evaluated. Fig.12 shows the clinical evaluation results. The average score of the artificially generated 140 kV images is the same as that of real 140 kV images because the generated images look almost the same as the real ones.

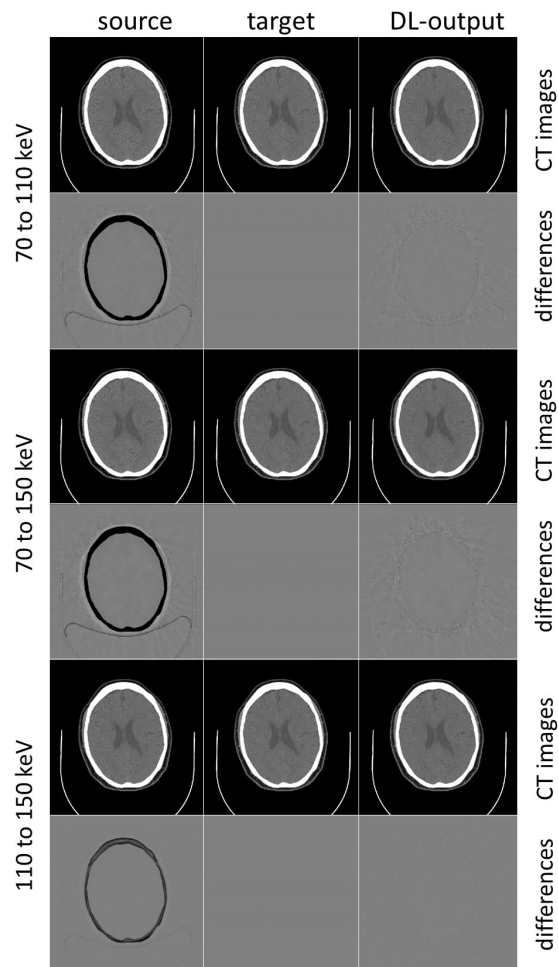


Figure 13: Example results on a Philips dataset testing slice and the difference with respect to the corresponding high-energy image. The CT images are displayed with a window width=300 HU and center=50 HU while the difference images are displayed under window width=300 HU and center=0 HU.

We also tested the proposed method on the Philips spectral CT image dataset. One independent model was trained between each energy level pair (e.g. between 70 keV and 110keV). Therefore, we got 3 models for the Philips dataset in total. For each model, 1828 slices were used as training data while 299 slices were used for validation and 336 for testing. All the hyper-parameters were set the same as those selected on the Siemens dataset. One difference that should be noticed is that we did not per-

form image denoising on Philips dataset because of its low noise level. The results are concluded in Table 7. For the task of estimating 150 keV images from 70 keV images, the average MSE between our results and the target images is only 13.45. Average PSNR value and SSIM value achieve 55.84 and 0.9985, respectively. The models for the other two tasks (estimating 110 keV images from 70keV images and 150 keV from 110 keV) perform even better on those metrics since the differences between real low- and high-energy images are smaller. Low noise level plays an important roll in the improvement in quantitative measurements. Besides, anatomical structures in non-enhanced brain images are much simpler than those in abdominal images. Results on one example slice are displayed in Fig.13. Brain tissue has similar CT values under different X-ray energy levels and the differences mainly come from the bone area. The proposed method performs well on the new Philips dataset acquired using a different DECT scanning protocol.

4. Discussion

There are substantial redundant information and correlation in both anatomical structure and energy-domain between the low- and high-energy DECT images. By incorporating the redundancy and the correlation into a deep learning model, it is possible to provide material- and energy-specific images using standard SECT scanners, which has the potential to alleviate the need for premium DECT scanners. Besides, compared to the standard fully low- and high-energy sampling DECT mechanism, the use of sparse sampling at the second energy level can significantly reduce the radiation dose of DECT imaging.

Our results show both the DL-DECT and FLESH-DECT methods achieve high-performance DECT imaging by using the input low-energy CT data, and quantitative analysis shows FLESH-DECT outperforms DL-DECT in terms of HU accuracy and calculation speed. The superior performance of the FLESH-DECT can be attributed to the additional single-view high-energy projection. Different from the DL-DECT method which directly infers a high-energy image using the incorporated prior knowledge, the proposed FLESH-DECT method uses the learned knowledge to fit the measured high-energy projection. Namely, the high-energy projection introduces a

penalty to constrain the projection generated by the predicted high-energy images to be consistent with the measurement, which in turn enhances the accuracy of the predicted images. The single-view high-energy projection can be obtained shortly before or after the standard SECT low-energy data acquisition, and existing SECT systems are able to implement these scanning protocols without modifying the hardware.

In the proposed method, the single-view high-energy projection and the corresponding low-energy projection view are supposed to be strictly aligned. We moved the X-ray source under high-energy to mimic patient movements and mismatches between projection source positions. The results are depicted in Fig.14. The model is robust to shift in axial direction to some extent. It outperforms DL-DECT model even when there is a 10mm-shift. The model is also robust concerning the movements in horizontal directions (projection source moving left or right). Changes in distance between patient and projection source are not accepted as they introduce severe mismatch on organ boundaries in projections. The proposed method is unlikely to introduce artifacts when there is mild mismatch between projections at different energy-levels. The material maps (in which structural information is included) are generated according to the given low-energy image. The anatomical structure will be well preserved when there is no artifact in the low-energy image. The projection difference only affects the CT values of each "material".

It should be noticed that the high-energy projection in the proposed method does not need to cover the whole area the corresponding low-energy projection scanned. Since in most cases, the number of equations in Eq.7 overwhelms the number of unknowns, reducing the number of equations (which means reducing the number of projections) has little influence on the results. From our experiments, MSE between the real image and the generated image will not change severely if we cut the projection channel number into half or even quarter, though full projections do perform slightly better (see Fig.15).

FLESH-DECT is suitable for different geometries. In this study, we have tested the method using 2D geometries (fan-beam and parallel beam). However, the method can be applied straightforwardly to 3D geometries by extending the networks and the projection operators into 3D scenarios. Since the reconstructed size is smaller than the

Table 7: Quantitative Results on Philips Dataset.

Source(keV)	Target(keV)	MSE	PSNR	SSIM
70	110	13.45 ± 1.94	55.84 ± 0.64	$0.9985 \pm 1.43 \times 10^{-4}$
70	150	22.27 ± 3.46	53.06 ± 0.66	$0.9980 \pm 2.65 \times 10^{-4}$
110	150	1.64 ± 0.21	64.22 ± 0.60	$0.9999 \pm 1.47 \times 10^{-5}$

projection view size in the rotation axis for a 3D case, projections generated using the reconstructed image volume cannot match the measured projection. To solve this issue, one can use the middle part of the measured projection which does not propagate through the region beyond the image volume.

The proposed method employs the MD-CNN to generate “material-decomposition” maps A from 100 kV images. Since basis materials in the image do not change under X-ray at different energy levels, accurate material-decomposition maps can provide CT images under any spectrum if combined with the projection view acquired using the corresponding spectrum (e.g. 120 kV image from 120 kV projection, 150 kV image from 150 kV projection). In our experiments, the models were trained and tested using DECT images under 100 kV/Sn 140 kV scanning protocol. Therefore, the generated “material-decomposition” maps are likely to be optimized for this specific protocol and may not be applied to images acquired using a different protocol (such as 80 kV/Sn 140 kV). An example is shown in Fig. 16. However, if the model was trained using images acquired from several different spectra, the “material-decomposition” maps would be much closer to the real ones and the model would have the potential to generate images under different spectra without re-training the MD-CNN. Also, regularizers may further be introduced and applied to matrix A during network training to enhance the robustness of the model.

There is a denoising-CNN included in the flowchart of FLESH-DECT. We use this network to reduce the impact of image noise and the results show its effectiveness. However, the denoising network is not mandatory and it can be replaced with other image denoising techniques (Ma et al., 2011), such as non-local mean (NLM) (Zhang et al., 2013), block-matching and 3D filtering (BM3D) (Salehjahromi et al., 2017), without see-

ing severe degradation in performance. The denoising step may also be removed when inputting extremely-high quality images.

In this paper, the proposed method was tested on the Siemens DECT dataset as well as Philips spectral CT dataset. For Siemens DECT, both low- and high-energy projections can be attained from a CT system while in Philips spectral CT dataset, the projection views of those virtual monoenergetic images are unable to acquire from real CT machines. Here, we are using those simulated data to demonstrate that the proposed method has the potential to deal with different energy-level settings.

Despite all the advantages and potentials mentioned above, the proposed method has limitations. FLESH-DECT relies highly on the deep neural network to perform the material-decomposition-like operation. Since the domain knowledge learned by the network highly depends on training data, it is unlikely to provide reasonable results when there is a huge difference between training and testing data. For example, models trained under 100 kV/Sn 140 kV protocol may not generate correct results when inputting low-energy images scanned under 70 kV protocol. However, these limitations can be solved by training different models for different protocols.

5. Conclusion

In this paper, we proposed a deep learning approach to perform DECT imaging using a low-energy image and a single-view high-energy projection. Compared to standard DECT imaging, the approach can provide superior material-specific images with significantly reduced noise. It also has the potential to simplify the system design and reduce the radiation dose and allows us to perform high-quality DECT imaging without the conventional hardware-based DECT solutions. The approach

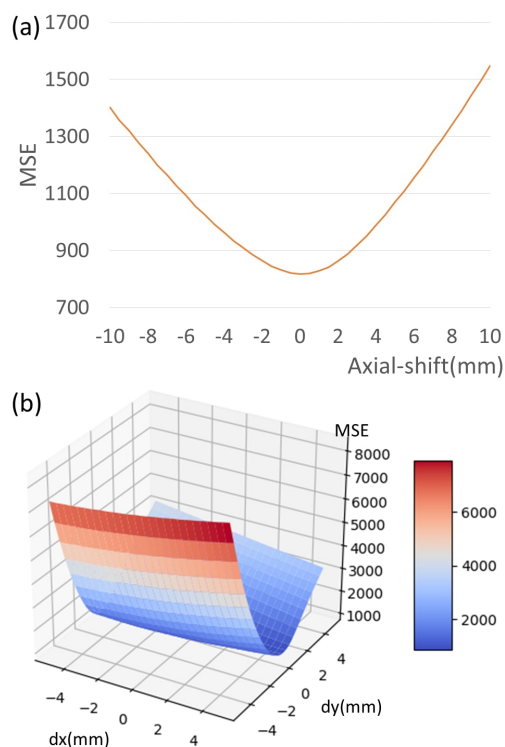


Figure 14: MSE values on test patient I when there is shift between projections at different energy-levels. (a) Shift in axial direction. (b) Shift in horizontal directions.

may significantly extend the usage of the widespread standard SECT scanners by providing advanced DECT clinical applications, such as urinary stone characterization and differentiating intracerebral hemorrhage from iodinated contrast, and thus lead to a new paradigm of SECT imaging.

References

Abadi M. et al., 2016. Tensorflow: Large-scale machine learning on heterogeneous distributed systems. *arXiv preprint arXiv:1603.04467*.

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

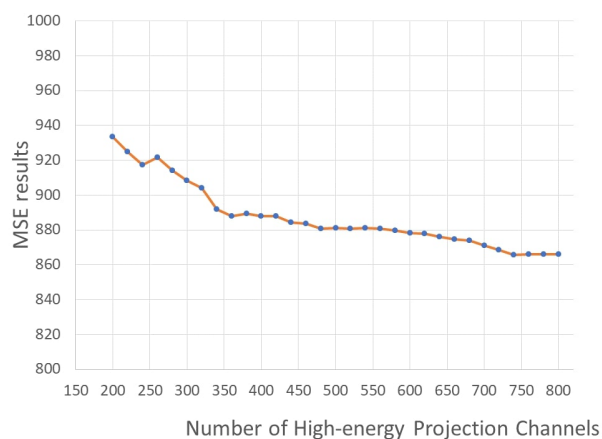


Figure 15: MSE values on test patient I under different high-energy projection channel numbers.

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

Ascenti G. et al., 2010. Stone-targeted dual-energy ct: a new diagnostic approach to urinary calculosis. *American Journal of Roentgenology*, 195(4), 953–958.

Boll D. T., Merkle E. M., Paulson E. K., Mirza R. A., and Fleiter T. R., 2008. Calcified vascular plaque specimens: assessment with cardiac dual-energy multidetector ct in anthropomorphically moving heart phantom. *Radiology*, 249(1), 119–126.

Boll D. T. et al., 2009. Renal stone assessment with dual-energy multidetector ct and advanced postprocessing techniques: improved characterization of renal stone composition—pilot study. *Radiology*, 250(3), 813–820.

Buerke B., Wittkamp G., Seifarth H., Heindel W., and Kloska S. P., 2009. Dual-energy cta with bone removal for transcranial arteries: intraindividual comparison with standard cta without bone removal and tomra. *Academic radiology*, 16(11), 1348–1355.

Chen H. et al., 2017. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36(12), 2524–2535.

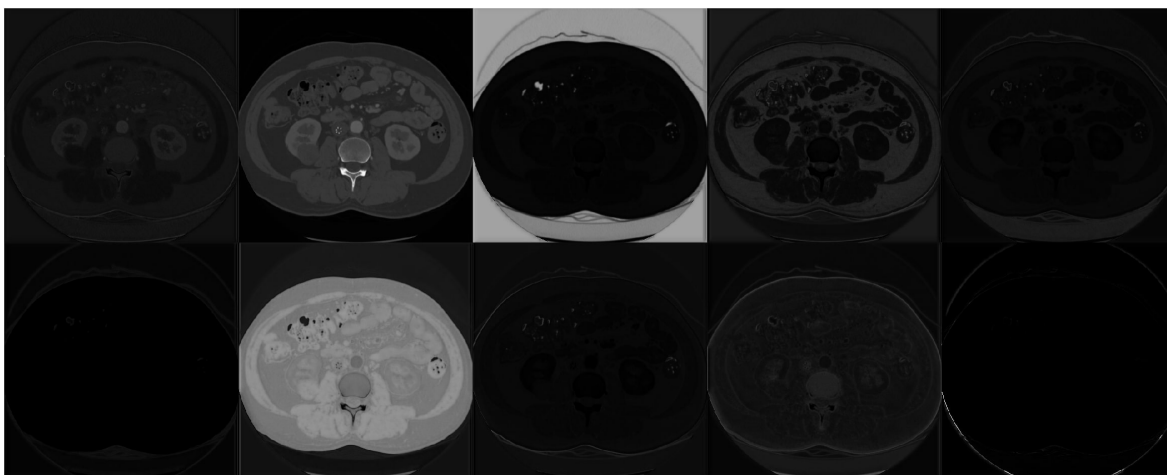


Figure 16: An example of the output of the MD-CNN. Each image represents a channel in the resulting matrix A_{DL} . All images are displayed under the same window with center=0.5 and width=1.0.

- Esteva A. et al., 2017. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 542 (7639), 115–118.
- Feng C., Kang K., and Xing Y., 2018. Fully connected neural network for virtual monochromatic imaging in spectral computed tomography. *Journal of Medical Imaging*, 6(1), 011006.
- Ferda J. et al., 2009. The assessment of intracranial bleeding with virtual unenhanced imaging by means of dual-energy ct angiography. *European radiology*, 19(10), 2518–2522.
- Flohr T. G. et al., 2006. First performance evaluation of a dual-source ct (dsct) system. *European radiology*, 16 (2), 256–268.
- Graser A. et al., 2009. Dual-energy ct in patients suspected of having renal masses: can virtual nonenhanced images replace true nonenhanced images? *Radiology*, 252(2), 433–440.
- Gulshan V. et al., 2016. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *Jama*, 316 (22), 2402–2410.
- Ho L. M. et al., 2012. Characterization of adrenal nodules with dual-energy ct: can virtual unenhanced attenuation values replace true unenhanced attenuation values? *American Journal of Roentgenology*, 198(4), 840–845.
- Johnson T. R. et al., 2007. Material differentiation by dual energy ct: initial experience. *European radiology*, 17 (6), 1510–1517.
- Kalender W. A., Perman W., Vetter J., and Klotz E., 1986. Evaluation of a prototype dual-energy computed tomographic apparatus. i. phantom studies. *Medical physics*, 13(3), 334–339.
- Kang E., Chang W., Yoo J., and Ye J. C., 2018. Deep convolutional framelet denoising for low-dose ct via wavelet residual network. *IEEE transactions on medical imaging*, 37(6), 1358–1369.
- Kingma D. P. and Ba J., 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.
- Lee D., Lee J., Kim H., Lee T., Soh J., Park M., Kim C., Lee Y. J., and Cho S., 2017. A feasibility study of low-dose single-scan dual-energy cone-beam ct in many-view under-sampling framework. *IEEE transactions on medical imaging*, 36(12), 2578–2587.

- Lee D., Kim H., Choi B., and Kim H.-J., 2019. Development of a deep neural network for generating synthetic dual-energy chest x-ray images with single x-ray exposure. *Physics in Medicine & Biology*, 64(11), 115017.
- Leng S., Yu L., Fletcher J. G., and McCollough C. H., 2015a. Maximizing iodine contrast-to-noise ratios in abdominal ct imaging through use of energy domain noise reduction and virtual monoenergetic dual-energy ct. *Radiology*, 276(2), 562–570.
- Leng S. et al., 2015b. Feasibility of discriminating uric acid from non-uric acid renal stones using consecutive spatially registered low-and high-energy scans obtained on a conventional ct scanner. *American Journal of Roentgenology*, 204(1), 92–97.
- Liao Y. et al., 2018. Pseudo dual energy ct imaging using deep learning-based framework: basic material estimation. In *Medical Imaging 2018: Physics of Medical Imaging*, volume 10573, page 105734N. International Society for Optics and Photonics.
- Liu F., Jang H., Kijowski R., Bradshaw T., and McMillan A. B., 2018. Deep learning mr imaging-based attenuation correction for pet/mr imaging. *Radiology*, 286(2), 676–684.
- Ma J., Huang J., Feng Q., Zhang H., Lu H., Liang Z., and Chen W., 2011. Low-dose computed tomography image restoration using previous normal-dose scan. *Medical physics*, 38(10), 5713–5731.
- Maier A. K. et al., 2019. Learning with known operators reduces maximum error bounds. *Nature machine intelligence*, 1(8), 373–380.
- Mangold S. et al., 2012. Virtual nonenhanced dual-energy ct urography with tin-filter technology: determinants of detection of urinary calculi in the renal collecting system. *Radiology*, 264(1), 119–125.
- 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(3), 637–653.
- Morhard D., Fink C., Graser A., Reiser M. F., Becker C., and Johnson T. R., 2009. Cervical and cranial computed tomographic angiography with automated bone removal: dual energy computed tomography versus standard computed tomography. *Investigative radiology*, 44(5), 293–297.
- Petrongolo M. and Zhu L., 2018. Single-scan dual-energy ct using primary modulation. *IEEE transactions on medical imaging*, 37(8), 1799–1808.
- Phan C., Yoo A., Hirsch J., Nogueira R., and Gupta R., 2012. Differentiation of hemorrhage from iodinated contrast in different intracranial compartments using dual-energy head ct. *American journal of neuroradiology*, 33(6), 1088–1094.
- Poirot M. G. et al., 2019. physics-informed deep learning for dual-energy computed tomography image processing. *Scientific reports*, 9(1), 1–9.
- Pomerantz S. R. et al., 2013. Virtual monochromatic reconstruction of dual-energy unenhanced head ct at 65–75 keV maximizes image quality compared with conventional polychromatic ct. *Radiology*, 266(1), 318–325.
- Primak A. N. et al., 2007. Noninvasive differentiation of uric acid versus non-uric acid kidney stones using dual-energy ct. *Academic radiology*, 14(12), 1441–1447.
- Ronneberger O., Fischer P., and Brox T., 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer.
- Salehjahromi M., Zhang Y., and Yu H., 2017. A spectral ct denoising algorithm based on weighted block matching 3d filtering. In *Developments in X-Ray Tomography XI*, volume 10391, page 103910G. International Society for Optics and Photonics.
- Schulz B. et al., 2012. Automatic bone removal technique in whole-body dual-energy ct angiography: performance and image quality. *American Journal of Roentgenology*, 199(5), W646–W650.
- Shen L., Zhao W., and Xing L., 2019. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nature biomedical engineering*, 3(11), 880–888.

- Sommer W. H. et al., 2009. The value of dual-energy bone removal in maximum intensity projections of lower extremity computed tomography angiography. *Investigative radiology*, 44(5), 285–292.
- Takahashi N. et al., 2008. Dual-energy ct iodine-subtraction virtual unenhanced technique to detect urinary stones in an iodine-filled collecting system: a phantom study. *American Journal of Roentgenology*, 190(5), 1169–1173.
- Ting D. S. W. et al., 2017. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *Jama*, 318(22), 2211–2223.
- Toepker M. et al., 2012. Virtual non-contrast in second-generation, dual-energy computed tomography: reliability of attenuation values. *European journal of radiology*, 81(3), e398–e405.
- Vedaldi A. and Lenc K., 2015. Matconvnet: Convolutional neural networks for matlab. In *Proceedings of the 23rd ACM international conference on Multimedia*, pages 689–692.
- Wolterink J. M., Leiner T., Viergever M. A., and Išgum I., 2017. Generative adversarial networks for noise reduction in low-dose ct. *IEEE transactions on medical imaging*, 36(12), 2536–2545.
- Xing L., Krupinski E. A., and Cai J., 2018. Artificial intelligence will soon change the landscape of medical physics research and practice. *Medical physics*, 45(5), 1791–1793.
- Xue Y., Jiang Y., Yang C., Lyu Q., Wang J., Luo C., Zhang L., Desrosiers C., Feng K., Sun X., et al., 2019. Accurate multi-material decomposition in dual-energy ct: A phantom study. *IEEE Transactions on Computational Imaging*, 5(4), 515–529.
- Yang Q. et al., 2018. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6), 1348–1357.
- Yu L., Leng S., and McCollough C. H., 2012. Dual-energy ct-based monochromatic imaging. *American journal of Roentgenology*, 199(5_supplement), S9–S15.
- Zhang H., Huang J., Ma J., Bian Z., Feng Q., Lu H., Liang Z., and Chen W., 2013. Iterative reconstruction for x-ray computed tomography using prior-image induced nonlocal regularization. *IEEE Transactions on Biomedical Engineering*, 61(9), 2367–2378.
- Zhang W. et al., 2019. Image domain dual material decomposition for dual-energy ct using butterfly network. *Medical physics*, 46(5), 2037–2051.
- Zhao W., Lv T., Lee R., Chen Y., and Xing L., 2020. Obtaining dual-energy computed tomography (ct) information from a single-energy ct image for quantitative imaging analysis of living subjects by using deep learning. In *Pac Symp Biocomput.* World Scientific.
- Zhao W. et al., 2016. Using edge-preserving algorithm with non-local mean for significantly improved image-domain material decomposition in dual-energy ct. *Physics in Medicine & Biology*, 61(3), 1332.
- Zhao W. et al., 2019a. Markerless pancreatic tumor target localization enabled by deep learning. *International Journal of Radiation Oncology* Biology* Physics*, 105(2), 432–439.
- Zhao W. et al., 2019b. Incorporating imaging information from deep neural network layers into image guided radiation therapy (igrt). *Radiotherapy and Oncology*, 140, 167–174.