Virtual computed-tomography system for deep-learning-based material decomposition

Daiyu Fujiwara[1], Taisei Shimomura[1], Wei Zhao[2], Kai-Wen Li[3], Akihiro Haga[1], and Li-Sheng Geng[2,4,5,6]

[1] Graduate School of Biomedical Sciences, Tokushima University, Tokushima 770-8503, Japan

[2] School of Physics, Beihang University, Beijing 102206, China

[3] Medical Management Department, CAS Ion Medical Technology Co., Ltd., Beijing 100190, China

[4] Beijing Key Laboratory of Advanced Nuclear Materials and Physics, Beihang University, Beijing 102206, China

[5] Beijing Advanced Innovation Center for Big Data Based Precision Medicine, School of Medicine and Engineering, Beihang University, Key Laboratory of Big Data Based Precision Medicine (Beihang University), Ministry of Industry and Information Technology, Beijing 100191, China

[6] School of Physics and Microelectronics, Zhengzhou University, Zhengzhou, Henan 450001, China

E-mails: kwli@cashim.cn (KL); haga@tokushima-u.ac.jp (AH); lisheng.geng@buaa.edu.cn (LG)

Abstract

Objective: Material decomposition (MD) evaluates the elemental composition of human tissues and organs via computed tomography (CT) and is indispensable in correlating anatomical images with functional ones. A major issue in MD is inaccurate elemental information about the real human body. To overcome this problem, we developed a virtual CT system model, by which various reconstructed images can be generated based on ICRP110 human phantoms with information about six major elements (H, C, N, O, P, and Ca).

Approach: We generated CT datasets labelled with accurate elemental information using the proposed generative CT model and trained a deep learning (DL)-based model to estimate the material distribution with the ICRP110 based human phantom as well as the digital Shepp–Logan phantom. The accuracy in quad-, dual-, and single-energy CT cases was investigated. The influence of beam-hardening artefacts, noise, and spectrum variations were analysed with testing datasets including elemental density and anatomical shape variations.

Main results: The results indicated that this DL approach can realise precise MD, even with single-energy CT images. Moreover, noise, beam-hardening artefacts, and spectrum variations were shown to have minimal impact on the MD.

Significance: Present results suggest that the difficulty to prepare a large CT database can be solved by introducing the virtual CT system and the proposed technique can be applied to clinical radiodiagnosis and radiotherapy.

Keywords: material decomposition, deep learning, computed tomography, ICRP110 human phantom

1. Introduction

In recent years, the problem of high-quality image-reconstruction for computed tomography (CT) has drawn increasing attention [1–3]. CT has numerous medical applications, including disease diagnosis, differential

diagnosis, radiotherapy treatment planning, and surgery navigation. An especially important application is material decomposition (MD), in which the spatial density distribution of each chemical element in an object is evaluated based on CT imaging. MD is very useful for clinical radiotherapy and radiodiagnosis (e.g., kidney-stone characterisation, liver-fat quantification, and the identification of contrast-filled agent vessels) [4–6]. In addition, MD can help differentiate normal tissues from tumours, which have higher concentrations of hydrogen [7]. In proton therapy, the proton stopping-power ratio relative to water must be precisely determined during treatment planning [8]; according to the Bethe–Bloch equation and Bragg additivity rule, the mass fractions of the constituent elements (especially the six 'macro-elements' H, C, N, O, P, and Ca) are the key inputs in this calculation. However, all current approaches to predicting the stopping-power ratio use the effective atomic number [9], leading to less accurate results.

MD is a nonlinear and nonconvex inverse problem [10]. Its theoretical frameworks can be classified into three categories: projection-domain processing [11–13], image-domain processing [14–16], and one-step-inversion processing [17,18]. In projection-domain methods, base material sinograms are obtained first; images for each material are then reconstructed using material-specific sinograms, reducing nonlinear effects that cause beam-hardening artefacts. However, these sinograms can suffer from mismatches in the projection data of each energy. In image-domain methods, which are widely applied in clinical and pre-clinical situations, materials are decomposed after image reconstruction. These methods are relatively simple because there is no requirement for projection registration. However, they depend heavily on the image reconstruction algorithm, and the results are sensitive to beam-hardening and other artefacts, as well as to resolution error. One-step-inversion approaches directly estimate the material information from spectral CT transmission data using iterative algorithms. Although the drawbacks of the first two methods can sometimes be avoided, the computational processing is slow owing to the complex analytical form of the models, and the accuracy relies on prior information and the choice of the regularisation term. In particular, the model's settings need to be changed when it is applied to different situations. Thus, existing approaches to CT MD clearly require improvement.

Deep learning (DL) has had a huge impact on science and industry. In particular, it is an ideal tool for image processing [19], because it integrates the linear and nonlinear elements in the inner products and activation functions. In recent years, DL-based approaches to medical imaging have obtained promising results for tasks such as sparse-view image reconstruction [20], cardiovascular-risk prediction [21], vessel segmentation [22], and lesion localisation [23]. Owing to their outstanding effectiveness in feature extraction and modelling,

deep convolutional neural networks (CNNs) based on the U-net [24,25] and VGG-16 [26] architectures have been proposed for multi-MD of spectral CT in the image domain. To apply DL-based approaches to MD, however, it is necessary to prepare a large database containing CT images and corresponding material information. In routine clinical tests, it may be unrealistic to collect these simultaneously for a large cohort.

To overcome these challenges, we developed a DL-based virtual CT system model that generates reconstructed images based on ICRP110 human phantoms composed of the six macro-elements (H, C, N, O, P, and Ca). First, the X-ray spectrum of a real CT machine was estimated using appropriate techniques [27–30]. Second, a virtual CT system was developed with the same geometry settings as a real CT machine. Third, various virtual phantoms composed of H, C, N, O, P, and Ca were prepared. With the information of the material component in each pixel of a phantom, realistic sinograms could be produced using the estimated X-ray spectrum in the virtual CT device [31,32]. After image reconstruction, each pixel in the image could be associated with its corresponding material components. Fourth, a DL-based model was trained and tested using the reconstructed images and their elemental compositions. Finally, the model was applied to real CT images.

U-net and VGG-based convolutional network models were used to estimate the density distributions of the six macro-elements. The training and testing were evaluated using reconstructed images of the digital Shepp–Logan (SL) phantom and ICRP110 human-phantom slices. The virtual CT system, implemented with four X-ray energy spectra, was applied to data acquisition and image reconstruction. Three different cases were considered: quad-, dual-, and single-energy CT (QECT, DECT, and SECT), meaning CT images generated from four, two, or one polychromatic X-ray beams, respectively. The influence of beam-hardening artefacts, noise, and spectrum variations were analysed. Our results demonstrate that DL can predict the mass fraction of macro-elements precisely, even when using SECT. Moreover, using the DL method, image quality ceases to be a drawback, as it is in image-domain approaches. These observations provide strong support for the implementation of this technique in radiodiagnosis and radiotherapy.

2. Methods

2.1 Data preparation

The study was based on two categories of two-dimensional digital phantoms, that is, 54 biological tissue (BT) phantoms (in SL shape) and two ICRP110 human phantoms. Each of the former was constructed with three random BT materials from ICRU Report No. 44 [33–35]. The elemental compositions are listed in Table 1 [36]. The ICRP110 human phantoms consist

of a total of 566 slices of the whole-body scan of a male (220 slices) and a female (346 slices). The detailed size and shape information of these two phantoms are shown in Fig. 1. As seen in Fig. 1, the structure of the SL phantom is simple, whereas that of the human phantom is more complicated, reflecting human anatomy. By using such different structures, the effectiveness of MD shape-information prediction could be tested.
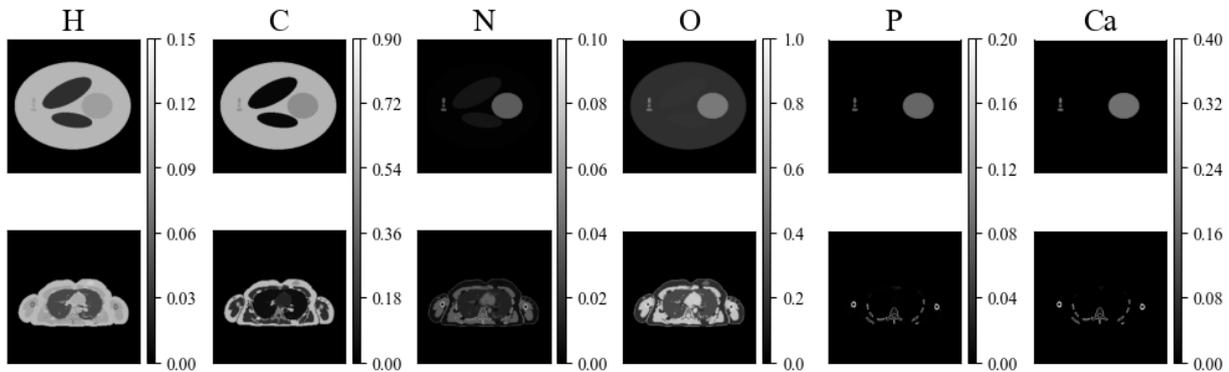
Fig. 1 Representative distribution images for six elements (H, C, N, O, P, and Ca) in the Shepp–Logan phantom (top) and ICRP110 human phantom (bottom).

In addition, we prepared human phantoms by deforming from the ICRP110 human phantom to real CT images using deformable image registration (DIR). The real CT image datasets were collected from The Cancer Imaging Archive (TCIA) [37]. Thirty CT images from the head, lung, and pelvis regions were sampled from TCIA and the CT-to-density conversion for 120 kVp was adapted. Note that the adapted conversion does not correspond to that in the sampled images, so the DIR accuracy might be reduced. This inconsistency in CT-to-density conversion, however, is not significant, because we only want a deformed image including elemental information from ICRP110, where the accuracy of DIR does not greatly matter. Sequential preprocessing, such as removing the couch and/or supporter, was performed. Then, elastix [38] was used to perform DIR; the moving and the targeting images were the density ones of ICRP110 phantom (female) and the preprocessed TCIA image, respectively. Representative deformed images are shown in Fig. 2, where the parameter set used in DIR is also shown. These deformed phantoms were only used for the evaluation for model testing.

The phantoms were projected via a virtual fan-beam CT implemented with four X-ray spectra: 80, 100, 120, and 6 MV. Images were reconstructed using an in-house program for the filter-back projection method with an SL filter; the sinogram size was 609×800. The details of the virtual CT system are described in the supporting information. Consequently, 54 images of the BT phantoms and 566 images of the human phantoms as well as 30 images of deformed phantoms of each energy CT were obtained.

Parameters	Value
Metric	Advanced Mattes Mutual Information
Optimizer	Adaptive Stochastic Gradient Descent
Transform	B-spline
Num. Iterations	1,000
Grid Spacing	32
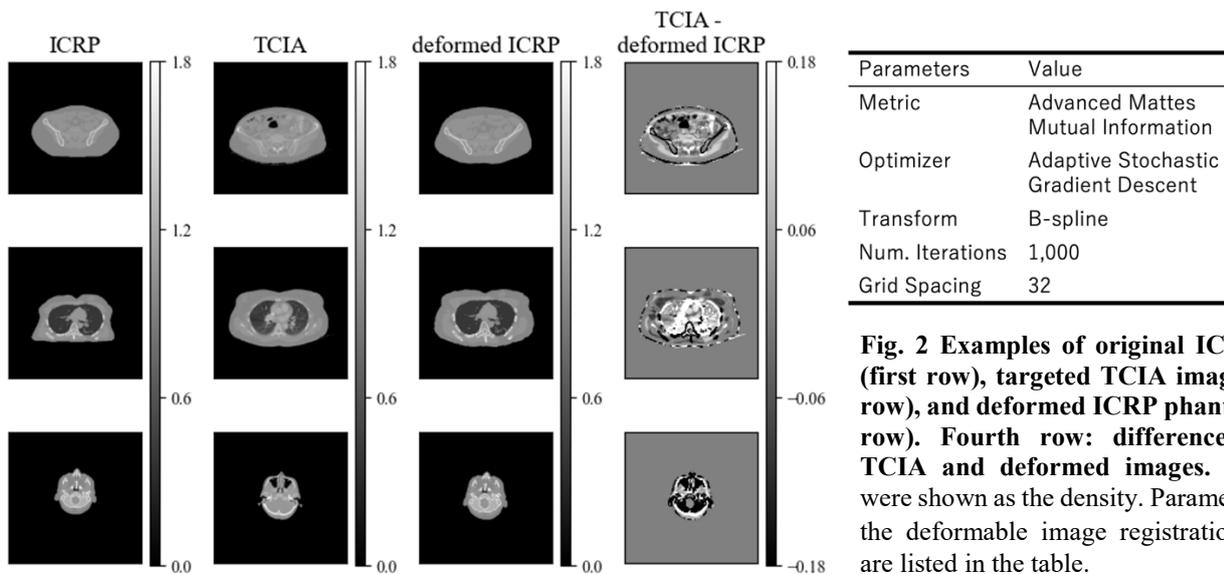
Fig. 2 Examples of original ICRP images (first row), targeted TCIA images (second row), and deformed ICRP phantoms (third row). Fourth row: differences between TCIA and deformed images. All images were shown as the density. Parameters used in the deformable image registration (elastix) are listed in the table.

Table 1. Specifications of the phantoms used in this study [36].

Name	Training		Testing		
	Biological tissue	ICRP110 humans	Biological tissue	ICRP110 humans	Deformed phantoms
Element	H, C, N, O, P, Ca	H, C, N, O, P, Ca	H, C, N, O, P, Ca	H, C, N, O, P, Ca	H, C, N, O, P, Ca
Shape	Shepp–Logan	Whole body	Shepp–Logan	Head/Lung/Pelvis	Head/Lung/Pelvis
Width (mm)	260	113–474	260	113–474	95-418
Height (mm)	197	150–210	197	150–210	153-231
Image (mm^2)	300×300	512×512	300×300	512×512	512×512
Slice(s)[a]	44	566	10	30	30
No. of images[b]	2112	2264	10	30	30

a: Number of phantoms (slices) used in training and testing.

b: Number of images after data augmentation.

2.2 Data pre-processing and augmentation

To investigate beam-hardening effects in MD, an empirical correction term was applied in the projection domain (mainly to remove the cupping artefacts) [32]. Furthermore, Gaussian distributed noise was added to the reconstructed images to obtain a signal-to-noise ratio of ~20. Data creation codes, including virtual projection, reconstruction, and beam-hardening correction, were developed using C++ and are available in the Github repository,
https://github.com/hagaakihiro/VirtualCT
An example comparison of the reconstructed images with and without noise and beam-hardening correction using the 120-kV spectrum is shown in Fig. 3. Augmentation was applied to

increase data diversity, which enables more effective model training and improves robustness. In this study, the training data of BT phantoms were augmented by horizontal and vertical flipping, rotation with a maximum angle of $\pm 20°$, and random image cropping and patching (RICAP) [39]. A pre-augmentation dataset of size N resulted in an augmented dataset of size $48N$. Note that flipping and rotation were applied in the image domain, while RICAP was used during the phantom-creation procedure so that the beam-hardening effects could be correctly treated. By contrast, the data of the ICRP110 human phantoms were augmented by adding a Gaussian deviation randomly in each anatomy, where a standard deviation of 5% for the mass density and 10% for the mass fraction of each element in a phantom slice were applied.
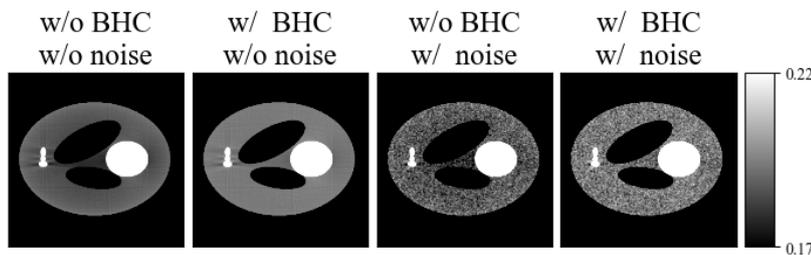
Fig. 3 Reconstructed images using the 120 kV spectrum for four different scenarios. BHC: beam-hardening correction.

2.3 Model development

In this study, we developed two MD models in the image domain based on the U-net and VGG architectures. CT images from SECT (120 kV), DECT (120 kV – 6 MV), and QECT (80 kV – 100 kV – 120 kV – 6 MV) were used as inputs, while

the outputs were density images of the six macro-elements. We employed a 6-MV CT spectrum as one of the spectra because it has been found that kV–MV X-ray-beam pairs outperform kV–kV DECT in effective atomic number (EAN or Z_{eff}), electron-density (ϱ_e) ratio (EDR), and stopping-power ratio (SPR) estimation. Note that MV CT has previously been applied in helical tomotherapy [40]; more

recently, it has been used in conjunction with a kV X-ray beam [41]. In light of this prior work, we selected the combination of 120 kV and 6 MV for the input CT energy in the MD model.

The U-net model had one contracting path with inputs of one, two, or four layers, which respectively corresponded to the reconstructed images of SECT, DECT, or QECT. After downsampling, the feature maps were concatenated in the bottommost layer. Upsampling and skip connection were subsequently performed to output six layers. Similarly, the VGG model had one stack of convolutional layers that extracted patches and features from one, two, or four input layers before the fully connected layers. Each convolutional layer was batch-normalised and activated by ReLU, except for the downsampling layers of the U-net model, where LeakyReLU was used after batch normalisation. The details of the two network architectures are shown in Fig. 4.

To create a baseline, we also applied a DECT-to-EDR-EAN lookup-table approach [36] to estimating the material density distribution, using the training sets for the human phantoms mentioned previously. The ϱ_e and corresponding Z_{eff} distributions were first reconstructed from the DECT analysis given in Ref. [32]; Z_{eff} in each pixel could be obtained by solving for Z in

$$\frac{<\mu>_L}{<\mu>_H}(Z)$$
$$= \frac{\sum_i \alpha_{i,L} \left(Z^4 f_1(E_{i,L}) + Z^3 f_2(E_{i,L}) + f_3(E_{i,L}) + Z f_4(E_{i,L}) \right)}{\sum_i \alpha_{i,H} \left(Z^4 f_1(E_{i,H}) + Z^3 f_2(E_{i,H}) + f_3(E_{i,H}) + Z f_4(E_{i,H}) \right)}, (1)$$

where $\alpha_{i,L(H)}$ is the energy fraction of the X-ray spectrum at energy $E_{i,L(H)}$ in the ith energy bin for lower (higher) tube voltage; $<\mu>_{L(H)}$ is the mean attenuation coefficient calibrated from the CT image for lower (higher) tube voltage;

and $f_1 \sim f_4$ come from the regression or theoretical formulae for the photoelectric, Compton, and pair-production effects. In the present study, these are

$$f_1(E) = 2.83 \times 10^{-8} E^{-3}, \quad (2)$$
$$f_2(E) = 3.45 \times 10^{-6} E^{-3}, \quad (3)$$
$$f_3(E) = 2\pi \left\{ \frac{1+E}{E^2} \left[\frac{2(1+E)}{1+2E} - \frac{ln(1+2E)}{E} \right] + \frac{ln(1+2E)}{2E} - \frac{1+3E}{(1+2E)^2} \right\}, (4)$$

$$f_4(E) = \frac{0.2545}{137}(E - 2.332). \quad (5)$$

Following Ref [36], we assumed that the MD was being applied to types of tissue with different compositions, separated by a certain threshold in EAN (Z_{eff}^{th}). In this study, we set $Z_{eff}^{th} = 7.5$ as the threshold value. The normalised elemental mass fractions w_i were obtained by fitting the EDR and EAN data to

$$w_i = a_i \varrho_e + b_i Z_{eff} + c_i \varrho_e Z_{eff} + d_i, \quad (6)$$

where $i \in$ H, C, N and O for $Z_{eff} < 7.5$, and $i \in$ H, C, N, O, P, and Ca for $Z_{eff} \geq 7.5$. That is, we assumed one component of the human body to be composed of four elements (H, C, N, and O) with Zeff < 7.5, and the other to be composed of all six macro-elements, with Zeff ≥ 7.5. The parameters $a_i \sim d_i$ were determined by fitting Eq. (6) with the dataset including $(w_i, \varrho_e, Z_{eff})$ by a linear least-squared method under the constraint that $\sum_i w_i = 1$.
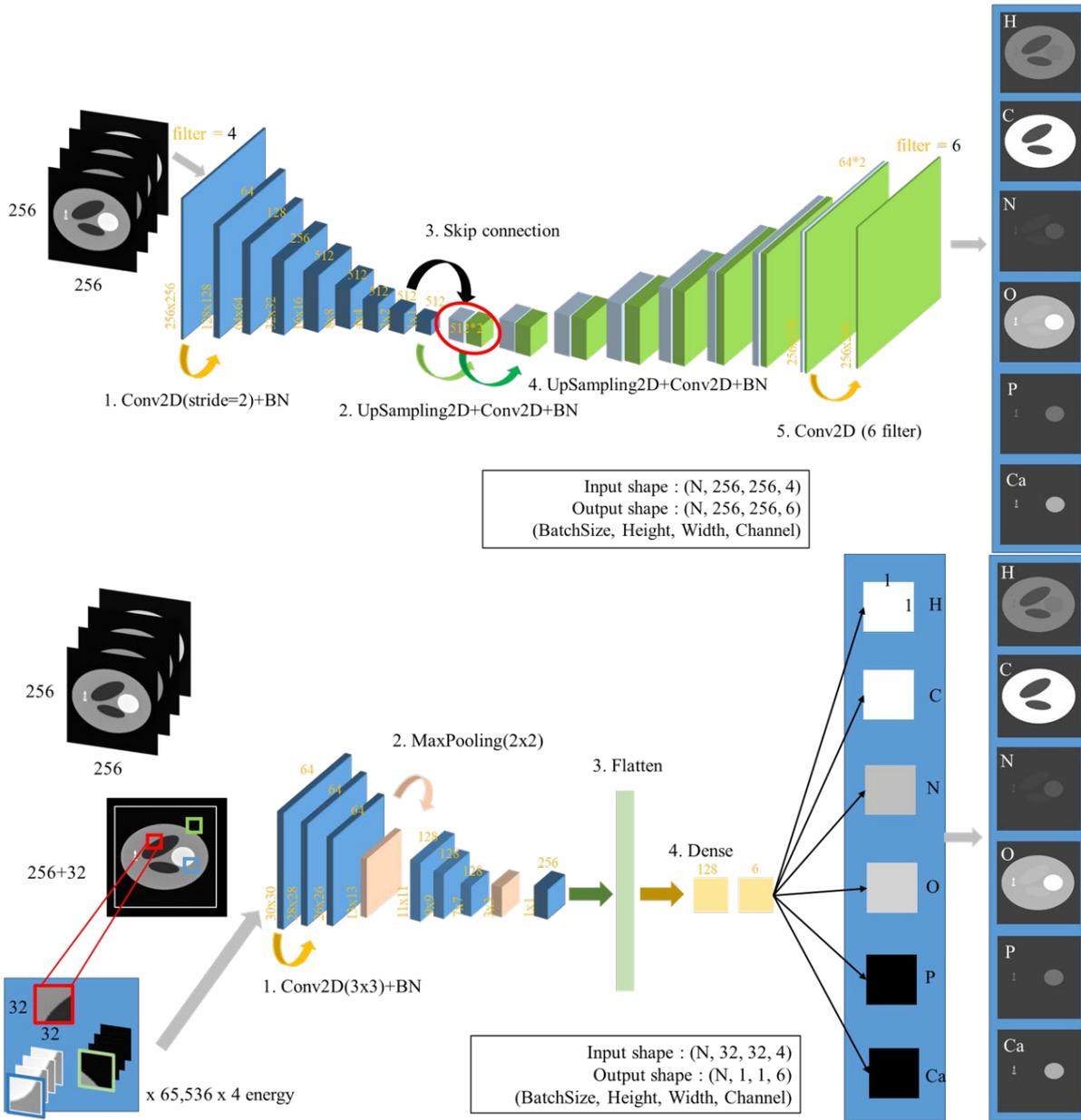
Fig. 4 Deep neural networks used for material decomposition. Top: U-net model, bottom: VGG model. The entire image (256×256) is input into the U-net model, whereas a patch (32×32) is used in the VGG model. Single-, dual-, and quad-energy CT images form a two-dimensional image with one, two, or four channels, respectively. The batch sizes N=32 and N=256 are used for U-net and VGG, respectively. The convolution kernel size is 3×3 for VGG and for upsampling in U-net, and 4×4 for downsampling in U-net.

2.4 Training details

The inputs of the aforementioned DL architectures were images with resolutions renormalised to 256×256 mm^2. The models were trained with the mean squared error (MSE) loss function, the Adam optimiser, an initial learning rate of 0.001 (which was reduced by a factor of 0.5 after five epochs without improvement), and a minimum learning rate of 0.00001. To validate the models and fine-tune the hyperparameters, 20% of the training data were used for cross-validation. Simulations of the data-preparation process and model training/testing were implemented using C++ and Python (with TensorFlow and Keras libraries), respectively. We used a Linux server equipped with an NVIDIA GeForce 2080Ti GPU with 64 GB of memory for the experiments in this study.

2.5 Model effectiveness

The performance of the models was evaluated using the root mean squared error (RMSE) and structural similarity (SSIM) between the predicted density distributions of the six macro-elements and their ground truth. These metrics are defined as

$$RMSE = \sqrt{\frac{\sum_N (\rho_{pred.} - \rho_{gt})^2}{N}}, \qquad (7)$$

$$SSIM(pred., gt) = \frac{(2\mu_{pred.}\mu_{gt} + c_1)(\sigma_{pred.,gt} + c_2)}{(\mu_{pred.}^2 + \mu_{gt}^2 + c_1)(\sigma_{pred.}^2 + \sigma_{gt}^2 + c_2)}, \qquad (8)$$

where $\rho_{pred.}$ and ρ_{gt} are the predicted and true density values, respectively. N denotes the pixel number in the region of interest (ROI). The shape of the ROI was chosen as a rectangle surrounding the objects in each image. Here, $\mu_{pred.}$ (μ_{gt}) and σ_{pred} (σ_{gt}) indicate the average value and standard deviation of the predicted (true) density in the ROI, respectively, and $\sigma_{pred.,gt}$ is their covariance. The values $c_1 = 0.0004$ and $c_2 = 0.0036$ were assumed in the calculations. We used segment masks to limit the SSIM calculation to the anatomical regions only.

One of the difficult tasks in MD is distinguishing between anatomies with similar attenuation coefficients (CT values) but different material weights. Therefore, it is worthwhile to verify the estimation accuracy in representative anatomies having such a relation. In this study, spongiosa in the sacrum and cartilage that appeared in the pelvic region, having similar CT values in the 120-kVp X-ray energy, were selected to demonstrate the performance of the DL-based MD.

3. Results

The RMSE and SSIM of the predicted MDs for both the SL phantom and ICRP110 human phantom are shown in Fig. 5. It was found that, in general, U-net predicted the MD more accurately in the SL and ICRP-human phantoms, while the VGG did better in the deformed human phantom.

The improvement of the U-net model compared with the VGG model in the SL and ICRP cases mainly originated from the estimations of C and O, as shown in Fig. 6 (using the ICRP human phantom as an example). This was because the VGG model only learned local information (patches), whereas the U-net model was trained by using global information (the whole image). The structure of the SL phantom was simple; thus, the predictions of the two DL-based models were similar. However, when dealing with an anatomically more complicated human phantom, the U-net model did better. In addition, the predictions from QECT, DECT, and SECT in the two DL-based approaches showed similar accuracies: the DL framework, even when using SECT images, could obtain precise MD results. The influence of noise was subtle for both the U-net and VGG models in the predictions of the SL phantom. Although noise influence could be seen in the case of the human phantom, this behaviour was less pronounced in the U-net model than in the VGG model. These findings indicate that shape information is important in the DL-based method, as explained previously.

However, with the deformed human phantom, the VGG model outperformed U-net, because the anatomical structures learned from the training data could be only partly used in the prediction (testing data). The fact that the RMSE in the VGG model for the deformed human phantom was comparable with that for the ICRP human phantom implies that the model training in VGG almost converged. On the other hand, the result also implies that U-net needs more training data, i.e., variety of anatomical shapes, to achieve its maximum potential performance. Overall, it can be said that, while both VGG and U-net predict elemental information for the human body, VGG is better for a small cohort size and U-net for a large one. The accuracy of both models slightly increased with the number of CT images with different energy spectra used in the prediction. More importantly, DL models can predict the material-density distribution with high accuracy using only SECT images.

The influence of beam-hardening effects was also compared for different models and phantoms. As expected, both DL-based models produced almost the same results by using the images with and without beam-hardening correction for the two phantoms. This result shows that the DL-based model is very robust for noise and/or image artefacts.

The kV–kV DECT method had the worst performance: the RMSE and single-SSIM dramatically increased, especially when noise was added. This method predicted the elemental information only from pixel-value information, never considering structural information; therefore, it was obviously at a disadvantage compared with the DL-based model.
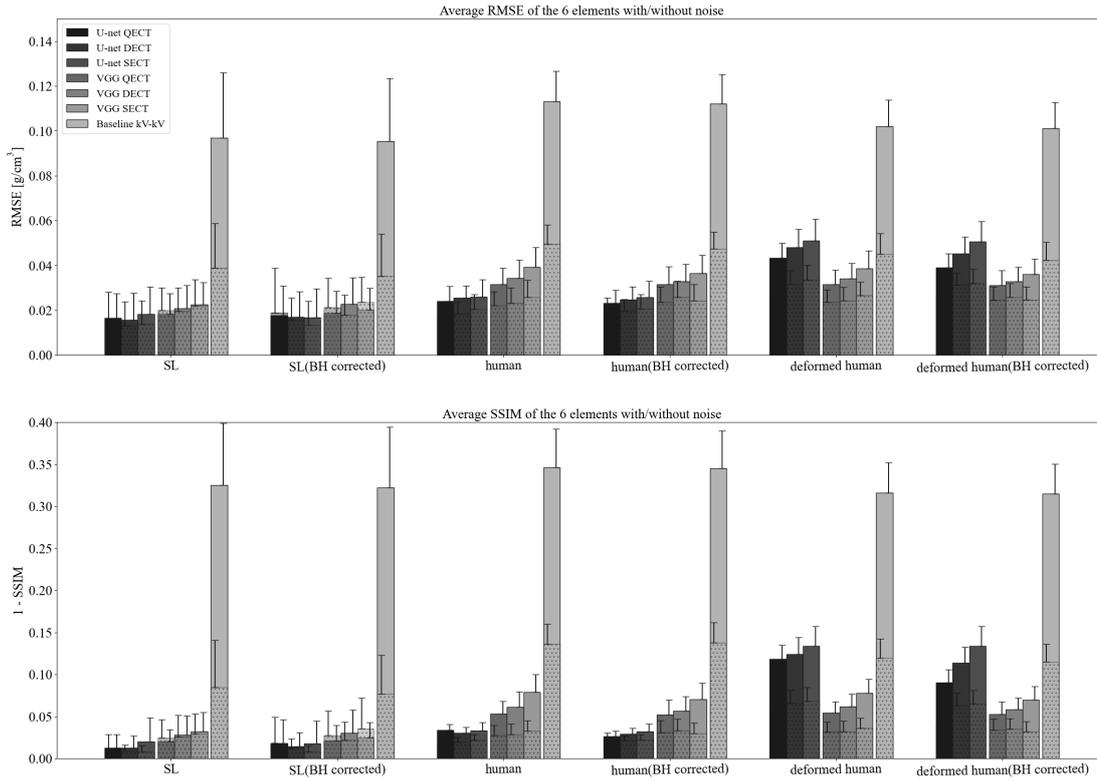
Fig. 5 Average root mean squared error (top: RMSE) and single-structural similarity (bottom: 1-SSIM) of the predicted material distributions for the Shepp–Logan and (ICRP and deformed) human phantoms. BHC indicates that the beam-hardening artefact was corrected in the reconstructed image. The dotted bars indicate the corresponding results without noise.
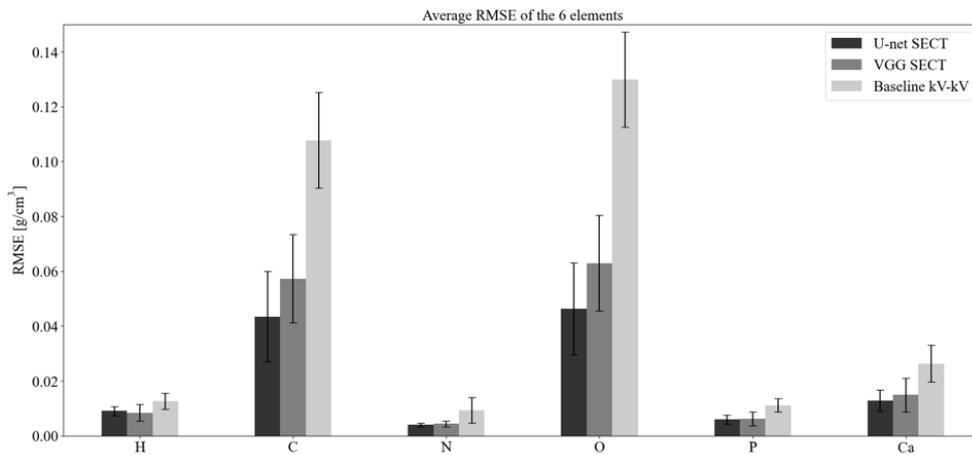
Fig. 6 Average root mean squared error of the predicted material distributions for the six elements of the ICRP110 human phantom (without beam-hardening correction and noise).

Figure 7 shows an example of the predicted material distributions (in male lung tissue) from the U-net and VGG models by SECT, DECT, QECT, and the baseline method with kV–kV DECT, as well as the ground truth. No noise or beam-hardening correction was applied. Qualitatively, the two DL-based models had good prediction results for the six macro-elements in all three CT cases. In particular, the detailed structure of the spinal cord could be reproduced well. However, the kV–kV DECT method generally yielded worse results and failed to predict the O density precisely in the spinal cord.
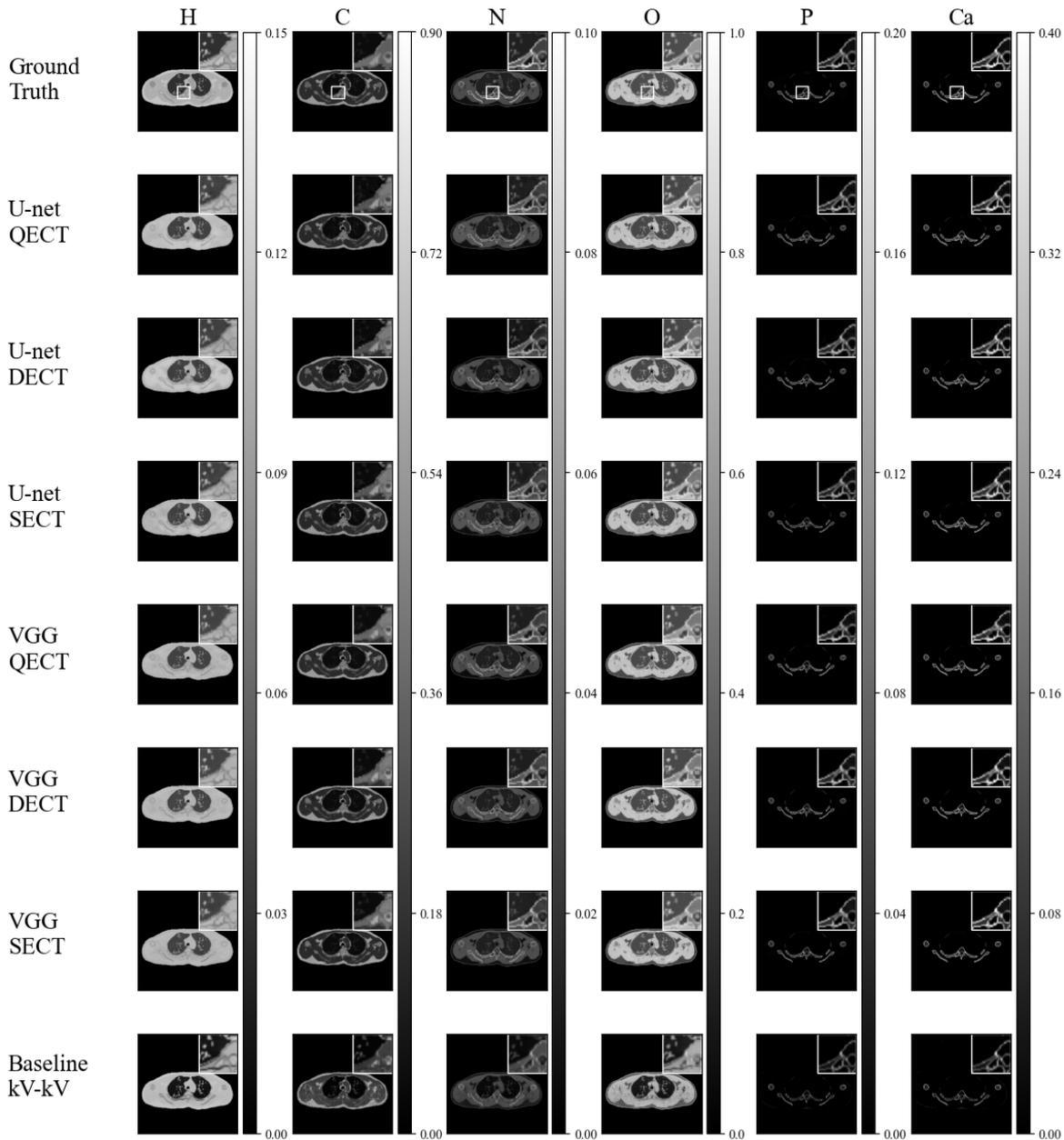
Fig. 7 Example of predicted material-composition distributions.

The residual maps between the predicted material distributions and ground truth in various situations are illustrated in Fig 8. The two DL-based approaches were insensitive to the number of X-ray energy spectra. The results from the U-net model generally had smaller errors compared with those from the VGG model. The enlarged panels show that the U-net model better preserved the shape. Similarly, relatively large errors were found for C and O compared to the other four elements. This is because there are large diversities in the density distributions of C and O in biological tissues: when the total density is the same, the densities of C and O in different tissues are quite different, as shown in Fig. 6. Thus, it is difficult to predict the density distributions accurately.

Figure 9 shows the residual map of DL-based MD (the U-net model) using SECT in the pelvis region, where the sacrum (spongiosa) and a sample of cartilage having similar CT values are the focus of attention. As shown in Table 2, DL-based MD can distinguish between them, even though the actual material weights differ significantly. This result also indicates that the accuracy of MD using SECT is still high and the DL-based method effectively uses not only CT value information but also global shape information.
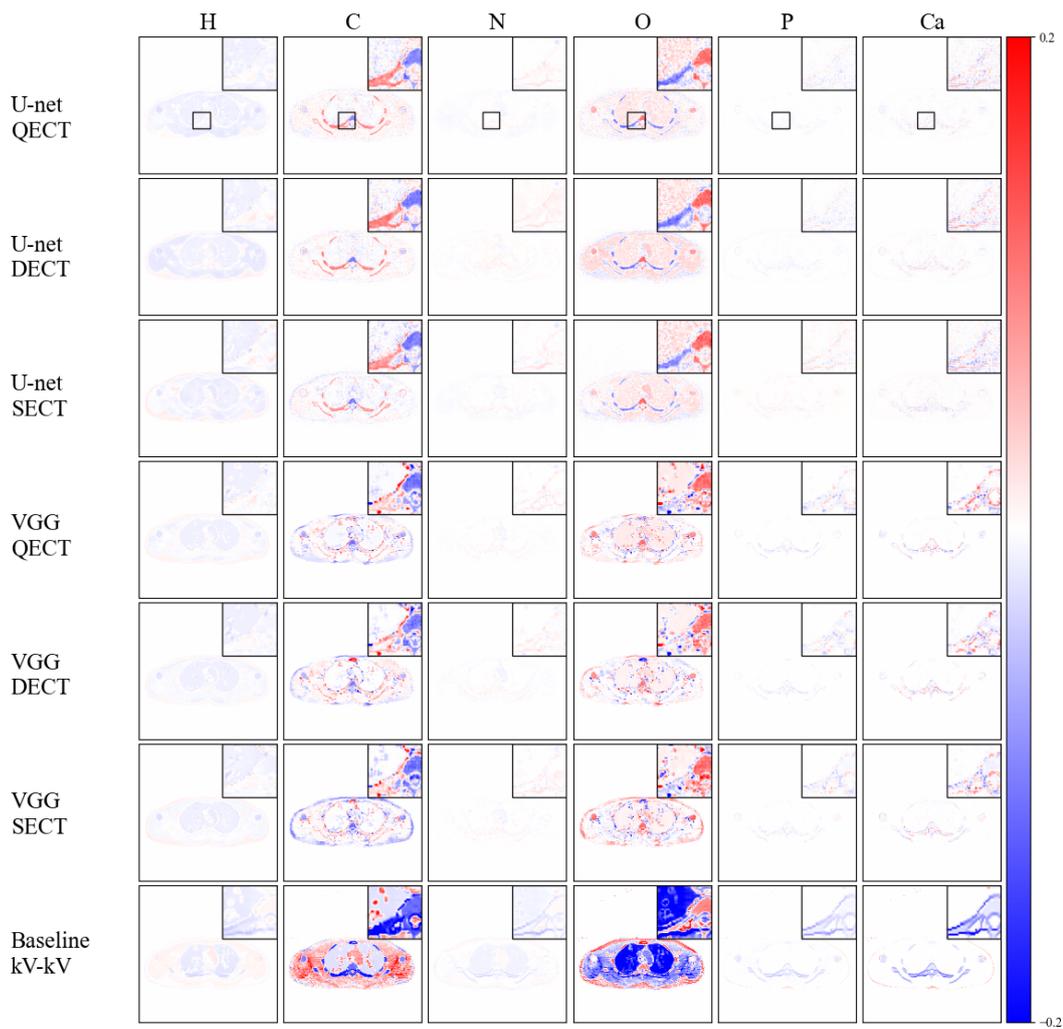
Fig. 8 Residual maps of the areas shown in the predicted material-composition distributions.
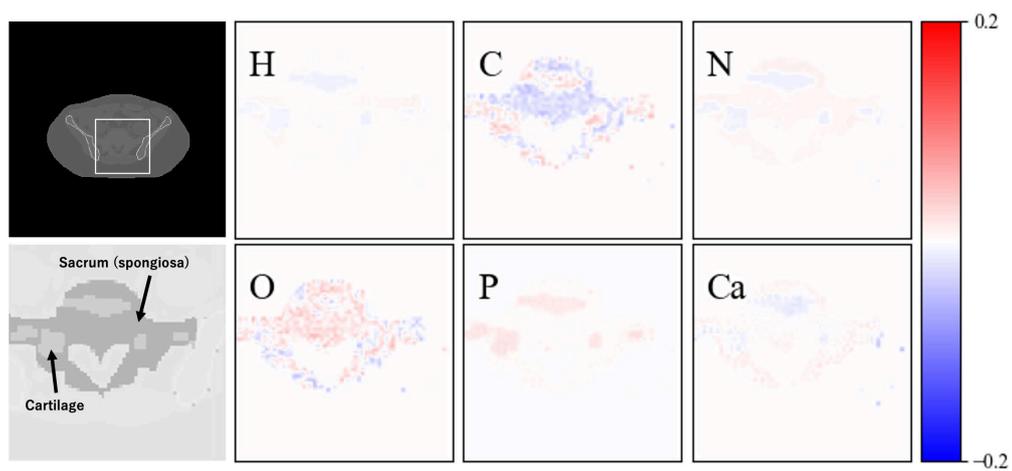
Fig. 9 Residual maps in deep-learning-based material decomposition (the U-net model) using single-energy CT for sacrum (spongiosa) and cartilage. Upper Left: whole image. Lower Left: image focused on sacrum (spongiosa) and cartilage. Remaining images: residual map of each element.

Table 2. Characteristics of sacrum (spongiosa) and cartilage, and the material density estimated by deep-learning-based material decomposition (the U-net model) using single-energy CT. Parentheses: root mean squared error.

Organ	Attenuation coefficient (120 kVp)	Total density [g/cm³]	H [g/cm³]	C [g/cm³]	N [g/cm³]	O [g/cm³]	P [g/cm³]	Ca [g/cm³]
Sacrum (spongiosa)	0.2129	1.05	0.111 (0.0256)	0.410 (0.0922)	0.030 (0.0238)	0.477 (0.1027)	0.007 (0.0264)	0.015 (0.0277)
Cartilage	0.2261	1.098	0.094 (0.0085)	0.114 (0.0889)	0.027 (0.0043)	0.840 (0.0984)	0.023 (0.0064)	0.000 (0.0131)

As mentioned above, the first step in connecting our strategy to the clinic is to estimate the X-ray spectrum of a real CT machine precisely. The mean absolute error (MAE) of the estimated spectrum presently can range up to 20% [27]. The MAE is defined as

$$MAE = \frac{1}{N} \sum_{i}^{N} |y_i - t_i|, \qquad (9)$$

where i denotes the energy bin (1 keV for the kV spectra), and y_i and t_i are the estimated and true fractions, respectively, of the ith bin. Thus, we studied the influence of the spectral variations in the case of 120 kV SECT for the human phantom. Using the SPEKTRA toolkit [42], various X-ray spectra were generated at a tube voltage of 120 kV by uniformly sampling the ripple in the range 0%–100%, Al-filtration thickness in the range 0–5 mm, and Cu-filtration thickness in the range 0–1 mm. Spectra with MAEs of 5%, 10%, 15%, and 20% compared with the base spectra (the 120 kV spectrum shown in the supporting information) were selected, as shown in Fig. 10.
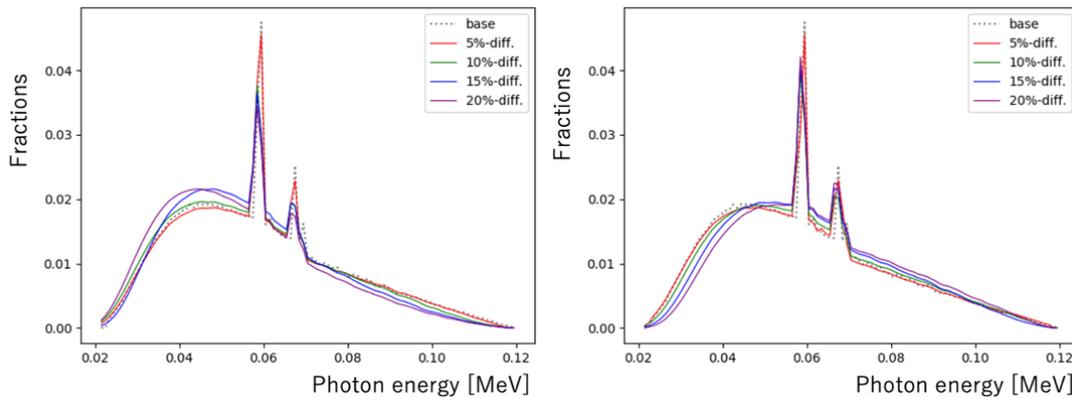
Fig. 10 Under- (left) and over- (right) estimated 120-kV X-ray spectrum (mean energy of the base spectrum: 59.3 keV). The mean energies for the underestimated spectra are 59.1 keV (5%), 57.7 keV (10%), 57.2 keV (15%), and 55.1 keV (20%), while those of the overestimated spectra are 59.4 keV (5%), 60.1 keV (10%), 60.7 keV (15%), and 62.5 keV (20%).

The training for the human phantom was performed using the base spectrum of 120 kV as described previously, while for testing, the images from the eight different 120-kV spectra that were applied during acquisitions and image reconstructions were used. We included noise and beam-hardening corrections. The predicted average RMSEs of the six elements are shown in Fig. 11. Compared with the case of the base spectrum, during testing, deviations of the spectrum (MAE) up to 20% had little impact on accuracy. Thus, even if the spectrum of a real CT machine is not precisely estimated, the DL-predicted results are still reliable. Again, the U-net model outperformed the VGG model.
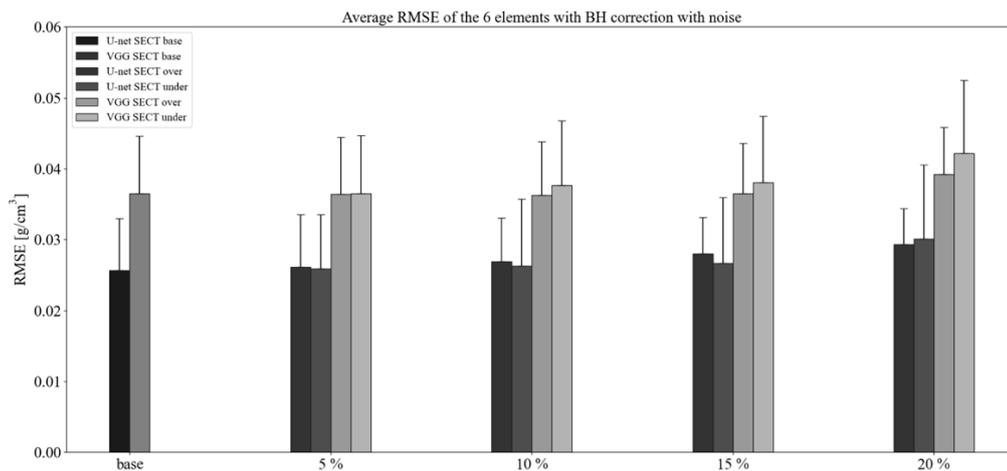
Fig. 11 Average root mean square errors of the six elements with eight different 120 kV energy spectra during the testing (with beam-hardening correction and noise) for the human phantom.

4. Discussion

In this study, we proposed a strategy for estimating elemental composition using DL without requiring a large cohort of real humans. Based on the virtual CT system and the estimated energy spectrum of a real CT machine, we performed training and testing using reconstructed images from digital phantoms. It was observed that, in the DL framework, the results were insensitive to noise, beam-hardening effects, and spectrum variations, indicating that the strategy was appropriate.

Extant model-based MD algorithms require either DECT or multi-energy CT images [26,36,43–47]; SECT's ability to differentiate soft tissue is inherently limited, because the attenuation coefficient can overlap considerably between different body materials at a fixed energy spectrum. Nevertheless, as photon-counting detectors have only recently been introduced into CT practice and the cost of DECT is relatively high, SECT is still the most commonly implemented form of CT in clinics, though the use of DECT has increased.

The DL-based approach in this study realised accurate MD using SECT. This is because DL extracts not only the values of the attenuation coefficients but also the shape information and even higher-order characteristics of human phantoms with complicated anatomical structures. Consequently, the proposed DL-based method has great potential for clinical applications, and it provides an additional solution for MD.

The above discussion should not be taken to imply that a higher-energy-multiplicity CT image does not significantly improve the prediction accuracy. In fact, as seen in Fig. 5, the DECT DL model was more accurate than the SECT DL model, and the QECT DL model was more accurate than either. Rather, we emphasize that the present study reveals the importance of structural information in addition to quantitative CT values in MD prediction.

Most MD methods only decompose tissues, such as fat, blood, and bone [26,43–45]. Very few investigations have focused on macro-elements until recently [36,47]. The proposed DL-based method has great potential for application in radiodiagnosis and radiotherapy treatment planning. In a previous study, Hünemohr et al. [36] proposed the derivation of elemental mass fractions by a linear fit of the EAN and EDR via DECT. However, the calculations were only performed using the reference CT numbers of the biological tissues. Moreover, as this study has shown, the use of DECT has weaknesses due to noise. This result may have significant implications for particle-therapy treatment planning.

By contrast, in this work, we have actually performed MD prediction, advancing the approach of Hünemohr et al. by considering the entire process, including phantom creation, CT scan, image reconstruction, model development, and model training and testing. The simulation environment makes it feasible to study the effects of noise, beam hardening, and other clinical issues, such as the influence of different CT machines, energy spectra, and projection numbers. Our approach can overcome the problem of small data cohorts, which appears frequently in DL-based approaches, to create various human phantoms via a Gaussian distribution model for material density in each anatomy. This can be further developed by combining it with non-rigid deformation to create a general human-phantom library [48–50] in the way that we created the model test datasets.

In the method proposed by Hünemohr et al. [36], the EAN and EDR were calculated using the known elemental compositions of each biological tissue. In a clinical situation, these quantities are obtained using the DECT method [51–53], which, however, has been proven to be sensitive to variations in CT number [32,54]. Therefore, MD becomes a two-step approach (CT number – EAN and EDR – elemental fractions). Error propagation will further render the results inaccurate. As

a result, image-domain methods always require high-quality reconstructed images as inputs. The DL-based approach, however, is insensitive to the influence of noise and beam-hardening effects, as shown in Fig. 5. Moreover, it is a one-step process that directly connects the CT image(s) to the elemental fractions. These two aspects ensure that our method is valuable for clinical applications.

Our study has several limitations. First, the estimated densities of C and O still had relatively large errors. This is because the CT images (reference biological tissues) of the whole body in the human phantom (in the BT phantom) were used for training and testing. The CT values of the brain white matter and prostate are almost the same, but the densities of C and O in these two organs differ greatly. The C (O) density of the former is approximately 0.20 g/cm^3 (0.69 g/cm^3), while for the latter, the density is 0.09 g/cm^3 (0.81 g/cm^3). As suggested by Fig. 10, a possible way to improve our results is to use a specific region only, that is, to use the limited images or biological tissues from the pelvic region in the training and testing. Second, the proposed strategy is still incomplete. The base spectra in our study were not obtained from real CT machines, and well-trained DL models have not yet been clinically applied. The present study only examined the feasibility of this strategy. Future studies should use a real CT machine; after training, the DL model should be tested in the clinic using real CT images as inputs to identify abnormal tissues and dose calculations for hadron therapy.

5. Conclusions

We proposed an MD strategy to estimate the density distributions of the macro-elements via deep learning. Through a simulation study, we found that the U-net model can estimate elemental compositions precisely, even with single-energy CT images as inputs. The results are insensitive to the influence of noise, beam-hardening effects, and spectrum variations. These observations demonstrate that our method is potentially useful for clinical applications.

Acknowledgements

Conflicts of Interest

The authors have no relevant conflicts of interest to disclose.

References

[1] Wu W, Hu D, An K *et al* 2021 A high-quality photon-counting CT technique based on weight adaptive total-variation and image-spectral tensor factorization for small animals imaging *IEEE Trans. Instrum. Meas.* 70 1–14. Art no. 2502114. 10.1109/TIM.2020.3026804.

[2] Xu M, Hu D, Luo F *et al* 2021 Limited-angle X-ray CT reconstruction using image gradient ℓ₀-norm with dictionary learning *IEEE Trans. Radiat. Plasma Med. Sci* 5(1) 78–87. 10.1109/TRPMS.2020.2991887.

[3] Luo F, Li W, Tu W, Wu W 2018 Adaptive weighted total variation minimization based alternating direction method of multipliers for limited angle CT reconstruction *IEEE Access* 6 64225–64236. 10.1109/ACCESS.2018.2873713.

[4] Joshi ML, Sahani DV, Kambadakone AR *et al* 2010 Effective atomic number accuracy for kidney stone characterization using spectral CT *J. Med. Imaging*: *Physics of Medical Imaging*, 7622 1136–1147. 76223K. 10.1117/12.844372.

[5] Deng K, Liu C, Ma R *et al* 2009 Clinical evaluation of dual-energy bone removal in CT angiography of the head and neck: Comparison with conventional bone-subtraction CT angiography *Clin Radiol.* 64(5) 534–541. 10.1016/j.crad.2009.01.007, Pubmed:19348851.

[6] Mendonça P R, Lamb P, Sahani D V 2014 A flexible method for multi-material decomposition of dual-energy CT images *IEEE Trans Med Imaging.* 33(1) 99–116. 10.1109/TMI.2013.2281719, Pubmed:24058018.

[7] Brown M, Parks P B 1969 Neutron radiography in biologic media. Techniques, observations, and implications *Am J Roentgenol Radium Ther Nucl Med.* 106(3) 472–485. 10.2214/ajr.106.3.472, Pubmed:5306140.

[8] Schneider U, Pedroni E, Lomax A 1996 The calibration of CT Hounsfield units for radiotherapy treatment planning *Phys Med Biol.* 41(1) 111–124. 10.1088/0031-9155/41/1/009, Pubmed:8685250.

[9] Wohlfahrt P, Richter C. 2020 Status and innovations in pre-treatment CT imaging for proton therapy *Br J Radiol.* 93(1107) 20190590. 10.1259/bjr.20190590.

[10] Ducros N, Abascal J F P J, Sixou B, Rit S, Peyrin F 2017 Regularization of nonlinear decomposition of spectral x-ray projection images *Med Phys.* 44(9) e174–e187. 10.1002/mp.12283, Pubmed:28901616.

[11] Le Huy Q, Molloi S 2011 Least squares parameter estimation methods for material decomposition with energy discriminating detectors *Med Phys.* 38(1) 245–255. 10.1118/1.3525840, Pubmed:21361193.

[12] Long Y, Fessler J A 2014 Multi-material decomposition using statistical image reconstruction for spectral CT *IEEE Trans Med Imaging.* 33(8) 1614–1626. 10.1109/TMI.2014.2320284, Pubmed:24801550.

[13] Zhao W, Vernekohl D, Han F *et al* 2018 A unified material decomposition framework for quantitative dual- and triple-energy CT imaging *Med Phys.* 45(7) 2964–2977. 10.1002/mp.12933, Pubmed:29679500.

[14] Clark D P, Badea C T 2014 Spectral diffusion: An algorithm for robust material decomposition of spectral CT data *Phys Med Biol.* 59(21) 6445–6466. 10.1088/0031-9155/59/21/6445, Pubmed:25296173.

[15] Persson M, Huber B, Karlsson S *et al* 2014 Energy-resolved CT imaging with a photon-counting silicon-strip detector *Phys Med Biol.* 59(22) 6709–6727. 10.1088/0022-3727/59/22/6709, Pubmed:25327497.

[16] Zhao W, Niu T, Xing L *et al* 2016 Using edge-preserving algorithm with non-local mean for significantly improved image-domain material decomposition in dual-energy CT *Phys*

Med Biol. 61(3) 1332–1351. 10.1088/0031-9155/61/3/1332, Pubmed:26796948.

[17] Mory C, Sixou B, Si-Mohamed S, Boussel L, Rit S 2018 Comparison of five one-step reconstruction algorithms for spectral CT *Phys Med Biol.* 63(23) 235001. 10.1088/1361-6560/aaeaf2.

[18] Rit S, Mory C, Noël P B 2020 Image formation in spectral computed tomography *Spectral Photon Counting Comput Tomogr Technol Appl.* 355.

[19] Wang G 2016 A perspective on deep imaging *IEEE Access.* 4 8914–8924. 10.1109/ACCESS.2016.2624938.

[20] Shen L, Zhao W, Xing L 2019 Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning *Nat Biomed Eng.* 3(11) 880–888. 10.1038/s41551-019-0466-4, Pubmed:31659306.

[21] Zeleznik R, Foldyna B, Eslami P *et al* 2021 Deep convolutional neural networks to predict cardiovascular risk from computed tomography *Nat Commun.* 12(1) 715. 10.1038/s41467-021-20966-2, Pubmed:33514711.

[22] Fu F, Wei J, Zhang M *et al* 2020 Rapid vessel segmentation and reconstruction of head and neck angiograms using 3D convolutional neural network *Nat Commun.* 11(1) 4829. 10.1038/s41467-020-18606-2, Pubmed:32973154.

[23] Shin H C, Roth H R, Gao M *et al* 2016 Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning *IEEE Trans Med Imaging.* 35(5) 1285–1298. 10.1109/TMI.2016.2528162, Pubmed:26886976.

[24] Clark D P, Holbrook M, Badea C T 2018 Multi-energy CT decomposition using convolutional neural networks *J. Med. Imaging*: *Physics of Medical Imaging* 10573 1057310.

[25] Abascal J F P J, Ducros N, Pronina V *et al* 2021 Material Decomposition in Spectral CT using deep learning: A Sim2Real transfer approach *IEEE Access.* 9 25632–25647. 10.1109/ACCESS.2021.3056150.

[26] Chen Z, Li L 2019 Robust multimaterial decomposition of spectral CT using convolutional neural networks *Opt Eng.* 58(1). 10.1117/1.OE.58.1.013104, 013104.

[27] Hasegawa Y, Haga A, Sakata D *et al* 2021 Estimation of X-ray energy spectrum of cone-beam computed tomography scanner using percentage depth dose measurements and machine learning approach *J Phys Soc Jpn.* 90(7). 074801.

[28] Sidky E Y, Yu L, Pan X, Zou Y, Vannier M 2005 A robust method of x-ray source spectrum estimation from transmission measurements: Demonstrated on computer simulated, scatter-free transmission data *J Appl Phys.* 97(12). 10.1063/1.1928312, 124701.

[29] Duan X, Wang J, Yu L, Leng S, McCollough C H 2011 CT scanner x-ray spectrum estimation from transmission measurements *Med Phys.* 38(2) 993–997. 10.1118/1.3547718, Pubmed:21452736.

[30] Ha W, Sidky E Y, Barber R F, Schmidt T G, Pan X 2019 Estimating the spectrum in computed tomography via Kullback–Leibler divergence constrained optimization *Med Phys.* 46(1) 81–92. 10.1002/mp.13257, Pubmed:30370544.

[31] Li K W, Fujiwara D, Haga A, Liu H, Geng L S 2021 Physical density estimations of single- and dual-energy CT using material-based forward projection algorithm: A simulation study *Br J Radiol.* 94(1128) 20201236. 10.1259/bjr.20201236.

[32] Li K-W, Fujiwara D, Haga A, Liu H, Geng L-S 2021 kV–kV and kV–MV DECT based estimation of proton stopping power ratio—A simulation study *Physcia Medica.* 89 182–192.

[33] International Commission on Radiation Units and Measurements *Tissue Substitutes in Radiation Dosimetry and Measurement* Volume 2. ICRU; 1989.

[34] White D R, Woodard H Q, Hammond S M 1987 Average soft-tissue and bone models for use in radiation dosimetry *Br J Radiol.* 60(717) 907–913. 10.1259/0007-1285-60-717-907, Pubmed:3664185.

[35] Woodard H Q, White D R 1986 The composition of body tissues *Br J Radiol.* 59(708) 1209–1218. 10.1259/0007-1285-59-708-1209, Pubmed:3801800.

[36] Hünemohr N, Paganetti H, Greilich S, Jäkel O, Seco J 2014 Tissue decomposition from dual energy CT data for MC based dose calculation in particle therapy *Med Phys.* 41(6) 061714. 10.1118/1.4875976.

[37] Clark K, Vendt B, Smith K *et al* 2013 The Cancer Imaging Archive (TCIA): Maintaining and operating a public information repository, *Journal of Digital Imaging.* 26 1045–1057. https://www.cancerimagingarchive.net/

[38] Klein S, Staring M, Murphy K *et al* 2009 Elastix: a toolbox for intensity-based medical image registration *IEEE Transactions on Medical Imaging*, 29(1) 196-205.

[39] Takahashi R, Matsubara T, Uehara K 2019 Data augmentation using random image cropping and patching for deep CNNs *IEEE Trans Circuits Syst Video Technol.* 30(9) 2917–2931. 10.1109/TCSVT.2019.2935128.

[40] Ruchala K J, Olivera G H, Schloesser E A *et al* 1999 Megavoltage CT on a tomotherapy system *Phys Med Biol* 44 2597–2621.

[41] Chen G P, Tai A, Puckett L *et al* 2021 Clinical implementation and initial experience of real-time motion tracking with jaws and MLC during helical tomotherapy delivery *Practical Radiation Oncology* 11(5) e486–e495. 10.1016/j.prro.2021.01.010

[42] Punnoose J, Xu J, Sisniega A, Zbijewski W, Siewerdsen JH 2016 Technical note: Spektr 3.0—A computational tool for x-ray spectrum modeling and analysis *Med Phys.* 43(8) 4711. 10.1118/1.4955438, Pubmed:27487888.

[43] Mechlem K, Ehn S, Sellerer T *et al* 2018 Joint statistical iterative material image reconstruction for spectral computed tomography using a semi-empirical forward model *IEEE Trans Med Imaging.* 37(1) 68–80.

[44] Shen C, Li B, Lou Y, Yang M, Zhou L, Jia X 2018 Multienergy element-resolved cone beam CT (MEER-CBCT) realized on a conventional CBCT platform *Med Phys.* 45(10) 4461–4470. 10.1002/mp.13169, Pubmed:30179261.

[45] Tilley S, Zbijewski W, Stayman J W 2019 Model-based material decomposition with a penalized nonlinear least-squares CT reconstruction algorithm *Phys Med Biol.* 64(3) 035005. 10.1088/1361-6560/aaf973.

[46] Zhang W, Zhang H, Wang L *et al* 2019 Image domain dual material decomposition for dual-energy CT using butterfly network *Med Phys.* 46(5) 2037–2051. 10.1002/mp.13489, Pubmed:30883808.

[47] Badea C T, Holbrook M, Clark D P 2018 Multi-energy CT decomposition using convolutional neural networks *Phys Med Imaging.*

[48] Choi C, Yeom YS, Lee H *et al* 2020 Body-size-dependent phantom library constructed from ICRP mesh-type reference computational phantoms *Phys Med Biol*. 65(12) 125014.

[49] Akhavanallaf A, Xie T, Zaidi H 2019 Development of a library of adult computational phantoms based on anthropometric indexes *IEEE Trans Radiat Plasma Med Sci*. 3(1) 65–75.

[50] Sun X, Wang H, Ristaniemi T 2020 Registration-based construction of a whole-body human phantom library for anthropometric modeling *IEEE 2020 42nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)* 2396–2399.

[51] Torikoshi M, Tsunoo T, Sasaki M *et al* 2003 Electron density measurement with dual-energy x-ray CT using synchrotron radiation *Phys Med Biol*. 48(5) 673–685. 10.1088/0031-9155/48/5/308, Pubmed:12696802.

[52] Tsunoo T, Torikoshi M, Ohno Y *et al* 2004 Measurement of electron density and effective atomic number using dual-energy X-ray CT *IEEE Nuclear Science Symposium Conference Record*.

[53] Bazalova M, Carrier JF, Beaulieu L, Verhaegen F 2008 Dual-energy CT-based material extraction for tissue segmentation in Monte Carlo dose calculations *Phys Med Biol*. 53(9) 2439–2456. 10.1088/0031-9155/53/9/015, Pubmed:18421124.

[54] Yang M, Virshup G, Clayton J, Zhu XR, Mohan R, Dong L 2011 Does kV-MV dual-energy computed tomography have an advantage in determining proton stopping power ratios in patients? *Phys Med Biol*. 56(14) 4499–4515. 10.1088/0031-9155/56/14/017, Pubmed:21719949.