



## Material Decomposition in Spectral CT using deep learning: A Sim2Real transfer approach

Juan F P J Abascal, Nicolas Ducros, Simon Rit, Pierre-Antoine Rodesch, Thomas Broussaud, Suzanne Bussod, Philippe Douek, Andreas Hauptmann, Simon Arridge, Françoise Peyrin

### ► To cite this version:

Juan F P J Abascal, Nicolas Ducros, Simon Rit, Pierre-Antoine Rodesch, Thomas Broussaud, et al.. Material Decomposition in Spectral CT using deep learning: A Sim2Real transfer approach. 2020. hal-02952707

HAL Id: hal-02952707

<https://hal.archives-ouvertes.fr/hal-02952707>

Preprint submitted on 29 Sep 2020

**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.

# Material Decomposition in Spectral CT using deep learning: A Sim2Real transfer approach

Juan FPJ Abascal<sup>1</sup>, Nicolas Ducros<sup>1</sup>, Simon Rit<sup>1</sup>, Pierre-Antoine Rodesch<sup>1</sup>, Thomas Broussaud<sup>1</sup>, Suzanne Bussod<sup>1</sup>, Philippe Douek<sup>1</sup>, Andreas Hauptmann<sup>2,3</sup>, Simon Arridge<sup>3</sup>, and Françoise Peyrin<sup>1</sup>

<sup>1</sup>Univ Lyon, INSA-Lyon, CNRS, Inserm, CREATIS UMR 5220, U1206, Lyon, France

<sup>2</sup>Research Unit of Mathematical Sciences, University of Oulu, Oulu, Finland

<sup>3</sup>Department of Computer Science, University College London, London, United Kingdom

September 29, 2020

## Abstract

The state-of-the art for solving the nonlinear material decomposition problem in spectral computed tomography is based on variational methods, but these are computationally slow and critically depend on the particular choice of the regularization functional. Convolutional neural networks have been proposed for addressing these issues. However, learning algorithms require large amounts of experimental data sets. We propose a deep learning strategy for solving the material decomposition problem based on a U-Net architecture and a Sim2Real transfer learning approach where the knowledge that we learn from synthetic data is transferred to a real-world scenario. In order for this approach to work, synthetic data must be realistic and representative of the experimental data. For this purpose, numerical phantoms are generated from human CT volumes of the KiTS19 Challenge dataset, segmented into specific materials (soft tissue and bone). These volumes are projected into sinogram space in order to simulate photon counting data, taking into account the energy response of the scanner. The network is trained to decompose the materials in the projection domain after which we apply any conventional tomographic method to reconstruct the different material volumes. The proposed decomposition method is compared to a regularized Gauss-Newton (RGN) method on synthetic data, experimental phantom data and human thorax data.

## 1 Introduction

The new generation of spectral computed tomography (SPCCT) scanners include photon-counting detectors (PCDs), which count single photons and resolve their energy [1]. With this extra dimension, SPCCT provides higher contrast with respect to conventional CT and allows for material decomposition, which opens up new diagnosis possibilities [2], [3]. PCDs can be exploited for material decomposition, including K-edge imaging, which uses the discontinuity at diagnostic energies of the linear attenuation coefficient of high-Z element-based contrast agents such as gadolinium, gold or bismuth. Material decomposition could be used to quantify calcium content to assess bone and teeth health, determine kidney stone composition, determine liver iron concentrations, determine bone marrow composition, characterize plaques [3], [4], abdominal imaging [5], evaluating the risk of breast cancer [6], and k-edge contrast agent quantification, among others.

A variety of approaches for material reconstruction have been investigated, which can be divided into projection-based [7], [8], image-based [9], [10], and one-step inversion approaches [11], [12]. Image-based methods reconstruct first each energy bin and then perform material decomposition on the image domain assuming that images are monochromatic, which may lead to beam hardening artefacts. Projection-based and one-step inversion methods are a natural choice as they take into account the inherent physics. One-step methods are mathematically the most elegant, as they solve the problem in one step, but they can be computationally expensive. Projection-based methods firstly decompose the energy-resolved sinograms into material-specific sinograms and then perform tomographic reconstruction of each of the decomposed sinograms [13]. An advantage of this approach is that it allows independent decomposition of each projection view, which can be parallelized, making the computational process more convenient [13].

Material decomposition in spectral CT is a nonlinear, nonconvex inverse problem [13], [14]. State-of-the art for solving this problem are variational approaches [13, 15–17] or iterative regularization schemes [14]. In [13], the authors proposed a regularized material-specific Gauss-Newton (RGN) method, which outperformed widely used unregularized maximum likelihood method. In [14], authors assessed an iterative regularization scheme based on the Bregman iteration. While variational methods are robust and have shown improved results with respect to unregularized approaches, they present several disadvantages. First, they can be computationally slow, given their iterative nature. Second, their performances rely on the prior knowledge of the scanner energy response and the choice of the regularization functional. With regard to the adopted model, for instance, the detector response can be degraded by charge sharing, pulse pileup, and energy loss due to K-escape [18]. The effect of assuming a perturbed detector response can have detrimental effects on image quality, as shown in [19]. With regard to the choice of the regularization functional, specific regularization for the decomposed materials have shown to improve image quality with respect to unregularized approaches [13], but the choice is far from being optimal.

Recently, deep convolutional neural networks (CNN) have shown outstanding results in several image processing tasks [20], [21]. CNNs have been also proposed for solving several inverse problems, including image restoration and image reconstruction [22–27]. Among the diverse CNN configurations, U-Net has shown outstanding results for image segmentation [28] and has been applied to postprocessing in CT [22] and to artifact removal of nonlinear CT images [29]. Several works have also proposed deep learning for solving the material decomposition, mainly in the image domain, using U-Net in [30] and VGG-16 in [31]. The previously proposed neural networks approaches in the projection domain have been based on multilayer perceptrons (using fully connected layers) for decomposing materials in a pixel-by-pixel basis [32], [33]. In [32], the authors used a neural network with two hidden layers followed by a denoising method to mitigate noise.

The recent success of CNNs for image processing tasks relies on large-scale annotated data sets such as ImageNet with millions of images [34], [20]. However, equivalent medical imaging data sets are not common. Transfer learning (TL) has emerged as a promising candidate technique to compensate the lack of large-scale medical imaging data sets. TL is the ability of a system or model to recognize and apply knowledge and skills learned in previous domains/tasks to novel domains/tasks. TL assumes that the model can generalize to new tasks or new domains by learning features or factors that are common between source and target tasks [35]. TL becomes relevant when labelled training data, time or computational power are limited, so it is particularly applicable to medical imaging. Different approaches for TL can be found in [36], [37]; most common is to pre-train a model on a large-annotated data set and then to fine tune the model on a more specific and smaller data set [38], [39]. Recent works proposed related TL approaches based on fine tuning for material decomposition [19, 33]. However, in this work we follow a Sim2Real transfer approach where the knowledge that one learns from synthetic data is transferred to a real-world scenario [40], [41]. The Sim2Real strategy is based on the fact that synthetic data is almost infinitely accessible and diverse, and it becomes pertinent when experimental annotated data are very limited or non-existent, which is the case of SPCCT. While learning-based approaches perform reasonably well given experimental training data, they may not necessarily work well when trained only on synthetic data. In cases in which synthetic data generalize poorly to real data, data augmentation strategies are generally followed.

In this work, we propose a deep learning strategy for solving the projection-based material decomposition problem based on a U-Net architecture and a Sim2Real transfer approach. In order for the Sim2Real approach to work, synthetic data must be realistic and representative of the experimental data. For this purpose, numerical phantoms are generated from human CT volumes of the KiTS19 Challenge dataset [42]. Numerical data are then computed by segmenting CT volumes into specific materials (soft tissue and bone), projecting the material volumes and simulating photon counting data, taking into account the energy response of the scanner. The U-Net is trained to decompose the materials in the projection domain and is compared to a regularized Gauss-Newton (RGN) method [13] using both numerical and experimental data. After material decomposition of the photon counting data, we apply a conventional tomographic method to reconstruct the different material volumes. The proposed approach combines a CNN model with training on human data, which allows to implicitly learn the real prior distribution of the decomposed materials [43]. This is particularly interesting from a Bayesian perspective to inverse problems as it motivates the choice of the regularization functional as a particular selection for the prior distribution of the unknown variable. Thus, our method differs from previous CNN methods based on the image domain and approaches that are trained on experimental phantom data only.

## 2 Methods

### 2.1 Forward model

We assume a 2D sensor with  $P$  pixels and  $I$  energy bins and an object with  $V$  voxels made of  $M$  materials. We image the object under  $\Theta$  projections. Let  $\mathbf{s} = (s_{1,1}^1, \dots, s_{i,p}^\theta, \dots, s_{I,P}^\Theta)^\top$  be the measurement vector, where  $s_{i,p}^\theta$  represents the photon counts measured in the  $i$ -th energy bin at the  $p$ -th pixel under the  $\theta$ -th projection, and  $\boldsymbol{\rho} = (\rho_{1,1}, \dots, \rho_{m,v}, \dots, \rho_{M,V})^\top$  be the (unknown) mass densities vector, where  $\rho_{m,v}$  is the mass density for the  $m$ -th material at the  $v$ -th voxel. Solving the spectral CT problem in one step requires the inversion of the nonlinear system

$$\mathbf{s} = \mathcal{G}(\boldsymbol{\rho}) \quad (1)$$

where  $\mathcal{G}$  represents the forward model that maps  $\boldsymbol{\rho}$  onto  $\mathbf{s}$ .

An alternative to solving the problem (1) in one step is the two-step approach [13, 15]. In this case, the forward model  $\mathcal{G}$  can be seen as the composition of the linear X-ray transform and a non-linear spectral mixing operator. The X-ray transform  $\mathcal{X}$  applies to each material independently, i.e.,

$$\mathbf{a}_m = \mathcal{X}(\boldsymbol{\rho}_m), \quad 1 \leq m \leq M \quad (2)$$

where  $\boldsymbol{\rho}_m = (\rho_{m,1}, \dots, \rho_{m,v}, \dots, \rho_{m,V})^\top$  and  $\mathbf{a}_m = (a_{m,1}^1, \dots, a_{m,p}^\theta, \dots, a_{m,P}^\Theta)^\top$  represent mass density and projected mass density for the  $m$ -th material, respectively. Spectral mixing applies to each view and detector pixel independently, i.e.,

$$\mathbf{s}^\theta = \mathcal{F}(\mathbf{a}^\theta), \quad 1 \leq \theta \leq \Theta \quad (3)$$

where  $\mathbf{a}^\theta = (a_{1,1}^\theta, \dots, a_{m,p}^\theta, \dots, a_{M,P}^\theta)^\top$  and  $\mathbf{s}^\theta = (s_{1,1}^\theta, \dots, a_{i,p}^\theta, \dots, a_{I,P}^\theta)^\top$ . In particular, we consider the following non linear mixing [15]

$$s_{i,p}^\theta = \int_{E \in \mathcal{E}} n_{0,p}(E) d_{i,p}(E) \exp \left[ - \sum_{m=1}^M a_{m,p}^\theta \tau_m(E) \right] dE \quad (4)$$

where  $\mathcal{E}$  is the range of energy delivered by the x-ray tube,  $n_{0,p}(E)$  is the x-ray source energy spectrum at the  $p$ -th pixel,  $d_{i,p}(E)$  is the energy-dependent detector response function at the  $p$ -th pixel for the  $i$ -th bin, and  $\tau_m(E)$  is the energy-dependent mass attenuation coefficient of the  $m$ -th material.

### 2.2 Spectral CT image reconstruction

Solving the inverse spectral CT image reconstruction problem in one step involves inverting the  $\mathbb{R}^{MV} \rightarrow \mathbb{R}^{IP\Theta}$  nonlinear mapping (1). In this work, we instead adopt the two-step approach, which splits the problem into  $\Theta$  inversions of  $\mathbb{R}^{MP} \rightarrow \mathbb{R}^{IP}$  mappings (3) followed by  $M$  inversions of  $\mathbb{R}^V \rightarrow \mathbb{R}^{P\Theta}$  (2). The two-step approach reduces the size of the subproblems, offering a natural parallelization scheme (e.g. decomposing all angles in parallel, reconstructing all materials in parallel), and allows to separate nonlinear (i.e. spectral) from linear (i.e. tomographic) mixing. Thus, the two-step approach allows to reduce the computational complexity with respect to the one-step method by first solving the material decomposition problem (3), projection by projection, and then solving the tomographic reconstruction problem for each material (2).

In this work, we consider the material decomposition problem with two different approaches: a regularized variational approach and a deep learning strategy, as described in the next two subsections. Material decomposition is done projection by projection, which allows to include an explicit regularization functional or to implicitly learn the prior distribution. A previous work found that a regularization approach led to significant improvement with respect to a unregularized pixel-by-pixel decomposition, which was very sensitive to noise [13]. An alternative to solving the material decomposition problem projection by projection would be to decompose all projection views at once. While this could allow to introduce further regularization on the angle dimension, we chose single projection decomposition as it allows for parallelization across angles, reducing computational complexity.

#### 2.2.1 Material decomposition with a variational method

Using a variational framework, material decomposition has been previously formulated as the minimization of the following cost function [13]:

$$\Phi(\mathbf{a}^\theta) = \frac{1}{2} \|\mathbf{s}^\theta - \mathcal{F}(\mathbf{a}^\theta)\|_{W^\theta}^2 + \alpha \sum_m \mathcal{R}_m(\mathbf{a}^\theta), \quad 1 \leq \theta \leq \Theta \quad (5)$$

where the data fidelity term is a weighted least square functional to approximate a Poisson distribution,  $W^\theta = \text{diag}(1/\sqrt{\mathbf{s}^\theta})$  is a weighting matrix,  $\alpha$  is a regularization parameter, and  $\mathcal{R}_m$  accounts for specific material regularization. We chose second-order and first-order Tikhonov regularization for soft tissue and bone, respectively, as suggested in [13]. While two different values of the regularization parameter could be used for different materials, it was previously found to not lead to significant improvement [13]. Cost function (5) is efficiently minimized by using a Gauss Newton algorithm (RGN); MATLAB code for RGN is available from the Spectral X-ray image reconstruction SPRAY toolbox [44].

### 2.2.2 Material decomposition with the proposed deep learning method

Deep learning has been recently proposed for inverse problems [22, 45]. In this work, we want to learn the mapping

$$\mathcal{H}_\beta : \tilde{\mathbf{s}}^\theta \mapsto \tilde{\mathbf{a}}^\theta, \quad (6)$$

where  $\beta$  indicates the parameters of the neural network,  $\tilde{\mathbf{s}} = \ln(\hat{\mathbf{s}}/\mathbf{s}^\theta)$  corresponds to the normalized data with  $\hat{\mathbf{s}} = \mathcal{F}(0)$  representing the measurement in the absence of the object, and  $\tilde{\mathbf{a}}$  corresponds to the normalized outputs. We normalize the output material per material across the entire training set, i.e.,  $\tilde{a}_{m,p}^{\theta,n} = a_{m,p}^{\theta,n}/\max_{p,\theta,n} a_{m,p}^{\theta,n}$ .

Normalization of input and output variables prior to training a neural network model is common practice to avoid a slow or unstable learning process. In this case, normalization of the outputs is motivated by the fact that different materials may have different range of values. For the inputs, photon-counting data have very large values; in addition; normalizing the data by the acquisition in the absence of the object,  $\hat{\mathbf{s}}$ , may decrease the sensitivity to modeling errors.

Learning means finding the parameters  $\beta$  of the network that minimizes the following loss function

$$\Xi(\beta) = \sum_{n=1}^N \|\mathcal{H}_\beta(\tilde{\mathbf{s}}^n) - \tilde{\mathbf{a}}^n\|^2 = \sum_{n=1}^N \sum_{m=1}^M \|\mathcal{H}_{\beta,m}(\tilde{\mathbf{s}}^n) - \tilde{\mathbf{a}}_m^n\|^2, \quad (7)$$

where  $(\tilde{\mathbf{s}}^n, \tilde{\mathbf{a}}^n)$  are  $N$  input-output vector pairs (3) that can be reshaped into input-output array pairs of size  $(P_x \times P_y \times I, P_x \times P_y \times M)$ ,  $N$  is the number of projection images in the training set, and each projection image is of size  $P_x \times P_y$ . Minimization of (7) was done with Adam method under TensorFlow, with learning rate  $10^{-5}$  and batch size of 16. Training and test losses were computed during training and early stopping was adopted to avoid overfitting.

We use a U-Net architecture [28], which consists in a contracting multi-scale decomposition path and a symmetric expanding path with skip connections at each scale. The contracting path comprises  $3 \times 3$  convolutions ('same' padding convolutions) each followed by a rectified linear unit (ReLU) and alternated with a  $2 \times 2$  max pooling operation with stride 2 for downsampling by half every two convolution operations. At each downsampling the number of feature channels is doubled, with 32 channels in the first scale and 128 channels in the last scale. The expansive path comprises the same convolution operations as in the contracting path but alternated with upsampling by two every two convolution operations. The final layer is a  $1 \times 1$  convolution used to map the 32 channel layer to the output with the desired  $M$  number of decomposed materials. This leads to a total of 14 hidden layers (figure 1).

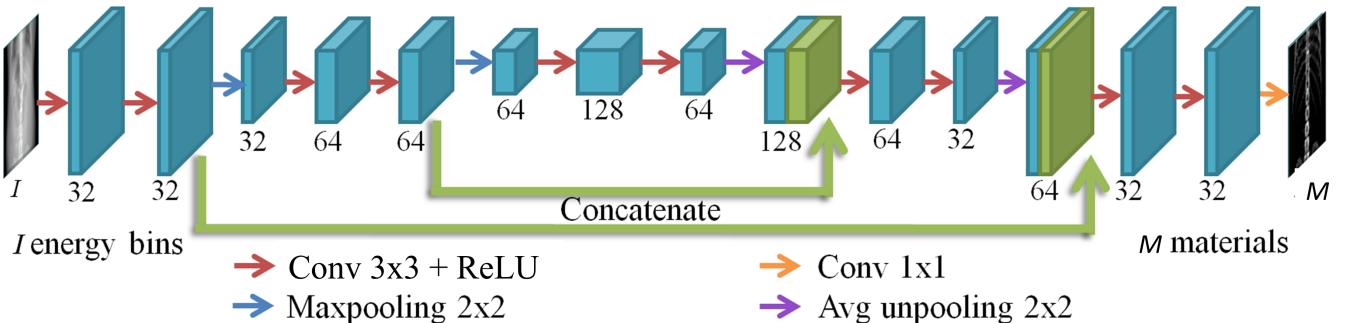


Figure 1: U-Net architecture for material decomposition. Input is a stack of photon counting projection images for  $I$  energy bins ( $P_x \times P_y \times I$ ), and output is a stack of the decomposed material projection images for  $M$  materials ( $P_x \times P_y \times M$ ) for a projection angle  $\theta$ .

### 2.2.3 Tomographic reconstruction

Tomographic reconstruction solves the problem (2). For numerical data, we assume parallel geometry and perform tomographic reconstruction using filtered back-projection algorithm. For experimental data, tomographic reconstruction is done using a regularized conjugate gradient method within the Reconstruction Toolkit (RTK) [46], which minimizes the following cost function:

$$\Psi(\boldsymbol{\rho}_m) = \|\mathcal{X}\boldsymbol{\rho}_m - \mathbf{a}_m\|_2^2 + \gamma\|\nabla\boldsymbol{\rho}_m\|_2^2, \quad (8)$$

where  $\gamma$  is a regularization parameter. The forward projection uses ray casting with trilinear interpolation and the backprojection is voxel-based with bilinear interpolation. Parameters were selected empirically as  $\gamma = 100$  and  $k = 25$ , where  $k$  is the number iterations of conjugate gradient.

## 2.3 Data and image evaluation

### 2.3.1 Numerical human phantoms

Numerical human phantoms are made of soft tissue and bone and are built from CT volumes obtained from the KiTS19 challenge data set (2019 Kidney Tumor Segmentation Challenge, <https://kits19.grand-challenge.org/data/>) [42] (figure 2). We used 50 phantoms for training, 3 for evaluation, and 9 for test. CT volumes have voxel size  $0.96 \text{ mm} \times 0.96 \text{ mm} \times 1.8 \text{ mm}$ . In order to facilitate numerical simulations, we processed CT volumes to get the same dimension: volumes were cropped to have 640 voxels in  $x$ - and  $y$ -dimensions and 100 voxels along  $z$ -axis. CT volumes were semi-automatically segmented into bone and soft-tissue using a combination of thresholding and morphological operations. After segmentation, we created mass density volumes by normalizing the segmented volumes to have median value in the kidneys equal to  $1.1 \text{ g}\cdot\text{cm}^{-3}$ . Then, mass densities for both materials were projected using the Radon transform (assuming parallel geometry) to create the projected mass densities.

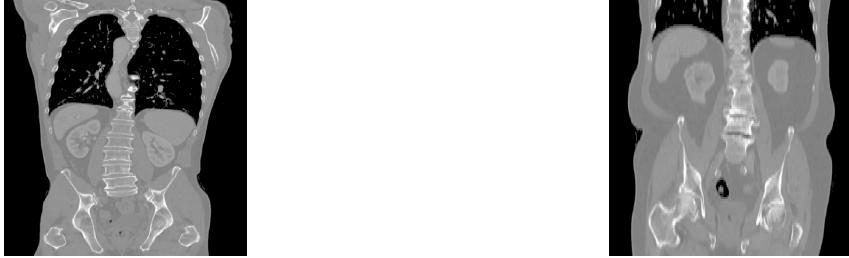


Figure 2: Two examples from the CT Kits19 data base.

Photon counting data were simulated taking into account the manufacturer source spectrum and detector response function and the linear attenuation coefficients (LACs) of the two materials. Both phantom creation and simulation were realized with the SPRAY toolbox [13, 44]. Noisy data was considered by assuming a Poisson distribution for a tube current of 200 mA, which corresponds to a total number of photons in a central pixel of the detector equal to  $N_0 = 6 \cdot 10^5$ . Projected data consisted of 360 projections over a 180 angle span and  $924 \times 8$  detector pixels.

Training data for U-Net then consisted of 18000 projections (360 projections for 50 phantoms). In order to perform further data augmentation, projected volumes were randomly cropped along the  $z$ -dimension and data simulation and noise corruption were performed at runtime during training.

### 2.3.2 Experimental data

A SPCCT prototype system (Philips Medical Systems) installed at CERMÉP, Lyon, was used for this study [47]. It is a 500 mm FOV system equipped with a conventional x-ray tube that can be set with a tube voltage at 80, 100, or 120 kVp and tube current values between 10 and 500 mA. Tube filtration absorbs low-energy x-rays so the subject is exposed to a spectrum that ranges from 35 to 120 keV. The system is based on photon counting detectors of 2-mm-thick cadmium zinc telluride with a pixel pitch of  $270 \mu\text{m} \times 270 \mu\text{m}$  at isocenter, ChromAIX2 application-specific integrated circuits combined with cadmium zinc telluride as sensor material, and operates in single photon-counting mode with energy discrimination. Photo-counting detectors allow up to 5 consecutive energy

thresholds between 30 and 120 keV. The scan field of view is 500 mm in-plane, with a z-coverage of 17.5 mm in the scanner isocenter. Axial and helical scans over 360° are performed with 2400 projections per rotation.

We consider two data sets: an experimental phantom and human thorax data. For the experimental phantom, we used the CIRS (Computerized Imaging Reference Systems, Inc, USA) Electron Density Phantom [48], which is made of a cylinder of size 330 mm × 270 mm × 50 mm and a density of 1.029 g·cm<sup>-3</sup> and inserts as surrogates of different biological tissues. We used a mixture of inserts from the CIRS phantom and the GAMMEX Tissue Characterization Phantom Model 467 (Gammex Middleton, USA). Data were acquired using a tube voltage of 120 kVp and tube current of 198 mA using axial geometry, with a standard set of energy thresholds (30, 51, 62, 72, 81 keV). Data were binned to a final size of 462 × 16 × 2400 × 5. Final reconstructed voxel size was 1 mm × 1 mm × 2 mm. In order to compare the data SNR with previous work [19], we compute the total number of photons provided by the source in a central pixel of the detector. For this data set, the total number of photons was approximately  $N_0 = 6 \cdot 10^5$ .

The human thorax data set corresponds to an adult male volunteer. Data were acquired using a tube voltage of 120 kVp and exposure of 62 mAs using helical geometry (85 mA, 0.75 s rotation time and 1.03 pitch). Data were binned to a final size of 924 × 8 × 24912 × 5. This corresponded to a total number of photons of  $N_0 = 1.8 \cdot 10^5$ .

We remark that data calibrations and corrections included in the scanner manufacturer pipeline have not been included in this work.

### 2.3.3 Image evaluation

U-Net and RGN were evaluated on synthetic data and on the two experimental data sets. The performance of RGN depends on the selection of the regularization parameter. Selecting the optimal regularization parameter in terms of MSE may lead to images with low noise but with loss of image details [14]. Thus, we provide results for RGN for two values of the regularization parameter:  $\alpha = 0.1$  (low regularization) and  $\alpha = 0.6$  (intermediate regularization). For the numerical test set, methods were assessed in terms of normalized Mean Squared Error (MSE) and Structural Similarity metric (SSIM) on both decomposed projections and reconstructed images. Normalized MSE was computed for each material as follows:

$$\text{MSE} = \frac{\|\mathbf{a}_m - \mathbf{a}_m^{\text{Target}}\|_2}{\|\mathbf{a}_m^{\text{Target}}\|_2}. \quad (9)$$

In addition, methods were assessed by visual inspection. For the experimental phantom data, methods were assessed in terms of quantification of soft tissue and bone inserts by computing the mean in a circular region with radius of 1 cm for a central slice. Both experimental phantom data and human thorax data were evaluated visually. For experimental data, we also computed a 'virtual' monochromatic image (for 60 keV) from the decomposed images in order to compare visually to the standard CT image provided by a conventional scanner. We chose 60 keV as a previous study found that monochromatic images between 60 and 70 keV could be considered as optimal surrogates of conventional CT images [49]. Monochromatic images were computed as the linear attenuation coefficient at energy  $E$  and point  $\mathbf{x}$  in the reconstructed volume:

$$\mu(E, \mathbf{x}) = \sum_{m=1}^M \rho_m(\mathbf{x}) \tau_m(E). \quad (10)$$

## 3 Results

### 3.1 Numerical thorax phantom data

Figure 3 shows decomposed images by RGN for two different values of  $\alpha$  and by U-Net. For  $\alpha = 0.1$ , which corresponds to little regularization, RGN provides accurate decomposition, recovering image details, but images are noisy. For  $\alpha = 0.6$ , RGN reduces noise but at the expense of excessive blurring, particularly in bone regions. This is well appreciated in the error images in figure 4. On the contrary, U-Net is able to reduce noise while maintaining image details.

Tomographic reconstructions of decomposed data with RGN and U-Net are shown in figure 5. These results are consistent with results obtained for the decomposed projections. For  $\alpha = 0.1$ , RGN recovers image details well but images are noisy. For  $\alpha = 0.6$ , RGN is able to remove noise but image details are lost. U-Net leads to similar reduction of noise than RGN for  $\alpha = 0.6$  while maintaining image quality, as it is well appreciated in the error images in figure 6.

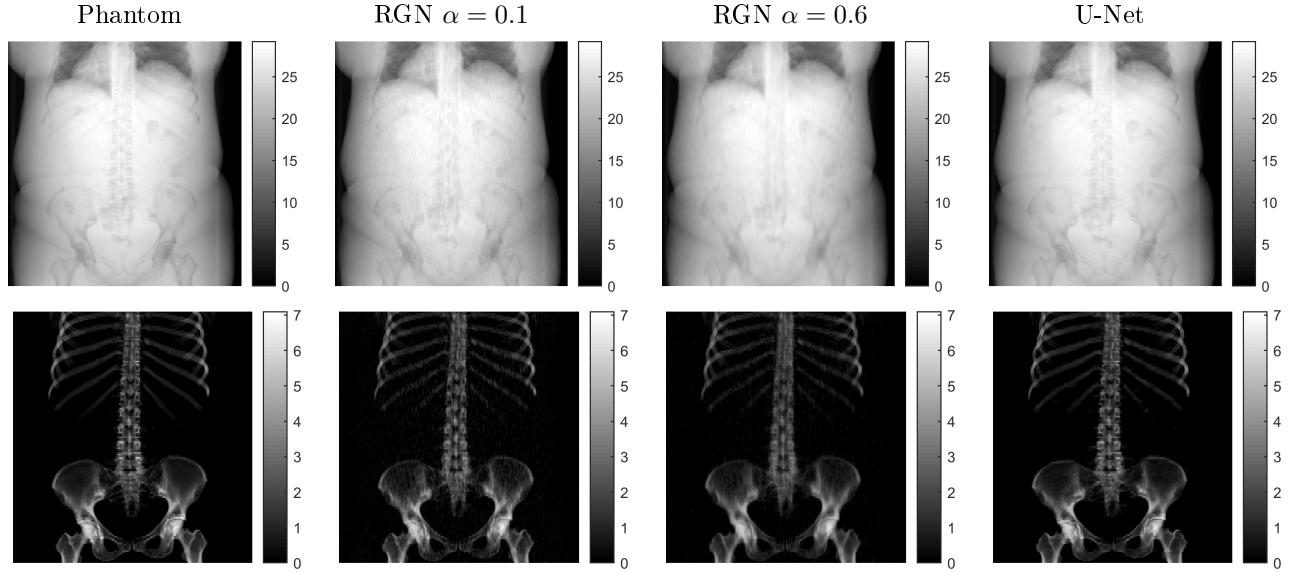


Figure 3: Phantom and decomposed projections (for  $0^\circ$  view) for RGN and U-Net for a synthetic phantom from the test set. Top: soft tissue. Bottom: Bone.

Figure 7 shows image details of the reconstructed images for soft tissue. RGN for  $\alpha = 0.6$  misses image details while U-Net presents better recovery than RGN for  $\alpha = 0.1$ . While the image quality of U-Net is generally superior, it presents some 'fake' structures in the soft tissue image that seem to originate from noisy data (shown by green arrows in figure 7). Artefacts also appear in the RGN image for the larger regularization parameter but they are slightly more apparent for U-Net because of the overall reduction of noise. In addition to these artefacts, RGN present large errors in some bone locations.

Figure 8 displays normalized MSE and SSIM for decomposed projections and tomographic reconstructions for U-Net and RGN. For soft tissue, U-Net presents a 29 % and 6 % decrease in normalized MSE on decomposed projections with respect to RGN for  $\alpha = 0.1$  and  $\alpha = 0.6$ , respectively. Normalized MSE for bone is larger than for soft tissue due to noise, as most part of the image corresponds to background. U-Net led to similar decrease in MSE for bone as for soft tissue. U-Net also led to increase in SSIM with respect to RGN for bone and soft tissue. For bone tissue, U-Net presents a 2.5- and 3-fold increase in SSIM on decomposed projections with respect to RGN for  $\alpha = 0.1$  and  $\alpha = 0.6$ , respectively. For tomographic reconstructions, U-Net leads to similar improvements in MSE and SSIM.

### 3.2 Experimental data

#### 3.2.1 Experimental phantom data

Figure 9 shows decomposed sinograms and figure 10 shows tomographic reconstructions, for the experimental phantom data. Methods lead to similar quantitative results, especially for soft tissue inserts. Decomposition of bone materials are slightly different. In terms of image quality, U-Net presents less noisy images and slightly less artefacts than RGN. The 60 keV images provided by the two methods are very similar, with U-Net presenting less noise.

Figure 11 shows quantification of the phantom inserts using the ROIs shown in figure 10. Methods provide similar quantification overall but there is a drift between RGN and U-Net for the recovery of bone like inserts. Figure 11(left) shows the quantification of all inserts from the decomposed soft image. Both methods show accurate quantification (Pearson's correlation coefficient  $r > 0.98$ ) for soft tissue-like inserts (displayed with symbol +). We remark that bone-like inserts (displayed with symbol  $\times$ ) are not close to zero, as they are comprised of both soft tissue and bone.

Figure 11(center) shows the quantification of all inserts from the decomposed bone image. Both methods show accurate quantification ( $r > 0.98$ ) for soft bone-like inserts ( $\times$ ). In this case, bone material is around zero for soft tissue-like inserts (+).

Figure 11(right) shows the quantification of all inserts from the 60 keV image. RGN and U-Net show almost

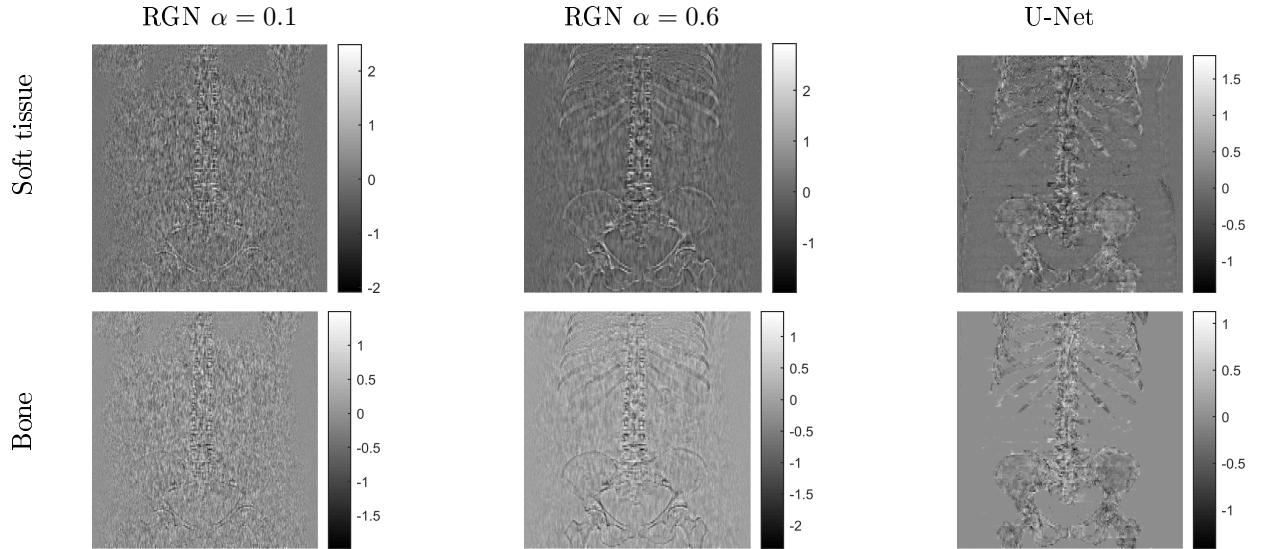


Figure 4: Error images of decomposed projections by RGN and U-Net for the same view as in figure 3.

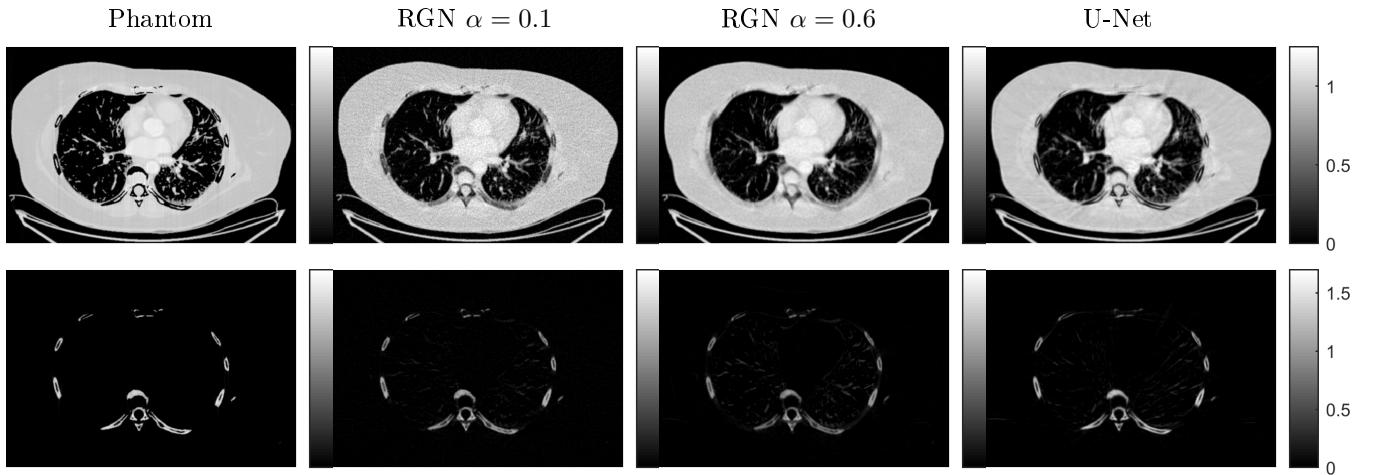


Figure 5: Tomographic reconstruction of a numerical human phantom for RGN and U-Net. Image reconstruction corresponding to decomposed projections shown in figure 3. Top: soft tissue. Bottom: Bone.

identical results.

### 3.2.2 Human thorax data

Figure 12 shows tomographic reconstructions and the 60 keV image for RGN and U-Net. RGN provides noisier decomposed images and appears more sensitive to ring artefacts than U-Net. In terms of decomposed materials, U-Net seems to have more bone material in the soft tissue image than RGN. Compared to the decomposed images, 60 keV images do not show ring artefacts. Zoomed details show that ring artefacts are more apparent for RGN than for U-Net. U-Net also appears to provide slightly finer recovery of bone details.

## 3.3 Computation time

We run our code on a Windows computer with a 64-bit operating system, Intel Xeon(R) E5-1650 v4 3.60 GHz CPU and 128 GB RAM. The U-Net is trained on a GeForce NVIDIA GTX 1080 Ti graphics card, which requires 6.5 days for 2000 epochs. Decomposition and tomographic reconstruction times for synthetic data and the human thorax data set are displayed in Table 1. Material decomposition with RGN is done on CPU by using straight-

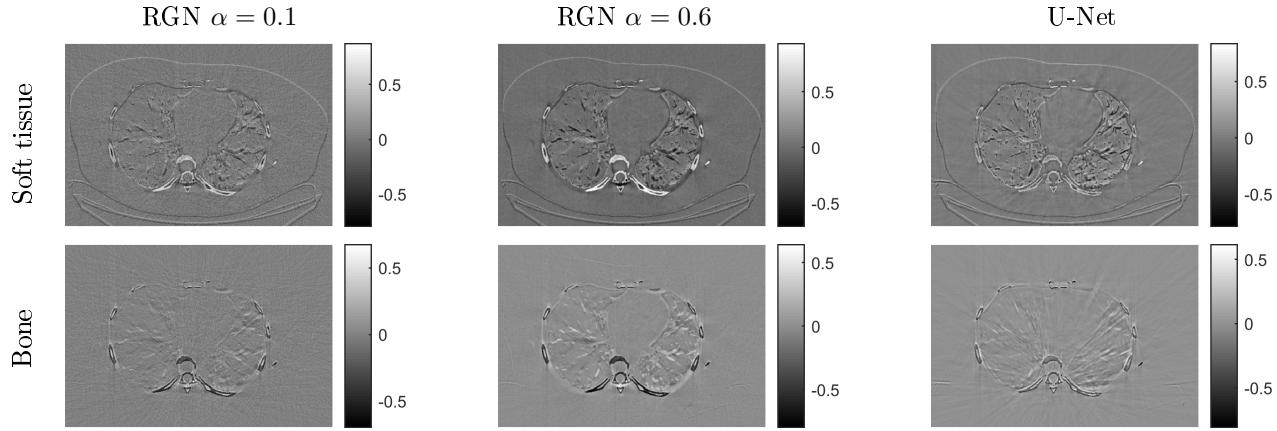


Figure 6: Error images of tomographic reconstructions of a numerical human phantom (corresponds to figure 5). Top: soft tissue. Bottom: Bone.

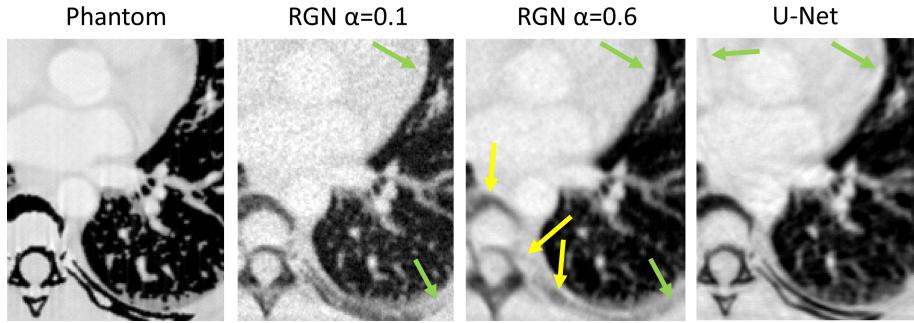


Figure 7: Detail of tomographic reconstruction of soft tissue for the images shown in figure 5. Yellow arrow points where bone details are not well recovered. Green arrow points to the appearance of 'fake' details.

forward parallelization on 4 cores. For U-Net, we provide decomposition times on CPU and GPU. Tomographic reconstruction is done on CPU for synthetic data and on GPU with RTK for the human thorax data set.

## 4 Discussion

We have proposed a deep learning approach based on a U-Net architecture for solving the material decomposition problem in spectral CT and compared it to a RGN method. Given that medical imaging data and, specifically, spectral CT data are scarce, we have followed a Sim2Real approach by training on simulated human phantoms and assessing on experimental phantom data and human data. We found that U-Net leads to improved image quality with respect to RGN, which presented a strong trade off between noise and regularization. Increasing the regularization parameter for RGN decreased noise but it led to loss of image details. On the contrary, U-Net was able to remove noise while maintaining image quality. In terms of quantification of decomposed materials, both methods led to similar results on the assessed experimental phantom data. Furthermore, U-Net led to a significant reduction in prediction time, specially for large data sets.

The superiority in terms of image quality of U-Net with respect to RGN must be discussed in detail. This can be explained by the fact that U-Net can learn implicitly the probability distribution of the decomposed materials. In fact, Bayesian deep learning interprets supervised learning as recovering the posterior distribution and it approximates the conditional mean when using the MSE loss [43]. This is particularly relevant for material decomposition in the projection domain, as in this case the choice of prior distributions is less clear. The RGN method included material specific regularization where regularization functionals for each material were selected as suggested in [13]. We used first- and second-order Tikhonov for bone and soft tissue, respectively. While this choice seems reasonable, it is not optimal. A second disadvantage of variational methods is the high dependence on the regularization parameter, which requires fine tuning. Here, we showed results for RGN for different values

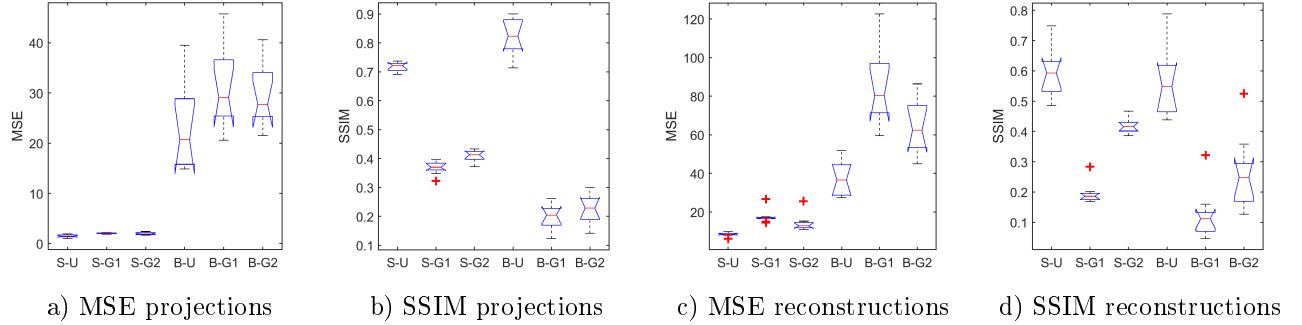


Figure 8: Quantitative results on the numerical thorax test set. Boxplots for normalized MSE and SSIM of decomposed projections and tomographic reconstructed images, given by U-Net (U), RGN for  $\alpha = 0.1$  (G1) and RGN for  $\alpha = 0.6$  (G2), for soft tissue (S) and bone (B). Boxplots show the median, 25-th and 75-th percentiles, and minimum and maximum values excluding outliers, which are displayed with red crosses. When notches do not overlap, one can conclude, with 95% confidence, that the true medians do differ.

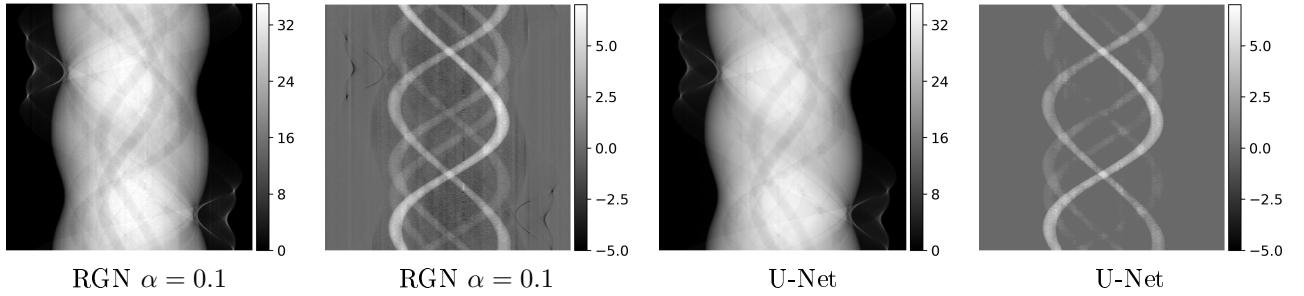


Figure 9: Material decomposed sinograms of soft tissue and bone for the experimental phantom with Gammex and Cirs inserts for RGN and U-Net.

of the regularization parameter. For further insight into the selection of the regularization parameter for RGN, we refer to [13] and [14]. Although fine tuning can generally lead to an acceptable compromise between noise and regularization for variational approaches, its benefits are mitigated by the fact that regularization functionals are not optimal.

In this work, we have proposed a learning approach for material decomposition in the projection domain which has been the preferred approach for empirical calibration methods and for model based approaches, as they are supported by the physics. Empirical calibration methods used low-order polynomials to perform material decomposition in a pixel-by-pixel basis. Similarly, model-based approaches such as Maximum-Likelihood estimation have been used for material decomposition in a pixel-by-pixel basis. However, these approaches do not exploit the information of neighbor pixels as in the regularization approaches. In [50], RGN method was shown to outperform a Maximum-Likelihood method, which was very sensitive to noise. In this work, we compare RGN to a CNN method, which implicitly regularizes the problem, as previously discussed.

Several deep learning approaches have been previously proposed for material decomposition. Most of them perform material decomposition in the image domain [30, 31]. These approaches reconstruct first each energy bin and then perform material decomposition on the image domain assuming that images are monochromatic, which may lead to beam hardening artefacts. However, CNN learning approaches may correct for this. A comparison between CNN projection-based and CNN image-based decomposition methods is still an open question but it is out of the scope of this work. Recently a multilayer perceptron was proposed for solving the material decomposition problem in the projection domain [33]. In this work, we proposed a CNN approach to assess the potential of deep learning not only for solving the material decomposition problem but for implicit regularization. We have also compared deep learning with a variational approach while previous works compared their methods with more classical approaches, eg. calibration methods. In addition, we have shown that a Sim2Real approach that learns on numerical human phantoms is feasible. This is particularly interesting for a CNN method, as it can implicitly learn the optimal regularization for human data. We remark that previous approaches have learned on experimental

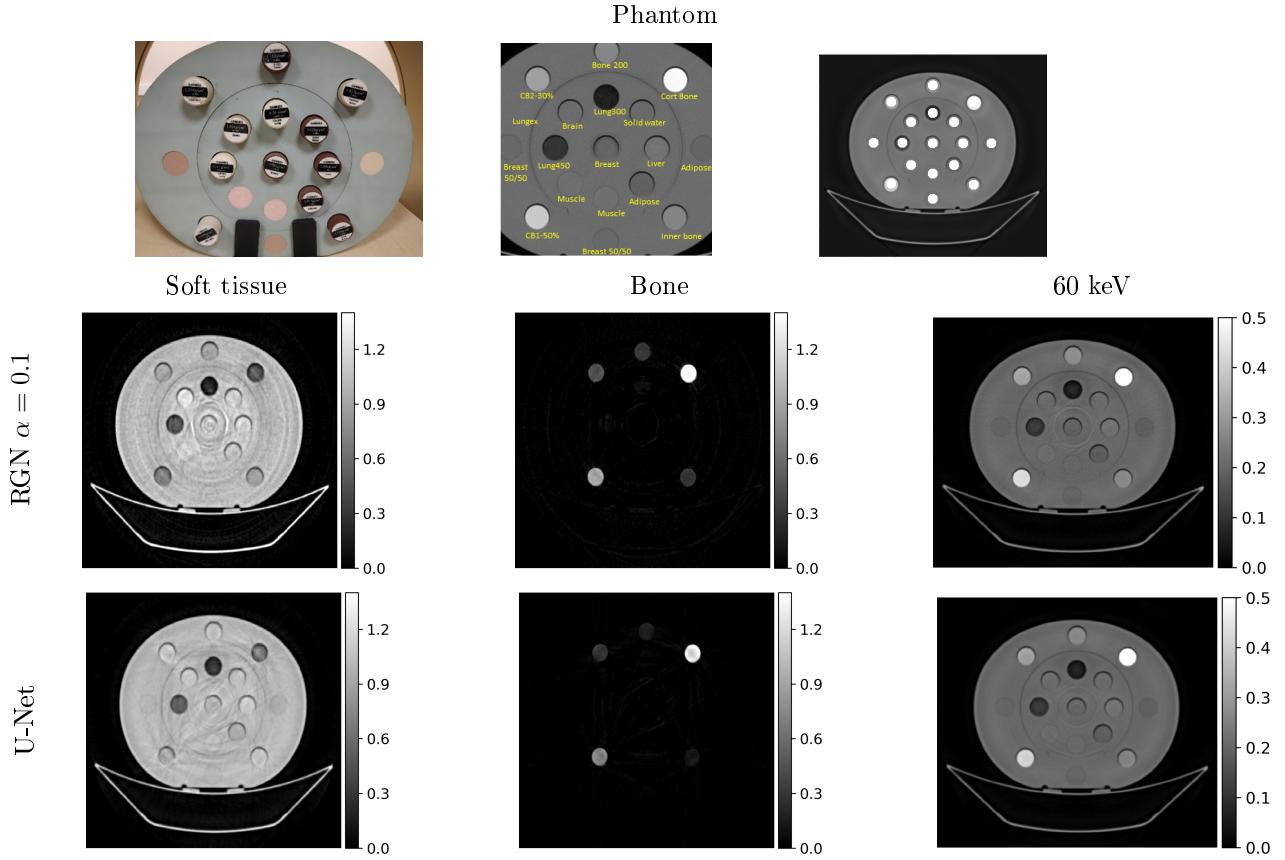


Figure 10: Experimental phantom with (Gammex and Cirs) inserts. Top row: Photo of the phantom, standard CT image with insert labels and standard CT image with ROIs used for quantification. Second and third rows: Tomographic reconstruction of soft tissue and bone and 60 keV image for RGN (second row) and U-Net (third row). We remark that data calibrations and corrections included in the scanner manufacturer pipeline have not been included in this work.

cylindrical-shape phantoms, which misses this feature of the CNN method.

The Sim2Real transfer approach belongs to the wider class of TL and domain adaptation approaches [36, 37, 51]. ‘Transferability’ has been shown to depend on the distance between source and target specific tasks [52]. In our case, the task is the same but domains may differ if the feature spaces or the marginal probability distributions between source and target are different. In the case of learning on numerical human phantoms and testing on human data, we do not expect large differences between source and target learned features and probability distributions, which may explain why we found positive results. In the case of testing on experimental phantoms, we found probability distributions to be different (results not shown). This could be expected as objects are different. We also remark that due to the fact that we are projecting mass densities, then the material decomposition problem in the projection domain depends also on the object size. Although distributions were different, the range of values of the thorax data set comprised those of the experimental phantom, which may explain why it works. In order to achieve better generalization, we also tried further data augmentation by modifying sizes of the thorax phantoms. This led to similar results, with an increase in training time, but this techniques could lead to models with improved generalization. The features that are learned from a human phantom or from a phantom with inserts should be different, so including these type of phantoms in the training data may improve the performance on these type of experimental phantom data sets. For experimental data, algorithms were subject to ring and band artefacts, as typical data calibrations and corrections included in the scanner manufacturer pipeline have not been included in this work. Interestingly, U-Net was more robust to the presence of artefacts than RGN. This may be because for U-Net we use normalized data while for RGN we used absolute data. Another explanation could be that U-Net learns the prior distribution of the decomposed materials, so it may penalize for these type of artefacts. In previous work, we also found this effect when decomposing simulated perturbed data. U-Net was found to be less sensitive to artefacts even when perturbed data has not been considered during training. In addition, including perturbed data

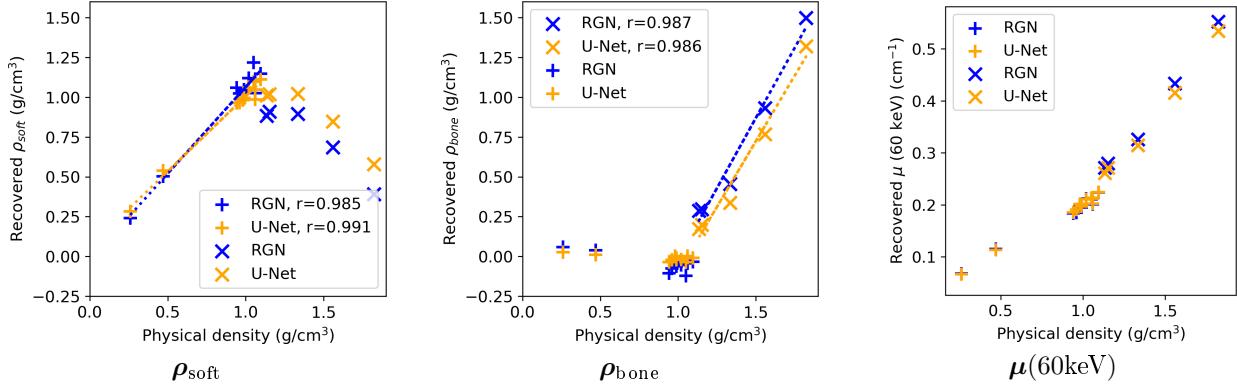


Figure 11: Quantification of (Gammex and Cirs) inserts for RGN (blue) and U-Net (orange) from decomposed soft tissue image  $\rho_{\text{soft}}$  (left), decomposed bone image  $\rho_{\text{bone}}$  (center) and 60 keV image  $\mu(60\text{keV})$  (right). Inserts are marked with the symbol + for soft tissue like materials (lung, breast, muscle, solid water, adipose tissue, brain, muscle) and with the symbol  $\times$  for bone like materials (cortical bone (CB), inner bone, CB 30%, CB 50%).

Table 1: Material decomposition (Mat. Dec.) and tomographic reconstruction (Tomo. Rec.) times for synthetic data and for the human thorax data set. CPU, 4 cores: Parallelization using 4 cores. BS = Batch Size.

	Simulated data		Human thorax	
	Mat. Dec.	Tomo. Rec.	Mat. Dec.	Tomo. Rec.
RGN (CPU, 4 cores)	44 min	25 s	155 min	50 min (GPU)
U-Net (CPU)	70 s	" "	25 min	" "
U-Net (GPU)	6 s (BS = 16)	" "	45 s (BS = 1024)	" "

during training led to almost complete removal of artefacts, as expected. Nevertheless, as most of these artefacts can be considerably reduced with appropriate data calibration, we do not address this issue in this work.

This work is subject to a few limitations. Results may depend on the learning model. Here we have used a U-Net architecture, which has led to outstanding results for diverse CT applications [22, 29]. However, for this problem a simpler architecture may be sufficient. The material decomposition has a pixel-wise nature, so it is possible that simpler ConvNets can learn the prior distributions of the materials without the need of a hierarchical-based architecture, which is more designed to interpret information across the entire image. Besides the architecture, deep iterative methods, also called model-based learning methods, proposed for solving inverse problems could provide further benefits [24, 27, 53]. In particular, these iterative methods include a data consistency condition, which could mitigate the appearance of artefacts that can potentially lead to mislead diagnosis. Actually, in this work we found that both RGN with large regularization or U-Net can lead to 'fake' artefacts that may look real, given the improved quality in the surrounding regions. Although the proposed method based on U-Net leads to overall high image quality, deep iterative methods may be required to provide learning methods that are robust for the clinical use. These methods can also reduce further the need of learning power as well as the required amount of training data. Another interesting solution could be to consider Bayesian neural networks that also provide an uncertainty estimate [43]. With regard to the amount of training data, we have not evaluated the minimum requirements for this model. In addition, we have assumed that the energy response of the scanner is perfectly known. Although this has been provided by the manufacturer, some deviations from ideality conditions may occur. Further work could also fine tune a pretrained network using experimental phantom data as suggested in [19]. However, a previous work found only small improvements using a fully-connected network [33]. In this work, we have assessed the feasibility of the proposed method for material decomposition for two materials only. However, this methodology can be easily extended to three or more materials, including k-edge imaging, as shown on numerical data in [19].

In conclusion, deep learning methods show a great potential for spectral CT as they provide almost real time material decomposition and yield competitive results in comparison with current variational approaches.

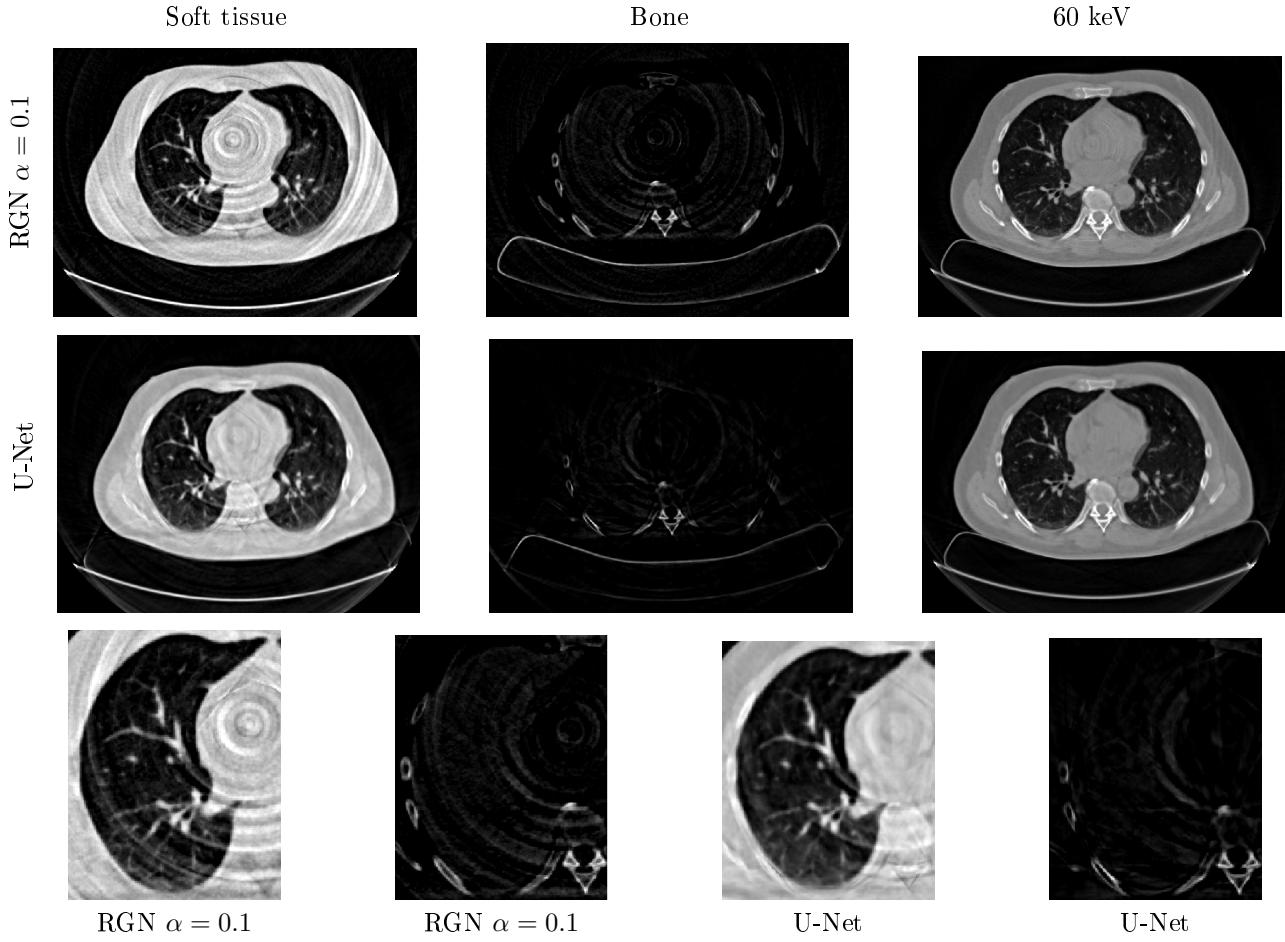


Figure 12: Top: Tomographic reconstruction of a human thorax for RGN and U-Net. Gray scales are (0, 35), (0, 15), and (0, 0.4) for soft tissue, bone and 60 keV, respectively. We remark that data calibrations and corrections included in the scanner manufacturer pipeline have not been included in this work. Bottom: zoomed detail for soft tissue and bone.

## Acknowledgment

We gratefully acknowledge support from the CNRS-IN2P3 Computing Center (Lyon - France) for providing computing and data-processing resources needed for this work. We thank Philips GmbH Innovative Technologies, Hamburg, Germany, for giving access to experimental data. We specially thank to Klaus Erhard, Bernhard Brendel and Yoad Yagil for their support.

This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 701915. It was also performed within the framework of the LabEx PRIMES (ANR-11-LABX-0063) of University de Lyon and under the support of the ANR project SALTO (ANR-17-CE19-0011-01). The project has also received funding from the European Union’s Horizon 2020 research and innovation programme under the grant agreement N 668142. This work was partly funded by France Life Imaging (grant ANR-11-INBS-0006) from the French Investissements d’Avenir. This work was partially supported by the Academy of Finland Project 312123 (Finnish Centre of Excellence in Inverse Modelling and Imaging, 2018–2025). SA acknowledges support of EPSRC grants EP/N022750/1 and EP/T000864/1.

## References

- [1] E Roessl and R Proksa. K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors. *Physics in Medicine and Biology*, 52(15):4679, 2007.

- [2] R E Alvarez and A Macovski. Energy-selective reconstructions in X-ray computerised tomography. *Physics in Medicine and Biology*, 21(5):733, 1976.
- [3] David P Cormode, Ewald Roessl, Axel Thran, Torjus Skajaa, Ronald E Gordon, Jens-Peter Schlomka, Valentin Fuster, Edward A Fisher, Willem J M Mulder, Roland Proksa, and Zahi A Fayad. Atherosclerotic Plaque Composition: Analysis with Multicolor CT and Targeted Gold Nanoparticles1. *Radiology*, 256(3):774–782, 2010.
- [4] P Baturin, Y Alivov, and S Mollo. SU-C-301-06: Assessment of Plaque Vulnerability Using Spectral CT and Two Contrast Materials. *Medical Physics*, 38(6Part2):3379–3379, jun 2011.
- [5] Salim Si-Mohamed, Arnaud Thivolet, Pierre-Emmanuel Bonnot, Daniel Bar-Ness, Vahan Képénékian, David P Cormode, Philippe Douek, and Pascal Rousset. Improved Peritoneal Cavity and Abdominal Organ Imaging Using a Biphasic Contrast Agent Protocol and Spectral Photon Counting Computed Tomography K-Edge Imaging. *Investigative Radiology*, page 1, may 2018.
- [6] Huanjun Ding, Bo Zhao, Pavlo Baturin, Farnaz Behroozi, and Sabee Mollo. Breast tissue decomposition with spectral distortion correction: A postmortem study. *Medical Physics*, 41(10):101901, sep 2014.
- [7] Huy Q Le and Sabee Mollo. Least squares parameter estimation methods for material decomposition with energy discriminating detectors. *Med. Phys.*, 38(1):245, 2011.
- [8] Y Long and J A Fessler. Multi-Material Decomposition Using Statistical Image Reconstruction for Spectral CT. *IEEE Transactions on Medical Imaging*, 33(8):1614–1626, 2014.
- [9] Darin P Clark and Cristian T Badea. Spectral diffusion: an algorithm for robust material decomposition of spectral CT data. *Physics in Medicine and Biology*, 59(21):6445–6466, oct 2014.
- [10] Mats Persson, Ben Huber, Staffan Karlsson, Xuejin Liu, Han Chen, Cheng Xu, Moa Yveborg, Hans Bornefalk, and Mats Danielsson. Energy-resolved CT imaging with a photon-counting silicon-strip detector. *Physics in Medicine and Biology*, 59(22):6709, 2014.
- [11] Cyril Mory, Bruno Sixou, Salim Si-Mohamed, Loic Boussel, Simon Rit, ic Boussel, Claude Bernard Lyon, and Salim Si-mohamed. Comparison of five one-step reconstruction algorithms for spectral CT. *Physics in Medicine and Biology*, 63(23), 2018.
- [12] Simon Rit, Cyril Mory, and Peter B. Noël. Image Formation in Spectral Computed Tomography. In *Spectral, Photon Counting Computed Tomography*, pages 355–372. CRC Press, jun 2020.
- [13] Nicolas Ducros, Juan Felipe Perez-Juste Abascal, Bruno Sixou, Simon Rit, and Françoise Peyrin. Regularization of nonlinear decomposition of spectral x-ray projection images. *Medical Physics*, 44(9):e174–e187, sep 2017.
- [14] J F P J Abascal, N Ducros, and F Peyrin. Nonlinear material decomposition using a regularized iterative scheme based on the Bregman distance. *Inverse Problems*, 34(12):124003, dec 2018.
- [15] J P Schlomka, E Roessl, R Dorscheid, S Dill, G Martens, T Istel, C Baumer, C Herrmann, R Steadman, G Zeitler, A Livne, and R Proksa. Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography. *Physics in Medicine and Biology*, 53(15):4031, 2008.
- [16] Okkyun Lee, Steffen Kappler, Christoph Polster, and Katsuyuki Taguchi. Estimation of Basis Line-Integrals in a Spectral Distortion-Modeled Photon Counting Detector Using Low-Rank Approximation-Based X-Ray Transmittance Modeling: K-Edge Imaging Application. *IEEE Transactions on Medical Imaging*, 36(11):2389–2403, nov 2017.
- [17] Bernhard Brendel, Frank Bergner, Kevin Brown, and Thomas Koehle. Penalized likelihood decomposition for dual layer spectral CT. In *Proc. 4th Intl. Mtg. on image formation in X-ray CT*, 2016.
- [18] Katsuyuki Taguchi and Jan S Iwanczyk. Vision 20/20: Single photon counting x-ray detectors in medical imaging. *Medical Physics*, 40(10):100901, 2013.

- [19] J Abascal, N Ducros, V Pronina, S Bussod, A Hauptmannz, S Arridge, P Douek, and F Peyrin. Material Decomposition Problem in Spectral CT: A Transfer Deep Learning Approach. In *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*, pages 1–4, jul 2020.
- [20] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. In *NIPS’12 Proceedings of the 25th International Conference on Neural Information Processing Systems*, pages 1097–1105, Lake Tahoe, Nevada, 2012.
- [21] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, may 2015.
- [22] Kyong Hwan Jin, Michael T. McCann, Emmanuel Froustey, and Michael Unser. Deep Convolutional Neural Network for Inverse Problems in Imaging. *IEEE Transactions on Image Processing*, 26(9):4509–4522, sep 2017.
- [23] Guang Yang, Simiao Yu, Hao Dong, Greg Slabaugh, Pier Luigi Dragotti, Xujiong Ye, Fangde Liu, Simon Arridge, Jennifer Keegan, Yike Guo, and David Firmin. DAGAN: Deep De-Aliasing Generative Adversarial Networks for Fast Compressed Sensing MRI Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1310–1321, jun 2018.
- [24] Andreas Hauptmann, Felix Lucka, Marta Betcke, Nam Huynh, Jonas Adler, Ben Cox, Paul Beard, Sébastien Ourselin, and Simon Arridge. Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography. *IEEE Transactions on Medical Imaging*, 37(6):1382–1393, jun 2018.
- [25] Saiprasad Ravishankar, Jong Chul Ye, and Jeffrey A Fessler. Image Reconstruction: From Sparsity to Data-adaptive Methods and Machine Learning. Technical report, 2019.
- [26] Alice Lucas, Michael Iliadis, Rafael Molina, and Aggelos K Katsaggelos. Using Deep Neural Networks for Inverse Problems in Imaging: Beyond Analytical Methods. *IEEE Signal Processing Magazine*, 35(1):20–36, jan 2018.
- [27] Simon Arridge, Peter Maass, Ozan Öktem, and Carola Bibiane Schönlieb. Solving inverse problems using data-driven models. *Acta Numerica*, 28:1–174, may 2019.
- [28] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. pages 234–241. Springer, Cham, 2015.
- [29] Hyung Suk Park, Sung Min Lee, Hwa Pyung Kim, and Jin Keun Seo. Machine-learning-based nonlinear decomposition of CT images for metal artifact reduction. aug 2017.
- [30] Darin P Clark, Matthew Holbrook, and Cristian T Badea. Multi-energy CT decomposition using convolutional neural networks. In Guang-Hong Chen, Joseph Y. Lo, and Taly Gilat Schmidt, editors, *Medical Imaging 2018: Physics of Medical Imaging*, volume 10573, page 59. SPIE, mar 2018.
- [31] Zhengyang Chen and Liang Li. Robust multimaterial decomposition of spectral CT using convolutional neural networks. *Optical Engineering*, 58(01):1, jan 2019.
- [32] Mengheng Touch, Darin P Clark, William Barber, and Cristian T Badea. A neural network-based method for spectral distortion correction in photon counting x-ray CT. *Physics in Medicine and Biology*, 61(16):6132, 2016.
- [33] Kevin C. Zimmerman, Gayatri Sharma, Abdul Kareem Parchur, Amit Joshi, and Taly Gilat Schmidt. Experimental investigation of neural network estimator and transfer learning techniques for K-edge spectral CT imaging. *Medical Physics*, 47(2):541–551, feb 2020.
- [34] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 248–255. IEEE, jun 2009.
- [35] Y Bengio, A Courville, and P Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, aug 2013.
- [36] Hoo-Chang Shin, Holger R Roth, Mingchen Gao, Le Lu, Ziyue Xu, Isabella Nogues, Jianhua Yao, Daniel Mollura, and Ronald M Summers. Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning. *IEEE Transactions on Medical Imaging*, 35(5):1285–1298, may 2016.

- [37] Garrett Wilson and Diane J Cook. A Survey of Unsupervised Deep Domain Adaptation. Technical report, 2020.
- [38] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R. Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning? *IEEE Transactions on Medical Imaging*, 35(5):1299–1312, may 2016.
- [39] Ravi K Samala, Heang-Ping Chan, Lubomir Hadjiiski, Mark A Helvie, Jun Wei, and Kenny Cha. Mass detection in digital breast tomosynthesis: Deep convolutional neural network with transfer learning from mammography. *Medical physics*, 43(12):6654, dec 2016.
- [40] Carl Doersch and Andrew Zisserman. Sim2real transfer learning for 3D human pose estimation: motion to the rescue. In *d Conference on Neural Information Processing Systems (NeurIPS 2019)*, 2019.
- [41] Alexander Pashevich, Robin Strudel, Igor Kalevatykh, Ivan Laptev, and Cordelia Schmid. Learning to Augment Synthetic Images for Sim2Real Policy Transfer. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), 2019*, 2019.
- [42] Nicholas Heller, Niranjan Sathianathan, Arveen Kalapara, Edward Walczak, Keenan Moore, Heather Kaluzniak, Joel Rosenberg, Paul Blake, Zachary Rengel, Makinna Oestreich, Joshua Dean, Michael Tradewell, Aneri Shah, Resha Tejpaul, Zachary Edgerton, Matthew Peterson, Shaneabbas Raza, Subodh Regmi, Nikolaos Paniklopoulos, and Christopher Weight. The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes. Technical report, 2019.
- [43] Jonas Adler and Ozan Oktem. Deep Bayesian Inversion Computational uncertainty quantification for large scale inverse problems. Technical report, 2018.
- [44] Nicolas Ducros. SPRAY - Spectral X-ray image reconstruction toolbox, 2017.
- [45] Eunhee Kang, Junhong Min, and Jong Chul Ye. A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction. *Medical Physics*, 44(10):e360–e375, oct 2017.
- [46] S Rit, M Vila Oliva, S Brousmiche, R Labarbe, D Sarrut, and G Sharp. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). *Journal of Physics: Conference Series*, 489:012079, 2014.
- [47] Salim Si-Mohamed, Daniel Bar-Ness, Monica Sigovan, David P Cormode, Philippe Coulon, Emmanuel Coche, Alain Vlassenbroek, Gabrielle Normand, Loic Boussel, and Philippe Douek. Review of an initial experience with an experimental spectral photon-counting computed tomography system. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 873:27–35, nov 2017.
- [48] P. Pemler, U. Schneider, and J. Besserer. Evaluation of the electron density phantom CIRS Model 62. *Z Med Phys*, 11(1):25–32, 2001.
- [49] Nils Große Hokamp, R Gilkeson, M K Jordan, K R Laukamp, Victor-Frederick Neuhaus, S Haneder, S S Halliburton, and A Gupta. Virtual monoenergetic images from spectral detector CT as a surrogate for conventional CT images: Unaltered attenuation characteristics with reduced image noise. *European Journal of Radiology*, 117:49–55, aug 2019.
- [50] N. Ducros, J.F.P.-J. Abascal, B. Sixou, S. Rit, and F. Peyrin. Regularization of nonlinear decomposition of spectral x-ray projection images. *Medical Physics*, 44(9), 2017.
- [51] Maithra Raghu, Chiyuan Zhang, Jon Kleinberg, and Samy Bengio. Transfusion: Understanding Transfer Learning for Medical Imaging. In *Neural Information Processing Systems (NeurIPS), 2019*, feb 2019.
- [52] Hossein Azizpour, Ali Sharif Razavian, Josephine Sullivan, Atsuto Maki, and Stefan Carlsson. From generic to specific deep representations for visual recognition. In *2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 36–45. IEEE, jun 2015.
- [53] Jonas Adler and Ozan Oktem. Learned Primal-Dual Reconstruction. *IEEE Transactions on Medical Imaging*, 37(6):1322–1332, jun 2018.