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

Juan F P J Abascal, Juan Fpj Abascal, Nicolas Ducros, Valeriya Pronina, Simon Rit, Pierre-Antoine Rodesch, Thomas Broussaud, Suzanne Bussod, Philippe Douek, Andreas Hauptmann, et al.

IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2017.DOI

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

JUAN FPJ ABASCAL[1], NICOLAS DUCROS[1], (Member, IEEE), VALERIYA PRONINA[2,1], SIMON RIT[1], PIERRE-ANTOINE RODESCH[1], THOMAS BROUSSAUD[1], SUZANNE BUSSOD[1], PHILIPPE DOUEK[1], ANDREAS HAUPTMANN[3,4], SIMON ARRIDGE[4], FRANCOISE PEYRIN[1], (Member, IEEE)

[1]Univ Lyon, INSA-Lyon, CNRS, Inserm, CREATIS UMR 5220, U1206, Lyon, France (e-mail: juan.abascal@creatis.insa-lyon.fr)
[2]Skolkovo Institute of Science and Technology, Moscow, Russia
[3]Research Unit of Mathematical Sciences, University of Oulu, Oulu, Finland
[4]Department of Computer Science, University College London, London, United Kingdom

Corresponding author: Juan FPJ Abascal (e-mail: juan.abascal@creatis.insa-lyon.fr).

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-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.

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. We compared projection- and image-based decomposition approaches where the network is trained to decompose the materials either in the projection or in the image domain. The proposed Sim2Real transfer strategies are compared to a regularized Gauss-Newton (RGN) method on synthetic data, experimental phantom data and human thorax data.

INDEX TERMS Spectral CT, inverse problem, deep learning, transfer learning

I. 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 spectral CT image reconstruction methods have been proposed. Some of these methods have focused on compressed sensing and algorithms that promote sparsity in the image domain, using total variation (TV) [7], vector-valued generalization of TV [8], nonlocal TV [9], tight frames [10], dictionary learning methods [11], prior-based methods [12], and low-rank [13]–[15]. In this work, we focus on reconstruction algorithms that perform material decomposition.

Material reconstruction algorithms can be divided into projection-based [16], [17], image-based [18], [19], and one-step inversion approaches [20], [21]. 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 [22]. 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 [22].

State-of-the art for solving the material decomposition problem are optimization methods, also named variational approaches [20], [22]–[26]. Image-based material decomposition is a convex problem for a least squares functional [21]. Material decomposition in the projection domain is a nonlinear, nonconvex inverse problem although it is convex in a well defined region of the solution space [22], [26]. For projection-based algorithms, a previous work [22] proposed a regularized material-specific Gauss-Newton (RGN) method, which outperformed widely used unregularized maximum likelihood method. In [26], 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 [27]. The effect of assuming a perturbed detector response can have detrimental effects on image quality, as shown in [28]. 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 [22], but the choice is far from being optimal.

Recently, deep convolutional neural networks (CNN) have shown outstanding results in several image processing tasks [29], [30]. CNNs have been also proposed for solving several inverse problems, including image restoration and image reconstruction [31]–[36]. Among the diverse CNN configurations, U-Net has shown outstanding results for image segmentation [37] and has been applied to postprocessing in CT [31] and to artifact removal of nonlinear CT images [38]. Several works have also proposed deep learning for solving the material decomposition, mainly in the image domain. In [39], the authors trained a U-Net network using simulated phantoms constructed from random combinations of materials and tested the model on mice data. In [9], the authors proposed a VGG-16 network and tested it on Shepp-Logan synthetic data and experimental cylindrical phantoms. 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 [40], [41], [42]. In [40], 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 [43], [29]. 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 [44]. 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 [45], [46]; 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 [47], [48]. Recent works proposed related TL approaches based on fine tuning for material decomposition [28], [42]. 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 [49], [50]. 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 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 [51]. 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. In order to compare projection- and image-based decomposition strategies, the U-Net is trained to decompose the materials either in the projection domain (U-Net-P) or in the image domain (U-Net-I). These are compared to a regularized Gauss-Newton (RGN) method in the projection domain [22] and all methods are assessed on both numerical and experimental data. 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 [52]. 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.

II. 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 Θ projections. Let $\mathbf{s} = (s^1_{1,1}, \dots, s^\theta_{i,p}, \dots, s^\Theta_{I,P})^\top$ be the measurement vector, where $s^\theta_{i,p}$ represents the photon counts measured in the i-th energy bin at the p-th pixel under the θ-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. The forward model in spectral CT can be represented as the mapping

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

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.,

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

where $\boldsymbol{\rho}_m = (\rho_{m,1}, \dots, \rho_{m,v}, \dots, \rho_{m,V})^\top$ and $\mathbf{a}_m = (a^1_{m,1}, \dots, a^\theta_{m,p}, \dots, a^\Theta_{m,P})^\top$ 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.,

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

where $\mathbf{a}^\theta = (a^\theta_{1,1}, \dots, a^\theta_{m,p}, \dots, a^\theta_{M,P})^\top$ and $\mathbf{s}^\theta = (s^\theta_{1,1}, \dots, s^\theta_{i,p}, \dots, s^\theta_{I,P})^\top$. In particular, we consider the following non linear mixing [23] (excluding pileup and charge sharing effects)

$$s^\theta_{i,p} = \int_{E \in \mathcal{E}} n_{0,p}(E) d_{i,p}(E) \, \exp\left[- \sum_{m=1}^{M} a^\theta_{m,p} \tau_m(E) \right] \mathrm{d}E \tag{4}$$

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. This assumes that the local linear attenuation coefficient (LAC) of the object at energy E and the voxel $\mathbf{x}$, $\mu(E, \mathbf{x}_v)$, can be decomposed as the sum of M basis functions that are separable in energy and space:

$$\mu(E, \mathbf{x}) = \sum_{m=1}^{M} \rho_m(\mathbf{x}) \tau_m(E), \tag{5}$$

where $\rho_m(\mathbf{x}_v) = \rho_{m,v}$ and $\mu_m(\mathbf{x}_v) = \mu_{m,v}$ for a voxel $\mathbf{x}_v$.

Data simulation involves solving the forward problem. For this, we first project the mass densities $\boldsymbol{\rho}_m$ following (2) to get the projected mass densities $\mathbf{a}_m$ and then we apply the spectral mixing operator (4) to get the photon-counting data at the different energy bins $\mathbf{s}$.

III. SPECTRAL CT IMAGE RECONSTRUCTION

Solving the spectral CT image reconstruction problem in one step involves inverting the $\mathbb{R}^{MV} \to \mathbb{R}^{IP\Theta}$ non-linear mapping (1). In this work, we instead adopt two different two-step approaches: projection- and image-based decomposition approaches.

A. PROJECTION-BASED APPROACH

The projection-based approach splits the problem into Θ inversions of $\mathbb{R}^{MP} \to \mathbb{R}^{IP}$ mappings (3), which decompose projections for different energy bins into projections for different materials, followed by M inversions of $\mathbb{R}^{V} \to \mathbb{R}^{P\Theta}$ (2), which provide tomographic reconstructed images for the different materials. This

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). 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 [22]. 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.

For the first step (material decomposition in the projection domain), we consider two different methods: a regularized variational framework and a deep learning strategy, as described below.

1) Projection-based material decomposition with a variational method

Using a variational framework, material decomposition has been previously formulated as the minimization of the following cost function [22]:

$$\Phi(\mathbf{a}^\theta) = \frac{1}{2}\|\mathbf{s}^\theta - \mathcal{F}(\mathbf{a}^\theta)\|^2_{W^\theta} + \alpha \sum_m \mathcal{R}_m(\mathbf{a}^\theta), 1 \leq \theta \leq \Theta \tag{6}$$

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, α 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 [22]. While two different values of the regularization parameter could be used for different materials, it was previously found to not lead to significant improvement [22]. Cost function (6) 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 [53].

2) Projection-based material decomposition with U-Net (U-Net-P)

Deep learning has been recently proposed for inverse problems [31], [54]. In this work, we want to learn the mapping

$$\mathcal{H}_\beta : \tilde{\mathbf{s}}^\theta \mapsto \tilde{\mathbf{a}}^\theta, \tag{7}$$

where β indicates the parameters of the neural network, $\tilde{\mathbf{s}}^\theta$ corresponds to the normalized data defined as

$$\tilde{\mathbf{s}}^\theta = \ln(\hat{\mathbf{s}}/\mathbf{s}^\theta) \tag{8}$$

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., $\tilde{a}^{\theta,n}_{m,p} = a^{\theta,n}_{m,p} / \max_{p,\theta,n} a^{\theta,n}_{m,p}$.

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 β of the network that minimizes the following loss function

$$\begin{aligned} \Xi(\beta) &= \sum_{n=1}^{N} \|\mathcal{H}_\beta(\tilde{\mathbf{s}}^n) - \tilde{\mathbf{a}}^n\|^2 \\ &= \sum_{n=1}^{N}\sum_{m=1}^{M} \|\mathcal{H}_{\beta,m}(\tilde{\mathbf{s}}^n) - \tilde{\mathbf{a}}^n_m\|^2, \end{aligned} \tag{9}$$

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 (9) was done with Adam method under TensorFlow, with learning rate 10^{-4} 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 [37], which consists in a contracting multi-scale decomposition path and a symmetric expanding path with skip connections at each scale. The contracting path comprises 3×3 convolutions (with the zero-padding) each followed by a rectified linear unit (ReLU) and alternated with a 2×2 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 1×1 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).

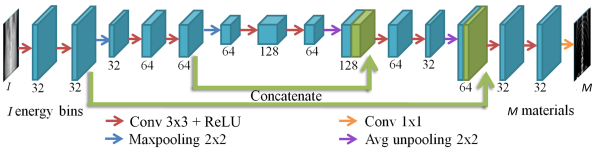

FIGURE 1. U-Net architecture for material decomposition in the projection domain (U-Net-P) and in the image domain (U-Net-I). For U-Net-P, the input is a stack of photon counting projection images for I energy bins ($P_x \times P_y \times I$), and output is a stack of the decomposed material projection images for M materials ($P_x \times P_y \times M$) for a projection angle θ. For U-Net-I, the input is a stack of reconstructed images for I energy bins ($Q_x \times Q_y \times I$), and output is a stack of the decomposed material reconstructed images for M materials ($Q_x \times Q_y \times M$) for a 2D slice.

3) Tomographic reconstruction of decomposed material mass densities

The second step corresponds to tomographic reconstruction for the different materials (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) [55], which minimizes the following cost function:

$$\Psi(\boldsymbol{\rho}_m) = \|\mathcal{X}\boldsymbol{\rho}_m - \mathbf{a}_m\|_2^2 + \gamma\|\nabla\boldsymbol{\rho}_m\|_2^2, \qquad (10)$$

where γ 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.

B. IMAGE-BASED APPROACH

The image-based approach is a widely used two-step approach for material decomposition in spectral CT [18], [19]. In the first step, images for the different energy bins, $\boldsymbol{\mu}_i^{\text{bin}}$, are reconstructed by solving

$$\tilde{\mathbf{s}}_i = \mathcal{X}(\boldsymbol{\mu}_i^{\text{bin}}), \quad 1 \le i \le I, \qquad (11)$$

where $\boldsymbol{\mu}_i^{\text{bin}} = (\mu_{i,1}^{\text{bin}}, \ldots, \mu_{i,v}^{\text{bin}}, \ldots, \mu_{i,V}^{\text{bin}})^\top$, $\mu_{i,v}^{\text{bin}}$ is the reconstructed image for the i-th energy bin at the v-th voxel, and $\tilde{\mathbf{s}}_i$ corresponds to the normalized data (8). The first step (11) corresponds to I inversions of $\mathbb{R}^V \to \mathbb{R}^{P\Theta}$ (11).

The second step is the material decomposition in the image domain

$$\boldsymbol{\mu}^{\text{bin}} = \mathcal{M}(\boldsymbol{\rho}), \qquad (12)$$

where $\boldsymbol{\mu}^{\text{bin}} = (\boldsymbol{\mu}_1^{\text{bin},\top}, \ldots, \boldsymbol{\mu}_i^{\text{bin},\top}, \ldots, \boldsymbol{\mu}_I^{\text{bin},\top})^\top$, $\boldsymbol{\rho} = (\boldsymbol{\rho}_1^\top, \ldots, \boldsymbol{\rho}_m^\top, \ldots, \boldsymbol{\rho}_M^\top)^\top$ and $\mathcal{M}$ is the spectral mixing operator in the image domain. We remark that $\mathcal{M}$ cannot be analytically described in a simple manner without taking an approximation and that we are only interested in learning its inverse mapping. Step (12) requires one inversion of the mapping $\mathbb{R}^{VM} \to \mathbb{R}^{VI}$.

The implementation of the two-step approach used in this work is described as follows.

1) Tomographic reconstruction of energy bins

Tomographic reconstruction for the different energy bins (11) is solved using the same methods and same parameters as described in section III-A3. For experimental data, the regularized conjugate gradient method minimizes

$$\Psi(\boldsymbol{\mu}_i^{\text{bin}}) = \|\mathcal{X}\boldsymbol{\mu}_i^{\text{bin}} - \tilde{\mathbf{s}}_i\|_2^2 + \gamma\|\nabla\boldsymbol{\mu}_i^{\text{bin}}\|_2^2. \qquad (13)$$

2) Image-based material decomposition with U-Net (U-Net-I)

We solve the material decomposition problem in the image domain (12) using deep learning. For this, we aim to learn the mapping

$$\mathcal{M}_\omega^{-1} : \boldsymbol{\mu}^{\text{bin},j} \mapsto \boldsymbol{\rho}^j, \qquad (14)$$

where $\mathcal{M}^{-1}$ is the inversion of the mixing operator in the image domain and $()^j$ stands for the j-th slice in the volume. In this case, input and outputs are not normalized because mass densities are already naturally normalized to water.

Learning means finding the parameters ω of the network that minimizes the following loss function

$$\begin{aligned}\Pi(\omega) &= \sum_{n=1}^{N} \|\mathcal{M}_\omega^{-1}(\boldsymbol{\mu}^{\text{bin},n}) - \boldsymbol{\rho}^n\|^2 \\ &= \sum_{n=1}^{N}\sum_{m=1}^{M} \|\mathcal{M}_{\omega,m}^{-1}(\boldsymbol{\mu}^{\text{bin},n}) - \boldsymbol{\rho}_m^n\|^2 \qquad (15)\end{aligned}$$

where $(\boldsymbol{\mu}^{\text{bin},n}, \boldsymbol{\rho}^n)$ are N input-output vector pairs (12) that can be reshaped into input-output array pairs of size ($Q_x \times Q_y \times I$,$Q_x \times Q_y \times M$), N is the number of image slices in the training set, and each image is of size $Q_x \times Q_y$. Minimization of (15) was done with the same method and the same U-Net architecture as described in the projection domain (figure 1).

C. DATA AND IMAGE EVALUATION

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/) [51] (figure 2). We used 50 phantoms for training, 3 for evaluation, and 9 for test. CT volumes have voxel size 0.96 mm × 0.96 mm × 1.8 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 g·cm^{-3}. Then, mass densities for both materials were projected using the Radon transform (assuming parallel geometry) to create the projected mass densities.

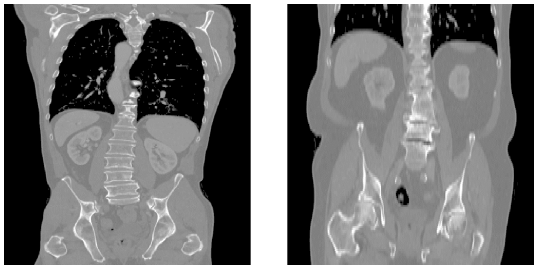

FIGURE 2. Two examples from the CT Kits19 data base.

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 [22], [53]. 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 924×8 detector pixels.

In this work, we follow a Sim2Real transfer approach, so training data for the deep learning strategies are based on the numerical human phantoms. Training data for U-Net-P then consisted of 19440 projections (360 projections for 54 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.

Training data for U-Net-I consisted of 5400 images (100 slices for 54 phantoms). Data augmentation was performed by random crop along all directions. In addition, data augmentation was performed to account for the appearance of bone tissue in the soft tissue image. For this, the bone image was first normalized and then scaled to have the maximum between 1.2 and 2.2 using random sampling from a uniform distribution. Then, this reference bone image was scaled again by κ_b to provide the final bone image. The reference bone image was also scaled by κ_s and added to the soft tissue image to provide the final soft tissue image. The values κ_b and κ_s were obtained from the ratios of the Gammex bone inserts (κ_b, κ_s) as $(0.77, 0.37)$, $(0.01, 1.82)$, $(0.80, 0.53)$, $(0.45, 1.15)$. We remark that these ratios account only for the maximum values in the image but larger variations of bone values and ratios are actually being considered because the bone image is not homogeneous. Given the computational cost of solving the two-step approach in the image domain, data augmentation and noise were done before training.

2) Experimental data

A SPCCT prototype system (Philips Medical Systems) installed at CERMEP, Lyon, was used for this study [56]. 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 μm $\times$ 270 μm at isocenter, ChromAIX2 application-specific integrated circuits combined with cadmiun 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 [57], which is made of a cylinder of size 330 mm $\times$ 270 mm $\times$ 50 mm and a density of 1.029 g·cm^{-3} 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 $462 \times 16 \times 2400 \times 5$. Final reconstructed voxel size was 1 mm $\times$ 1 mm $\times$ 2 mm. In order to compare the data SNR with previous work [28], 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 $924 \times 8 \times 24912 \times 5$. This corresponded to a total number of photons of $N_0 = 1.8 \cdot 10^5$.

Image reconstruction of experimental data with both RGN and deep learning strategies take into account the source spectrum and the detector response function of

the scanner, following the spectral mixing definition (4). We have not included charge sharing and pulse pileup effects, which can lead to artefacts. We also remark that data calibrations and corrections included in the scanner manufacturer pipeline have not been included in this work.

3) Image evaluation

U-Net-P, U-Net-I 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 [26]. 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, bias on reconstructed images for soft tissue and bone, and noise for soft tissue. Normalized MSE was computed for each material as follows:

$$\text{MSE}_m = \frac{\|\mathbf{a}_m - \mathbf{a}_m^{\text{Target}}\|_2}{\|\mathbf{a}_m^{\text{Target}}\|_2}. \tag{16}$$

Relative bias was computed for each material as follows:

$$\text{bias}_m = 100\frac{|\bar{\boldsymbol{\rho}}_m - \bar{\boldsymbol{\rho}}_m^{\text{Target}}|}{\bar{\boldsymbol{\rho}}_m^{\text{Target}}}, \tag{17}$$

where the mean was taken across a circular region in a homogeneous area (of radius equal to ten pixels) for soft tissue and in a mask for bone (comprised of pixels that belong to the support of bone in the phantom). Noise for soft tissue was computed as the standard deviation in the circular homogeneous region. 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 [58]. Monochromatic images were computed as the LACs at energy E and point $\mathbf{x}$ in the reconstructed volume following (5).

IV. RESULTS

A. NUMERICAL THORAX PHANTOM DATA

Figure 3 shows decomposed images by RGN for two different values of α and by U-Net-P. 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-P is able to reduce noise while maintaining image details, as can be appreciated in figure 5.

Tomographic reconstructions of decomposed data with RGN, U-Net-P and U-Net-I are shown in figure 6. 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-P 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 7. U-Net-I provides much higher image quality than projection-based methods, with less noise, better recovery of image details and excellent preservation of image texture.

Figure 8 shows image details of the reconstructed images for soft tissue. RGN for $\alpha = 0.6$ misses image details while U-Net-P presents better recovery than RGN for $\alpha = 0.1$. U-Net-I provides best results with almost perfect decomposition. The improved recovery of image details by U-Net-I can be explained by the presence of these details in the tomographic reconstructions of the different energy bins (figure 9), which constitute the input for U-Net-I.

Figure 10 (a, b) displays normalized MSE and SSIM for decomposed projections for RGN and U-Net-P. For soft tissue, U-Net-P 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-P led to similar decrease in MSE for bone as for soft tissue. U-Net-P also led to increase in SSIM with respect to RGN for bone and soft tissue. For bone tissue, U-Net-P 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.

Figure 10(c, d) shows normalized MSE and SSIM for tomographic reconstructions for all methods. U-Net-P leads to better results than RGN, and U-Net-I leads to the best results overall, with up to 85 % and three-fold increase in SSIM with respect to RGN and U-Net-P, respectively.

Figure 10(e, f) show normalized bias for tomographic reconstructions for all methods. Among projection-based methods, RGN leads to lower bias

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3056150, IEEE Access

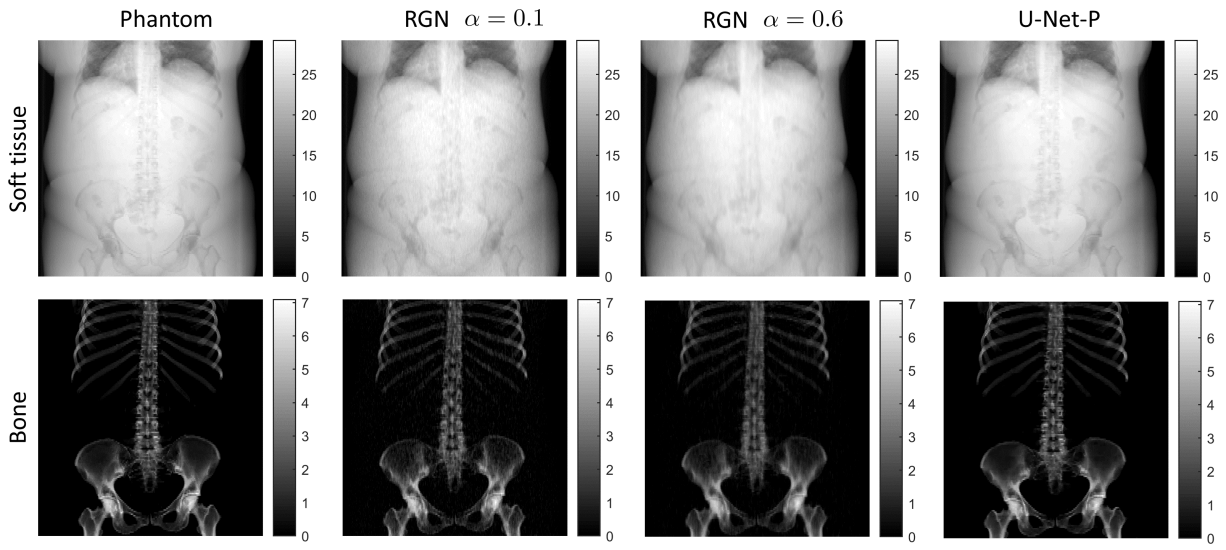

FIGURE 3. Phantom and decomposed projections (for 0° view) for RGN and U-Net-P for a synthetic phantom from the test set.

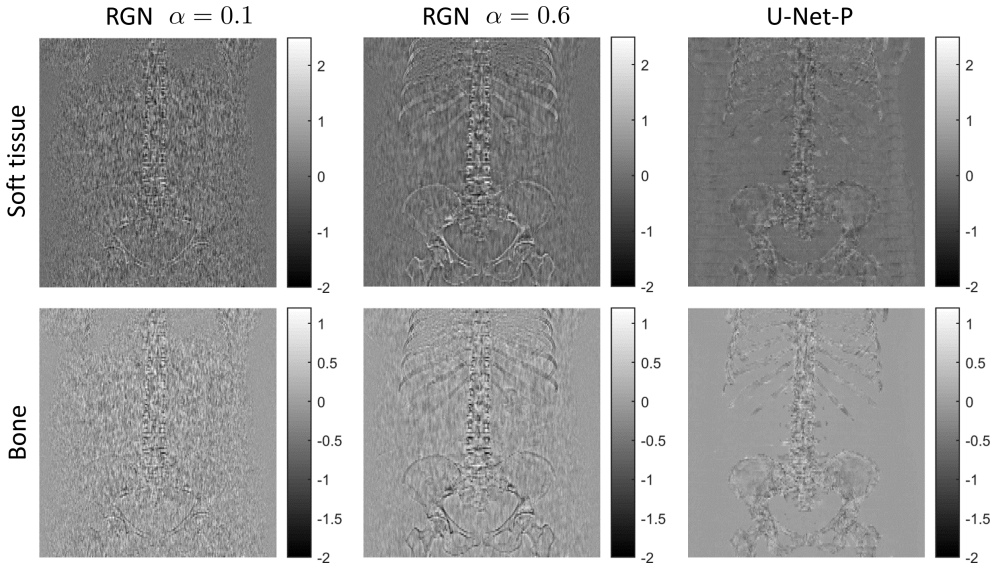

FIGURE 4. Error images of decomposed projections by RGN and U-Net-P for the same view as in figure 3.

for soft tissue but much higher bias for bone than U-Net-P. U-Net-I presents the lowest bias overall. For all methods, bias for soft tissue is smaller than for bone because it is computed over a larger region.

Figure 10(g) shows noise (SD) in soft tissue. Adopting RGN for $\alpha = 0.6$ as reference, U-Net-P and U-Net-I lead to 27 % and 85 % decrease in noise, respectively.

B. EXPERIMENTAL DATA

1) Experimental phantom data

Figure 11 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. U-Net-I leads to a clearer recovery and closer estimation for most inserts. In terms of image quality, U-Net-P presents less noisy images and slightly less ring artefacts than RGN. U-Net-I leads to high image quality with suppression of most ring artefacts. The 60 keV images provided by the two methods are very similar, with U-Net-P and U-Net-I presenting less noise.

Figure 12 shows quantification errors of the phantom inserts. Figures 12 (left and center) shows errors for all inserts for soft tissue and bone images. For soft tissue-like inserts, RGN leads to larger errors than deep learning methods, as it fails to provide zero densities for the bone image. For bone-like inserts, all methods

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3056150, IEEE Access

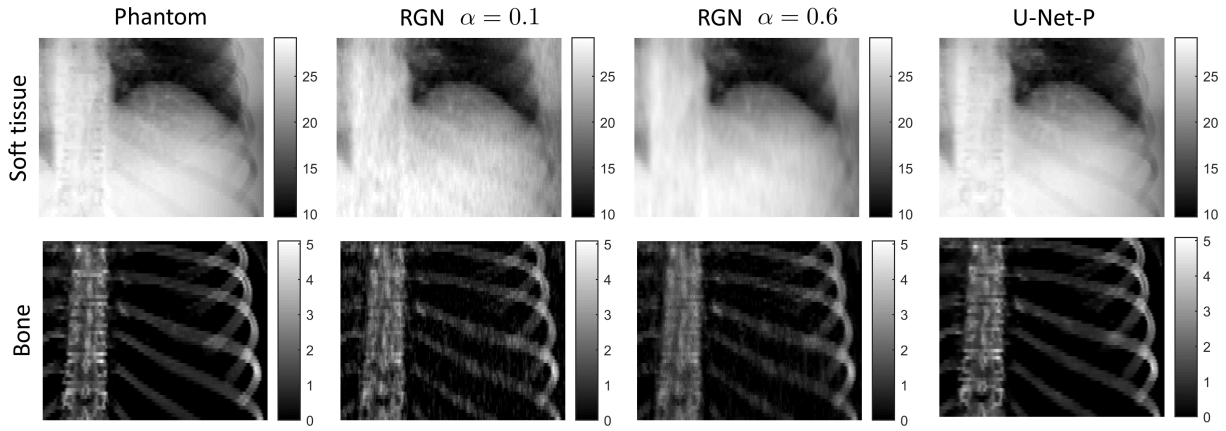

FIGURE 5. Image detail for phantom and decomposed projections for the images shown in figure 3.

tend to present a negative bias in the bone image (part of bone placed in the soft tissue image). Overall, U-Net-I leads to lowest quantification errors, as shown in figure 12 (right). Among the projection methods, U-Net-P provides lower median but larger variance (U-Net-P has almost zero error for soft tissue-like rods).

2) Human thorax data

Figure 13 shows tomographic reconstructions of the decomposed materials and the 60 keV image for RGN, U-Net-P and U-Net-I. Among the projection methods, RGN provides noisier decomposed images and appears more sensitive to ring artefacts than U-Net-P. U-Net-I provides higher image quality than projection-based methods, with better recovery of details for both soft tissue and bone images. Compared to the decomposed images, 60 keV images show highly reduced ring artefacts, with RGN presenting few artefacts, U-Net-P correcting for these artefacts, and U-Net-I providing the best image quality overall, with reduced noise and better recovery of texture for the soft tissue image.

C. COMPUTATION TIME

We run our code on a Windows computer with 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. U-Net-P requires 3.2 days for 1200 epochs and U-Net-I requires 33 hours for 600 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-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. U-Net-I presents higher tomographic reconstruction times than projection based methods because it requires reconstruction of five volumes (five energy bins) instead of two (two materials).

V. DISCUSSION

We have proposed a deep learning method based on a Sim2Real approach and a U-Net architecture for solving the material decomposition problem in spectral CT. 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 compared projection- and image-based material decomposition by training the U-Net on either the projection domain (U-Net-P) or in the image domain (U-Net-I), and both approaches were compared to a RGN method in the projection domain. Among projection-based methods, we found that U-Net-P 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, for synthetic data, U-Net was able to remove noise while maintaining image quality. Differences in terms of image quality between U-Net-P and RGN were smaller for experimental data. In terms of quantification of decomposed materials, both methods led to similar results on the assessed experimental phantom data. However, U-Net-P was less affected by ring artefacts than RGN. Furthermore, U-Net led to a significant reduction in prediction time, specially for large data sets. U-Net-I presented a significantly superior improvement in image quality than projection based methods. For synthetic data, U-Net-I led to almost perfect material decomposition, and for experimental data, it removed noise and ring artefacts while recovering image details.

The superiority in terms of image quality of U-Net-P with respect to RGN must be discussed in detail. This can be explained by the fact that U-Net can learn

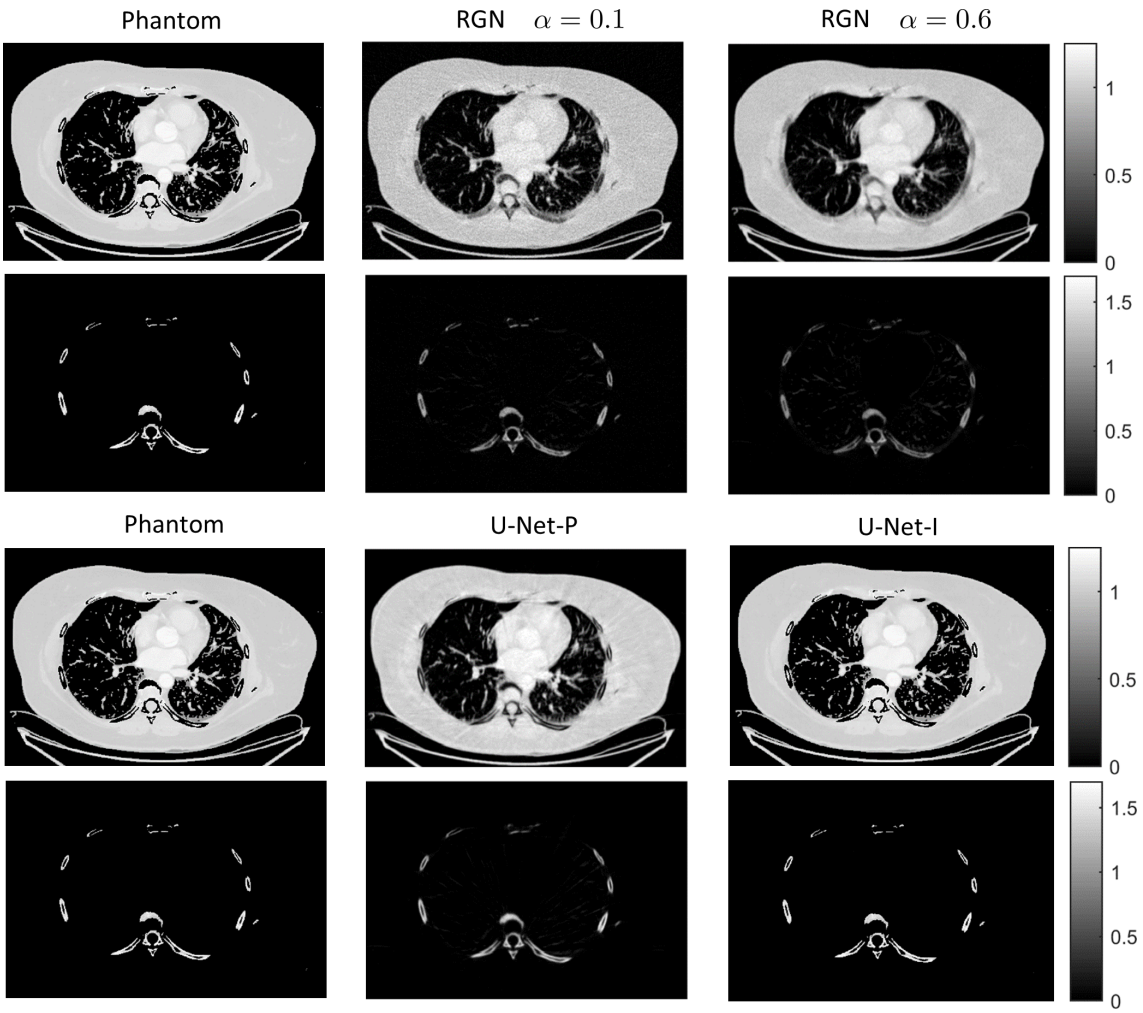

FIGURE 6. Tomographic reconstruction of a numerical human phantom for RGN, U-Net-P and U-Net-I. Image reconstruction corresponding to decomposed projections shown in figure 3.

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-P (CPU)	70 s	""	25 min	""
U-Net-P (GPU)	6 s (BS = 16)	""	45 s (BS = 1024)	""
U-Net-I (GPU)	6.6 s (BS = 16)	63 s	45 s	125 min

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 [52]. 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 [22]. 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 of the regularization parameter. For further insight into the selection of the regularization parameter for RGN, we refer to [22] and [26]. Although fine tuning can generally lead to an acceptable compromise between

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3056150, IEEE Access

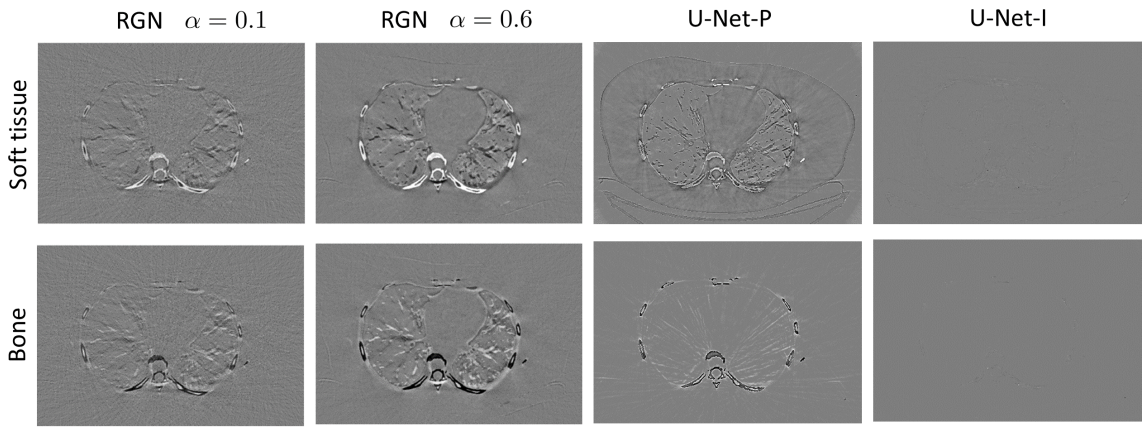

FIGURE 7. Error images of tomographic reconstructions of a numerical human phantom (corresponds to figure 6).

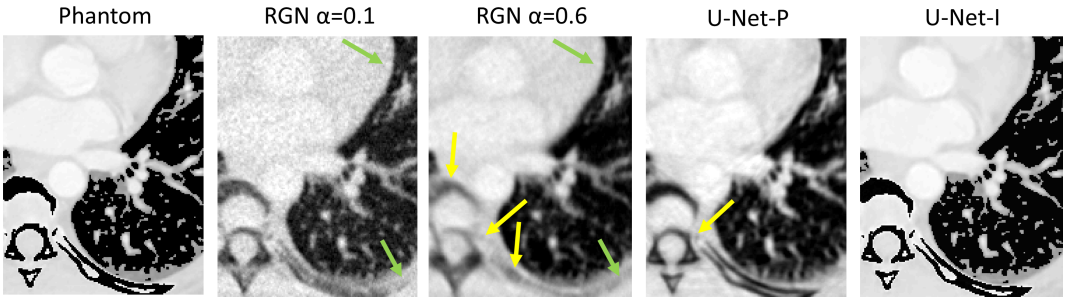

FIGURE 8. Detail of tomographic reconstruction of soft tissue for the images shown in figure 6. Yellow arrow points where bone details are not well recovered. Green arrow points to the appearance of 'fake' details.

FIGURE 9. Tomographic reconstructed images for energy bins 1, 3 and 5. Images are displayed with the same colorbar $(0, 0.53)$. Random crops of the five energy bins images are the input for U-Net-I. The output for these images are provided in figure 6.

noise and regularization for variational approaches, its benefits are mitigated by the fact that regularization functionals are not optimal.

The comparison between projection- and image-based methods is still an open question, particularly for deep learning approaches. In this work, we found that U-Net-I led to superior results than U-Net-P. Specifically, we found that U-Net-P led to high quality decomposition in the projection domain, as shown in the image detail in figure 5. However, this did not translate into an equivalent image quality after tomographic reconstruction. Tomographic reconstructed images suffered from some noisy structure or streak-like artefacts in the soft tissue image, a loss of small image details on bone parts and slightly blurred small details. This can be due to several reasons. First, learning the probability distribution of the soft tissue and bone seems easier in the image domain than in the projection domain. In the image domain, distributions could be simplified to some given distributions centered around their mean values. In the projection domain, values depend on both material density and length. For a given value of a pixel in the projection domain, it is then not possible to distinguish between different densities or thickness. Using this reasoning, material decomposition taking into account other projections may improve results. Second, for imaged-based methods, tomographic reconstruction of the different energy bins provide images that are noisy but for which image details are preserved (figure 9). Thus, U-Net-I learns

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/ACCESS.2021.3056150, IEEE Access

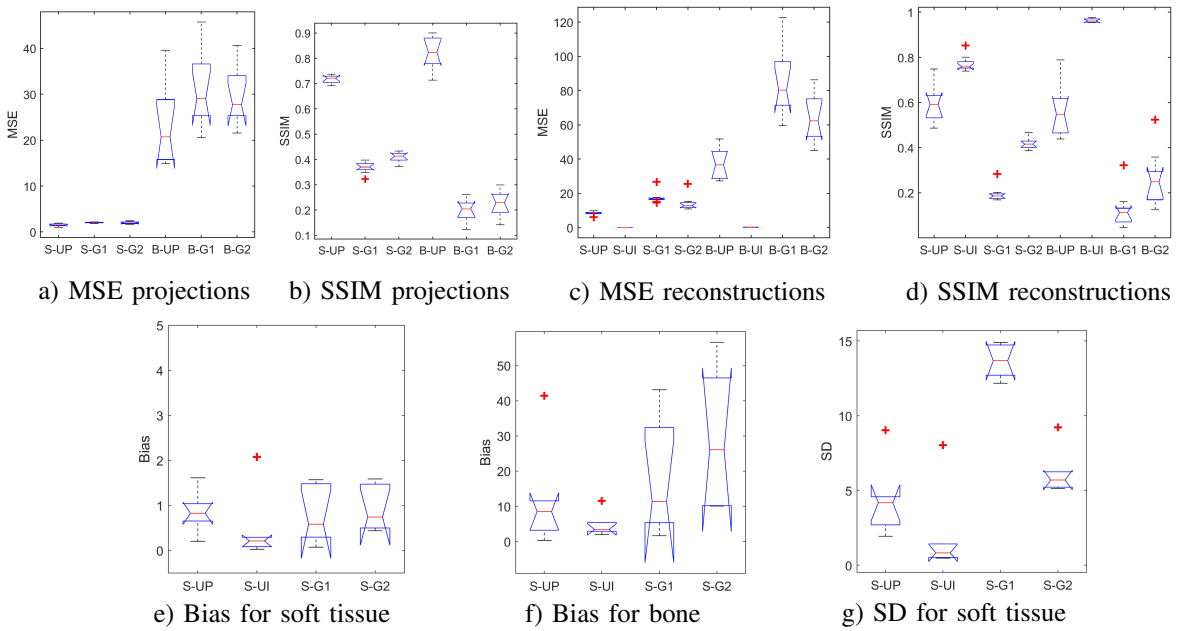

FIGURE 10. Quantitative results on the numerical thorax test set. Boxplots for normalized MSE and SSIM of decomposed projections and tomographic reconstructed images, given by U-Net-P (UP), U-Net-I (UI), RGN for $\alpha = 0.1$ (G1) and RGN for $\alpha = 0.6$ (G2), for soft tissue (S) and bone (B). UI is provided only on reconstructions. 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.

to simultaneously decompose materials and perform image restoration. Third, U-Net-I learns implicitly the prior distributions of the decomposed materials on the image domain while U-Net-P learns the prior distributions on the intermediate projection domain, on which small errors may translate in artefacts or noise in the image domain. Projection-based decomposition methods may thus require a further postprocessing step. In addition, we also found that training was easier in the image domain, requiring a lower number of epochs to start providing good results (results not shown).

For experimental data, projection-based algorithms were subject to ring and band artefacts. Interestingly, U-Net-P was more robust to the presence of artefacts than RGN. This may be because for U-Net-P we used normalized data while for RGN we used absolute data. Another explanation could be that U-Net-P 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 [28]. More interesting, U-Net-I did not present ring artefacts on the experimental data.

Several deep learning approaches have been previously proposed for material decomposition. Most of them perform material decomposition in the image domain [9], [39]. Recently a multilayer perceptron was proposed for solving the material decomposition problem in the projection domain [42]. 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. 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 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 [45], [46], [59]. 'Transferability' has been shown to depend on the distance between source and target specific tasks [60]. 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. For projection-based methods, we also remark that due to the fact that we are projecting mass densities, then the material decomposition problem in the projection

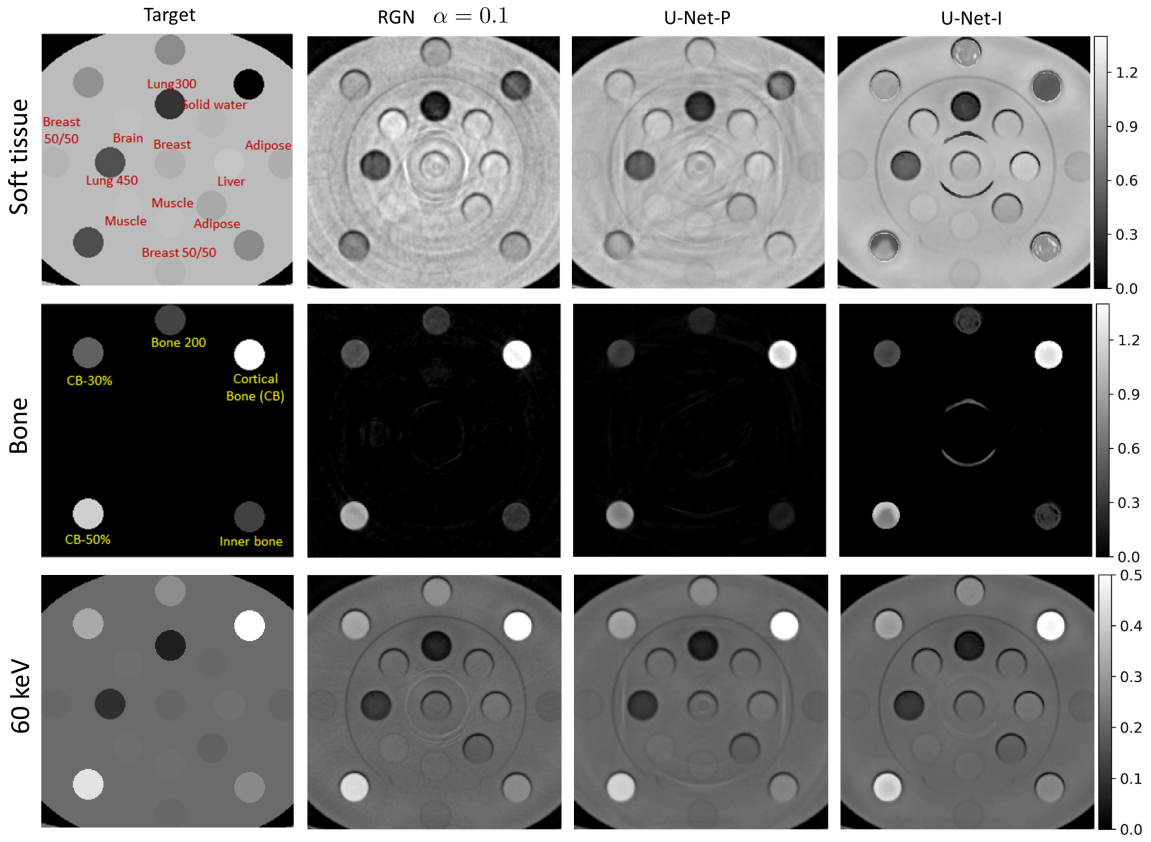

FIGURE 11. Experimental phantom with (Gammex and Cirs) inserts. Left: Estimated target images. Second to fourth columns: Tomographic reconstruction of soft tissue and bone and 60 keV image for RGN, U-Net-P and U-Net-I. We remark that data calibrations and corrections included in the scanner manufacturer pipeline have not been included in this work.

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.

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 [31], [38]. 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. In order to assess this, we compared U-Net-P with two ConvNets without max-pooling operations, one with four convolutional layers, which corresponds to the main path or the first level of the U-Net, and a second one with the same number of convolutional layers as U-Net. We found that U-Net-P allows better removal of artefacts and provides improved image texture for soft tissue (results not shown). Besides the architecture, deep iterative methods, also called model-based learning methods, proposed for solving inverse problems could provide further benefits [33], [36], [61]. In particular, these iterative methods include a data consistency condition, which could mitigate the appearance of artefacts that can potentially lead to mislead diagnosis. Although 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 [62]. 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 [52]. With regard to the amount of training data, we found that using less than ten phantoms for training U-Net-

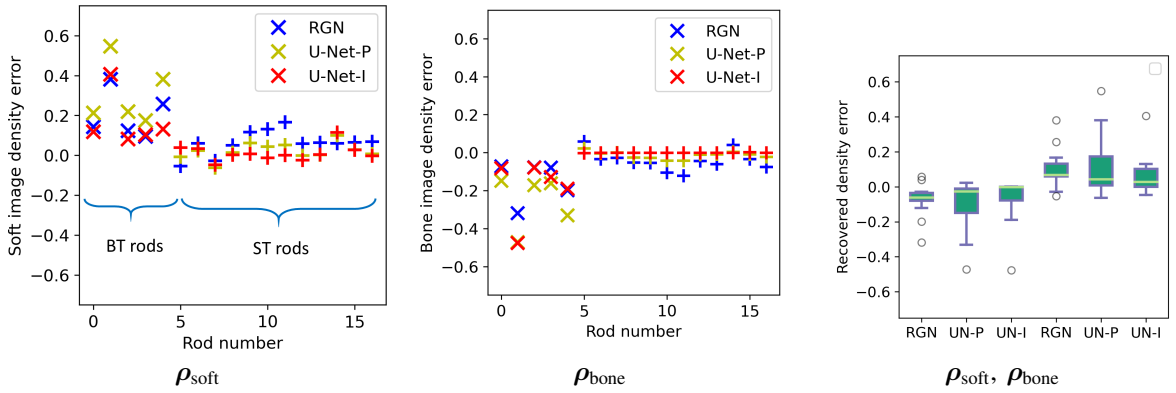

FIGURE 12. Quantification error of (Gammex and Cirs) inserts for RGN (blue) and U-Net-P (yellow) and U-Net-I (red) from decomposed soft tissue image ρ_{soft} (left) and decomposed bone image ρ_{bone} (center). Inserts are marked with the symbol $+$ for soft tissue like materials (lung, breast, muscle, solid water, adipose tissue, brain, muscle) and with the symbol $\times$ for bone like materials (cortical bone (CB), inner bone, CB 30%, CB 50%). Right: Boxplot of density errors where median and standard deviation are taken across all inserts for soft tissue and bone like inserts.

P led to lower image quality (results not shown); however, fewer data were not compensated for by more extensive data augmentation strategies such as translation and rotation. Further investigation could set the minimum amount of data for training 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 [28]. However, a previous work found only small improvements using a fully-connected network [42]. 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 [28].

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.

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 Klaus Erhard, Bernhard Brendel and Yoad Yagil for their support.

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 Molloi. 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 Molloi. Breast tissue decomposition with spectral distortion correction: A postmortem study. Medical Physics, 41(10):101901, sep 2014.

[7] Qiaofeng Xu, Alex Sawatzky, Mark A Anastasio, and Carsten O Schirra. Sparsity-regularized image reconstruction of decomposed K-edge data in spectral CT. Physics in Medicine and Biology, 59(10):N65, 2014.

[8] David S Rigie and Patrick J La Rivière. Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization. Physics in Medicine and Biology, 60(5):1741–1762, mar 2015.

[9] Zhengyang Chen and Liang Li. Robust multimaterial decomposition of spectral CT using convolutional neural networks. Optical Engineering, 58(01):1, jan 2019.

[10] Bo Zhao, Hao Gao, Huanjun Ding, and Sabee Molloi. Tight-frame based iterative image reconstruction for spectral breast CT. Medical Physics, 40(3):031905, feb 2013.

[11] Weiwen Wu, Peijun Chen, Shaoyu Wang, Varut Vardhanabhuti, Fenglin Liu, and Hengyong Yu. Image-domain Material Decomposition for Spectral CT using a Generalized Dictionary Learning. IEEE Transactions on Radiation and Plasma Medical Sciences, pages 1–1, may 2020.

[12] Yan Xi, Yi Chen, Rongbiao Tang, Jianqi Sun, and Jun Zhao. United Iterative Reconstruction for Spectral Computed Tomography. IEEE Transactions on Medical Imaging, 34(3):769–778, mar 2015.

[13] Hao Gao, Hengyong Yu, Stanley Osher, and Ge Wang. Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM). Inverse Problems, 27(11):115012, 2011.

[14] Kyungsang Kim, Jong Chul Ye, William Worstell, Jinsong Ouyang, Yothin Rakvongthai, Georges El Fakhri, and Quanzheng Li. Sparse-view spectral CT reconstruction using spectral patch-based low-

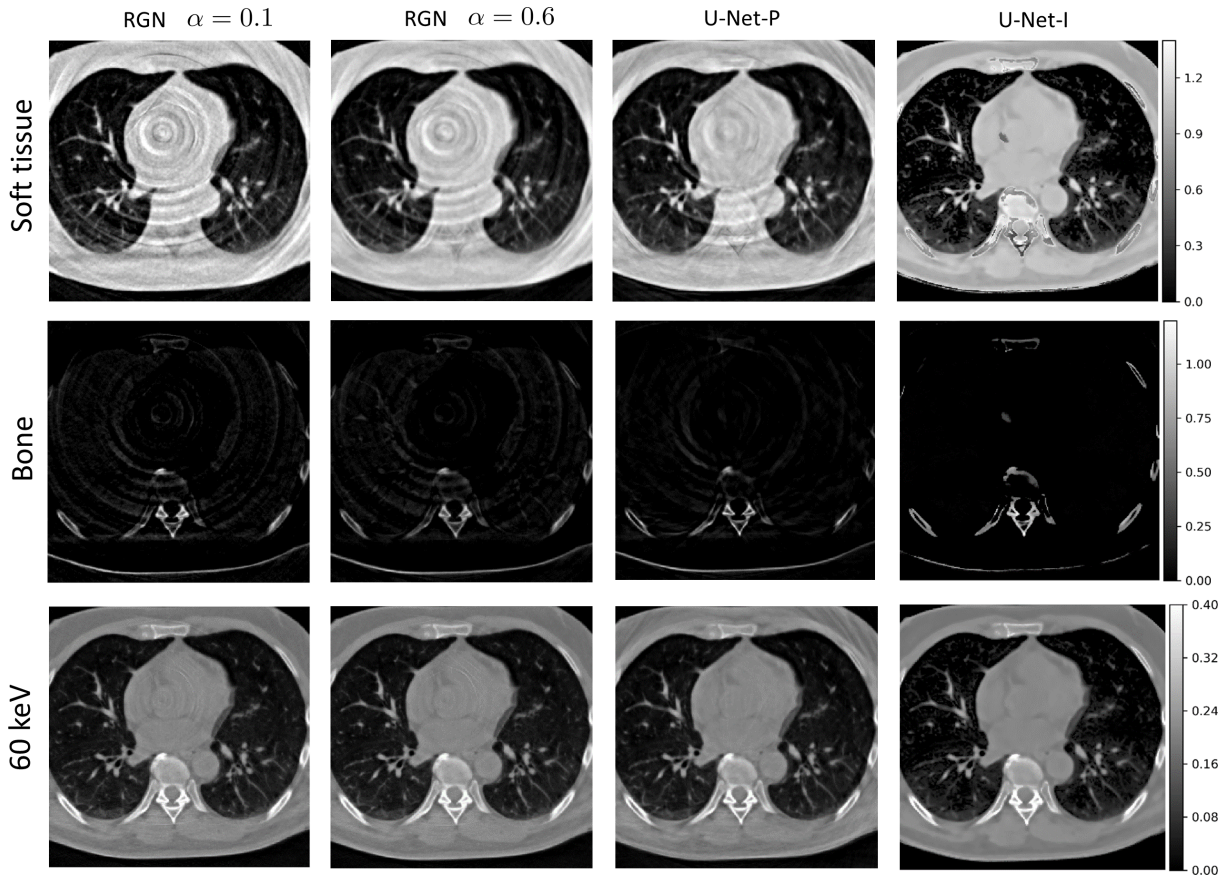

FIGURE 13. Tomographic reconstruction of a human thorax for RGN, U-Net-P and U-Net-I.

rank penalty. IEEE transactions on medical imaging, 34(3):748–760, 2015.

[15] Oguz Semerci, Ning Hao, Misha E Kilmer, and Eric L Miller. Tensor-based formulation and nuclear norm regularization for multienergy computed tomography. IEEE transactions on image processing : a publication of the IEEE Signal Processing Society, 23(4):1678–93, apr 2014.

[16] Huy Q Le and Sabee Molloi. Least squares parameter estimation methods for material decomposition with energy discriminating detectors. Med. Phys., 38(1):245, 2011.

[17] 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.

[18] 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.

[19] 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.

[20] Cyril Mory, Bruno Sixou, Salim Si-Mohamed, Loic Boussel, Simon Rit, ıc 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.

[21] 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.

[22] 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.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] Katsuyuki Taguchi and Jan S Iwanczyk. Vision 20/20: Single photon counting x-ray detectors in medical imaging. Medical Physics, 40(10):100901, 2013.

[28] 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.

[29] 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.

[30] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. Nature, 521(7553):436–444, may 2015.

[31] 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.

[32] 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.

[33] Andreas Hauptmann, Felix Lucka, Marta Betcke, Nam Huynh, Jonas Adler, Ben Cox, Paul Beard, Sebastien 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.

[34] Saiprasad Ravishankar, Jong Chul Ye, and Jeffrey A Fessler. Image Reconstruction: From Sparsity to Data-adaptive Methods and Machine Learning. Technical report, 2019.

[35] 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.

[36] Simon Arridge, Peter Maass, Ozan Öktem, and Carola Bibiane Schönlieb. Solving inverse problems using data-driven models. Acta Numerica, 28:1–174, may 2019.

[37] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. pages 234–241. Springer, Cham, 2015.

[38] 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.

[39] 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.

[40] 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.

[41] Kevin C Zimmerman and Taly Gilat Schmidt. Experimental comparison of empirical material decomposition methods for spectral CT. Physics in Medicine and Biology, 60(8):3175, 2015.

[42] 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.

[43] 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.

[44] 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.

[45] 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.

[46] Garrett Wilson and Diane J Cook. A Survey of Unsupervised Deep Domain Adaptation. Technical report, 2020.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] Nicholas Heller, Niranjan Sathianathen, 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 Papanikolopoulos, and Christopher Weight. The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes. Technical report, 2019.

[52] Jonas Adler and Ozan Oktem. Deep Bayesian Inversion Computational uncertainty quantification for large scale inverse problems. Technical report, 2018.

[53] Nicolas Ducros. SPRAY - Spectral X-ray image reconstruction toolbox, 2017.

[54] 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.

[55] 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.

[56] 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.

[57] P. Pemler, U. Schneider, and J. Besserer. Evaluation of the electron density phantom CIRS Model 62. Z Med Phys, 11(1):25–32, 2001.

[58] 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.

[59] 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.

[60] 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.

[61] Jonas Adler and Ozan Oktem. Learned Primal-Dual Reconstruction. IEEE Transactions on Medical Imaging, 37(6):1322–1332, jun 2018.

[62] Weiwen Wu, Dianlin Hu, Hengyong Yu, Hongming Shan, Shaoyu Wang, Wenxiang Cong, Chuang Niu, Pingkun Yan, Vince Vardhanabhuti, and Ge Wang. Stabilizing Deep Tomographic Reconstruction Networks. aug 2020.

J. ABASCAL received the B.S. degree in physics from The Autonomous University of Madrid, Spain, in 2002, the M.Sc. degree in applied mathematics from The University of Manchester, U.K., in 2003, and the Ph.D. in Medical Physics from University College London, U.K., in 2007.

After his Ph.D., he has been at Central Supelec, France, in 2007, and at University Carlos III of Madrid, Spain, from 2008 till 2016. He is currently a postdoctoral research fellow at CREATIS, France. He has contributed in inverse problems and compressed sensing for different medical imaging modalities. He is currently working on solving inverse problems using deep learning. His interest lies in inverse problems, image reconstruction, optimization, image processing, and machine learning.

N. DUCROS received an MSc in Electrical Engineering and an MSc in Biomedical Engineering, both from Strasbourg University, France, in 2006. He obtained a PhD in Electrical Engineering from Lyon University, in 2009, for which he received the Best Thesis Award of the French Section of the IEEE EMBS.

From 2009 to 2012, he held a Postdoctoral Fellowship in the Department of Physics of the Politecnico Milan in Italy. Since 2014, he has been an Associate Professor in the Electrical Engineering Department of Lyon University and with the Biomedical Imaging Laboratory CREATIS.

He is a member of the IEEE Bio Imaging and Signal processing (BISP) technical committee. His research interests include medical imaging and applied inverse problems, with particular emphasis on single-pixel imaging and spectral computerized tomography.

V. PRONINA received a B.Sc. degree in Biomedical Engineering and an M.Sc. Degree in Biomedical Systems and Technologies from Bauman Moscow State Technical University, Russia, in 2015 and 2017 respectively. In 2018, she obtained an M.Sc. Degree in Biomedical Engineering and Design from École Nationale supérieure des Mines de Saint-Étienne, France. Currently, Mrs. Pronina is pursuing her Ph.D. in Computational and Data Science Engineering at Skolkovo Institute of Science and Technology, Russia.

In 2020, Valeriya was awarded Ostrogradski Scholarship from the French Embassy, Russia, which allowed her to conduct part of her research at CREATIS Biomedical Imaging Laboratory, France. Her research interests include deep learning, image reconstruction, inverse problems, and optimization methods. Currently, she is working on solving biomedical image restoration problems using deep learning.

S. RIT completed in 2007 his PhD in computer science from the Lumière University, Lyon, France, at the LIRIS laboratory and on respiratory motion correction for cone-beam computed tomography (CT). He was then a postdoctoral fellow at the Netherlands Cancer Institute (Amsterdam) on the same topic. He is currently a CNRS research associate at the CREATIS laboratory in Lyon, working on cone-beam CT reconstruction for image-guided radiotherapy (IGRT), spectral CT, proton CT and single photon emission computed tomography (SPECT).

PIERRE-ANTOINE RODESCH was born in 1990 in France. He has studied mechanical engineering at Ecole Nationale des Ponts et Chaussées combined with a Master of Science in biomechanics at Université VI Paris Descartes in 2014. He completed his Doctor of Philosophy (PhD) at Université Grenoble Alpes in the field of applied mathematics. In 2018 he joined the CREATIS laboratory in Lyon as a postdoctorant in the frame of the SPCCT H2020 project. His research of interest are photon-counting detectors, CT reconstruction, material decomposition and spectral CT for medical diagnosis.

T. BROUSSAUD received a M.Sc. in Medical physics & image processing in 2001 from UPS Toulouse 3. He is interested in the validation/evaluation of Medical Imaging techniques in clinical setting. He worked in CERMEP / CREATIS Labs within SPCCT European project to provide support for data acquisition, image reconstructions and statistical analysis. He has joined lately a Radiomics project aiming at building & evaluating the performances of a diagnostic/prediction model in oncology.

S. BUSSOD received an engineering degree in image processing at Télécom Saint-Étienne, and a MSc in advanced imaging at the Jean Monnet University of Saint-Étienne, France, in 2018.

Currently, she is pursuing her year Ph.D in deep learning algorithms development for spectral X-ray computed tomography at the CREATIS lab, INSA Lyon, France. Her research interests are image and signal processing and analysis, medical imaging and deep learning.

A. HAUPTMANN received his B.Sc. and M.Sc. in Mathematics from the Technical University Munich in 2011 and 2012, respectively. He received his PhD in 2017 from the University of Helsinki in Applied Mathematics.

He worked from 2013 to 2014 as R&D Scientist at the X-ray sensor manufacturer AJAT Oy in Espoo, Finland. Since 2017 he holds a position as Research Associate in Medical Image Computing at the Centre for Medical Imaging, University College London and is an Assistant Professor (tenure track) of Computational Mathematics since 2019 at the Research Unit of Mathematical Sciences at the University of Oulu. His research interest is in inverse problems and tomographic imaging with a focus on medical applications, his current focus is on combining inverse problems and data driven methods such as deep learning and neural networks.

S. ARRIDGE received a BA Hons in Physics from Cambridge University in 1982, after which he moved to UCL Medical Physics where he completed a PhD in 1990. He subsequently joined the department of Computer Science UCL as a lecturer and has been professor of image processing since 2001 and visiting professor in the department of Mathematics since 2011. He was a founding member of the UCL Centre for Medical Image Computing in 2005 and the founder and director of the UCL Centre for Inverse Problems since 2013.

He has been a member of the editorial board of the Institute of Physics journal 'Inverse Problems' since 2000, and Editor-In-Chief 2015-2019. He is widely known as one of the originators of the field of Diffuse Optical Tomography (DOT). A special focus of his research in the last decade has been Imaging from Coupled Physics and in particular Photoacoustic Tomography (PAT). He has worked more generally in the application of inverse problems to other medical imaging areas including Diffusion Tensor MRI, SPECT, PET and digital tomosynthesis.

F. PEYRIN is Director of Research (INSERM) in the CREATIS Laboratory (INSERM U1206, CNRS 5220, INSA-Lyon, Université Lyon I, Université de Lyon) specialized in medical imaging. She has been leading the team 'Tomographic imaging and Radiotherapy' in Creatis from 2009 to 2020.

Her research interest is in 3D biomedical imaging particularly in X-ray tomography, image reconstruction, inverse problems in biomedical imaging and wavelet-based methods. She has particularly been developing imaging methods for the characterization of bone tissue up to the nano scale. She is the author of more than, 200 peer-reviewed papers, 21 books chapters, 400 conferences (>180 as 4-pages IEEE style proceedings) and 100 invited talks.

She has been leading the LabEx PRIMES (Physique Radiobiology Medical Imaging and Simulation) from the Université de Lyon since 2012. She led the French GdR Stic Santé CNRS-INSERM from 2011 to 2014. Since 2002, she has been involved in the IEEE BISP (Bio Imaging and Signal processing) and since 2016 in the BIIP (Bio Imaging and Image processing) committees. She is currently chair of the EURASIP (European Association for Signal Processing) TAC (Technical Area Committee) BISA (Biomedical Image & Signal Analytics). She is an EAMBES Fellow.