

MATERIAL DECOMPOSITION PROBLEM IN SPECTRAL CT: A TRANSFER DEEP LEARNING APPROACH

*J Abascal**, *N Ducros**, *V Pronina**, *S Bussod**, *A Hauptmann^{‡,◇}*, *S Arridge[◇]*, *P Douek**, *F Peyrin**

* Univ Lyon, INSA-Lyon, CNRS, Inserm, CREATIS UMR 5220, U1206, Lyon, France

‡ Research Unit of Mathematical Sciences, University of Oulu, Oulu, Finland

◇ Department of Computer Science, University College London, London, United Kingdom

ABSTRACT

Current model-based variational methods used for solving the non-linear material decomposition problem in spectral computed tomography rely on prior knowledge of the scanner energy response, but this is generally unknown or spatially varying. We propose a two-step deep transfer learning approach that can learn the energy response of the scanner and its variation across the detector pixels. First, we pretrain U-Net on a large data set assuming ideal data, and, second, we fine-tune the pretrained model using few data corresponding to a non-ideal scenario. We assess it on numerical thorax phantoms that comprise soft tissue, bone and kidneys marked with gadolinium, which are built from the kits19 dataset. We find that the proposed method solves the material decomposition problem without prior knowledge of the scanner energy response. We compare our approach to a regularized Gauss-Newton method and obtain a superior image quality.

Index Terms— Spectral CT, inverse problem, deep learning, transfer learning

1. INTRODUCTION

The new generation of spectral computed tomography (SCT) scanners include photon-counting detectors, which count single photons and resolve their energy [1]. With this extra dimension, SCT allows for material decomposition, which opens up new diagnostic possibilities [2].

Material decomposition is a nonlinear inverse problem [3]. State-of-the art for solving this problem is based on variational approaches [4, 3] but the disadvantage of these techniques is that they rely on accurate knowledge of the forward model, including the scanner energy response, which is generally unknown.

Deep learning-based methods are expected to overcome these hurdles. On the one hand, they can learn the inverse problem without *a priori* assumptions about the scanner energy response. On the other hand, deep convolutional neural networks (CNN) have shown outstanding results in several image processing tasks [5] and are also promising for learning several inverse problems [6, 7, 8]. Deep

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.

learning has been also proposed for solving the material decomposition in the image domain [9].

The success of CNNs depends in part on the availability of large-scale annotated data sets [10]. However, in medical imaging and specifically in emerging imaging modalities such as spectral CT data are very limited. For this, transfer learning appears as relevant technique for medical imaging [11].

In this work we propose a deep learning approach in the projection domain based on a U-Net architecture [12] and a transfer learning technique that can learn the energy response and compare it to a previously used regularized Gauss-Newton (RGN) method [4] using numerical human thorax phantoms.

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 Θ projections. Let $\mathbf{s} = (s_{1,1}^1, \dots, s_{i,p}^\theta, \dots, s_{I,P}^\Theta)^\top$ be the measurement vector, where $s_{i,p}^\theta$ is the data measured in the i -th energy bin at the p -th pixel for 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. We denote

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

where G represents the forward model that maps $\boldsymbol{\rho}$ onto \mathbf{s} . The goal of SCT is to invert (1).

The forward model G can be seen as the composition of the linear X-ray transform and a non-linear spectral mixing operator. The X-ray transform X applies to each material independently, i.e.,

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

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 independently, i.e.,

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

where $\mathbf{a}^\theta = (a_{1,1}^\theta, \dots, a_{m,p}^\theta, \dots, a_{M,P}^\theta)^\top$ and $\mathbf{s}^\theta = (s_{1,1}^\theta, \dots, s_{i,p}^\theta, \dots, s_{I,P}^\theta)^\top$. In particular, we consider the following non linear mixing [13]

$$s_{i,p}^\theta = \int_{\mathcal{E}} n_0(E) d_{i,p}(E) \exp \left[- \sum_{m=1}^M a_{m,p}^\theta \tau_m(E) \right] dE \quad (4)$$

where \mathcal{E} is the energy range of the source, n_o is the source energy spectrum, $d_{i,p}$ is the detector response function at the p -th pixel for the i -th bin, and τ_m is the m -th material mass attenuation coefficient.

2.2. Image reconstruction

We solve the inverse spectral CT image reconstruction problem in two steps: material decomposition and tomographic reconstruction. Material decomposition is the inversion of (3), which is done projection by projection, while tomographic reconstruction is the inversion of (2). In this work, we consider the material decomposition problem with two different approaches, as described below. Assuming parallel geometry, we perform tomographic reconstruction using filtered back-projection algorithm with Ramp filter.

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 [4]:

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

where the data fidelity term is a weighted least square functional to approximate a Poisson distribution, $W^\theta = \text{diag}(1/\sqrt{s^\theta})$ is a weighting matrix, α is a regularization parameter, and R_m accounts for specific material regularization. We chose second-order and first-order Tikhonov regularization for soft tissue and bone, respectively, and Huber functional for contrast agent as suggested in [4]. Cost function (5) is efficiently minimized by using Gauss Newton algorithm (RGN); MATLAB code for RGN is available from Spectral X-ray image reconstruction SPRAY toolbox [14].

2.2.2. Material decomposition by deep learning method

Deep learning has been recently proposed for learning inverse problems [15, 16]. In this work, we want to learn

$$h_\beta : \mathbf{s}^\theta \mapsto \mathbf{a}^\theta \quad (6)$$

where β indicates the parameters of the neural network. We use a U-Net architecture [12] that consists of 14 hidden layers and three levels as shown in figure 1.

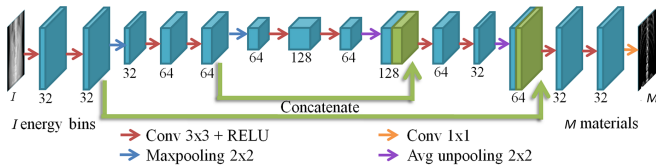


Fig. 1: U-Net architecture. Input is a stack of photon counting projections (I energy bins) and output is a stack of the decomposed material projection images (M materials) for a projection angle.

Learning means finding the parameters β that minimizes the following loss function

$$L(\beta) = \sum_{n=1}^N \|\mathbf{h}(\mathbf{s}^n; \beta) - \mathbf{a}^n\|^2 = \sum_{n=1}^N \sum_{m=1}^M \|\mathbf{h}_m(\mathbf{s}^n; \beta) - \mathbf{a}_m^n\|^2, \quad (7)$$

where $(\mathbf{s}^n, \mathbf{a}^n)$ are 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^{-4} and batch size of 16. Training and test losses were computed during training and early stopping was adopted to avoid overfitting. Both input and output projections are normalized before training.

2.3. Numerical phantom data

Numerical human thorax phantoms are comprised of soft tissue, bone and kidneys marked with gadolinium (Gd) and are built from CT volumes obtained from the kits19 challenge data set [17]. The distribution of concentration of gadolinium within the kidneys is generated as homogeneous, where the concentration for each phantom is obtained by sampling from a normal distribution centered in 0.1 g.cm^{-3} and standard deviation of 0.02 g.cm^{-3} . Phantoms have $640 \times 640 \times 100$ pixels and after projection each data set consists of 360 projections over a 180 angle span with size 909×100 detector pixels. Spectral CT data is simulated using SPRAY toolbox with the modeling parameters $(n_o(E), \tau_m(E), d_i(E))$ (4) used in [4]. Noisy data is considered by assuming a Poisson distribution for a total number of photons $N_0 = 10^7$.

To assess the effect of model deviations, we perturb the detector response function considering a pixel-dependent energy shift

$$\tilde{d}_{i,p}(E) = d_i(E - \Delta E_{i,p}), \quad (8)$$

where $\Delta E_{i,p}$ is the energy shift at pixel p and bin i . This shift is sampled from a normal distribution with a standard deviation of 3 keV. The data simulated with the ideal energy response $d_i(E)$ (which is constant across detector pixels) are referred to as *ideal*, while the data simulated with the perturbed energy response $\tilde{d}_{i,p}(E)$ are referred to as *non-ideal*.

For the ideal case, we split phantoms in 40 phantoms for training, 10 for validation and 10 for test, where there are 360 projection images per phantom. We pretrain U-Net on ideal data and compare it to RGN. In order to assess the effect of non-ideal data, we assess U-Net on a non-ideal version of the test data. Then, in order to account for deviations from ideality conditions, we applied U-Net with Transfer Learning (U-Net+TL), for which we fine tune the pretrained network (on ideal data) using one phantom for non-ideal data.

3. RESULTS

3.1. Material decomposed projections

Figure 2 shows phantom and decomposed projections. For an ideal detector, RGN leads to excessively smooth images for $\alpha = 1.4$; smaller values of α led to noisier images. Hence, RGN presents a strong trade-off between noise and regularization. On the contrary, U-Net removes noise while maintaining image quality.

For a non-ideal detector, both methods fail to provide accurate material decomposition (results shown only for U-Net). Assuming perfect data conditions translate into noisy projections. U-Net with transfer learning leads to a significantly reduction in noise.

3.2. Tomographic reconstructed images

Figure 3 shows phantom and tomographic reconstructed images. For ideal data, RGN leads to loss of details in soft tissue and bone. U-Net is able to decrease noise while maintaining image quality. For non-ideal data, U-Net presents ring artifacts and large crosstalk between soft tissue and Gd. However, U-Net+TL, which accounts for

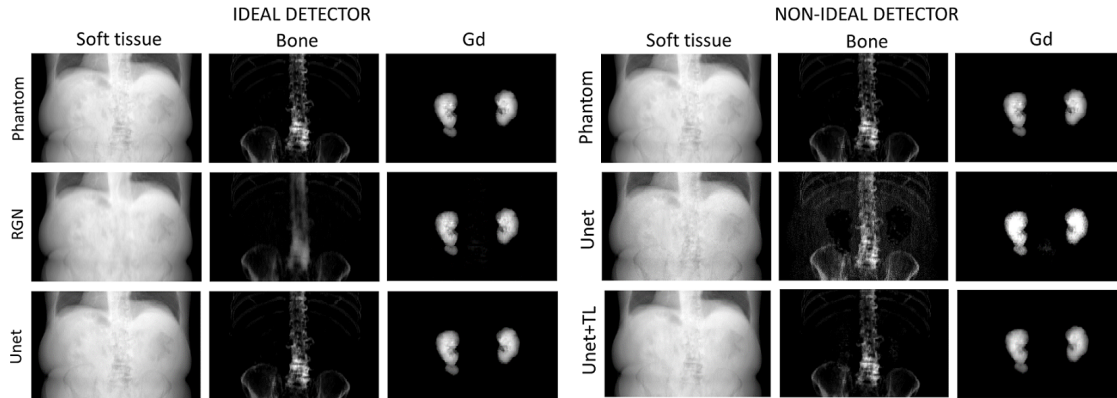


Fig. 2: Left: Phantom and material decomposed projections with RGN and U-Net for ideal data. Right: Phantom and material decomposed projections with U-Net with and without transfer learning for non-ideal data. Projection correspond to an angle view of 0° .

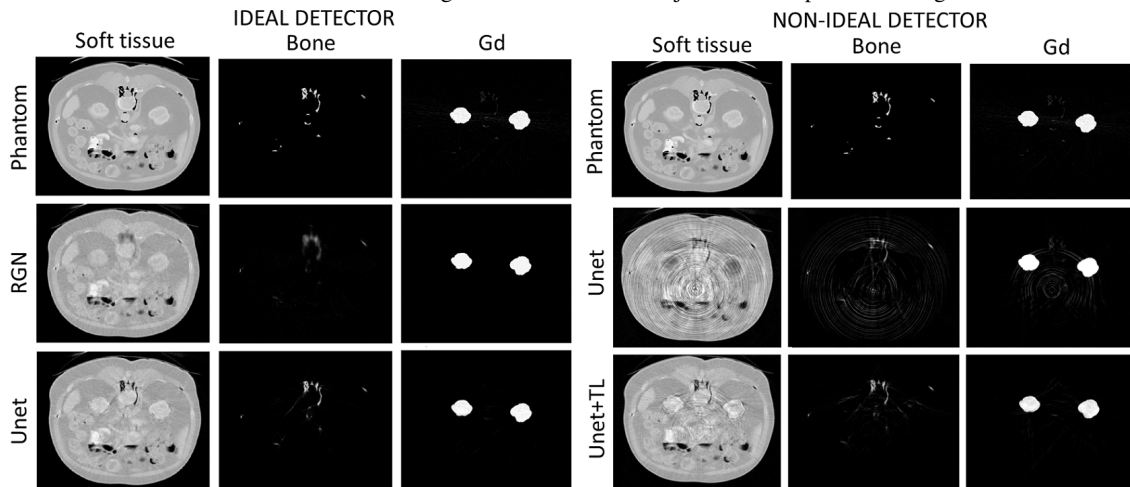


Fig. 3: Left: Phantom and tomographic reconstructed images for RGN and U-Net for ideal data. Right: Phantom and reconstructed images for U-Net with and without transfer learning for non-ideal data.

deviations from the ideal detector assumption, leads to reduction of most artefacts.

Figure 4 shows quantitative results. For ideal data, U-Net leads up to a three fold reduction on MSE on decomposed projections and reconstructions with respect to RGN on the three materials (MSE on the projections not shown). With respect to quantification of Gd, errors were less than 2% with both methods. RGN provides lower mean quantification error but larger variance is present in the histogram of Gd image (the latter not shown).

For non-ideal data, U-Net leads up to 4-fold increase in MSE and 20-fold increase in quantification error. Nevertheless, the proposed U-Net+TL yields 4-fold error reduction.

4. DISCUSSION

We have proposed a deep learning approach for solving the material decomposition problem that can learn the detector response of a SCT scanner. The approach is based on a U-Net architecture and a transfer learning approach to account for non ideal data.

For ideal data, U-Net provided superior image quality than RGN. 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 and it approximates the conditional mean when using the MSE loss [18]. This is particularly relevant for material

decomposition in the projection domain, as in this case the choice of prior distributions is less clear. A second disadvantage of variational methods is the high dependence on the regularization parameter.

In terms of quantification of gadolinium, for ideal data, both methods led to less than 2 % quantification error but RGN led to lower error than U-Net. However, inspection of the histogram reveals that RGN presents higher variance. High quantification performance for RGN could be expected as we have considered a low noise scenario and a perfect forward model for RGN. We found that RGN performance decreased for higher noise settings (results not shown).

For non ideal data, methods failed to provide satisfactory material decomposition, with the apparition of large ring artefacts and a 20-fold increase in quantification error. Ring artefacts arise from bias errors in the detector pixels. To mitigate these errors calibration is generally performed in practice. Nevertheless, even after calibration, errors may remain. Here, data discrepancy was modelled as a perturbation of the detector response where the size of the perturbation can be considered as moderate. Lower errors should be encountered in practice after recalibration.

To account for non ideal data, we proposed a transfer learning methodology based on pretraining on numerical phantoms and then fine tuning with one new data set corresponding to non-ideal data. These results show that few data are needed to correct for model deviations. Fine tuning can be understood as recalibrating the model.

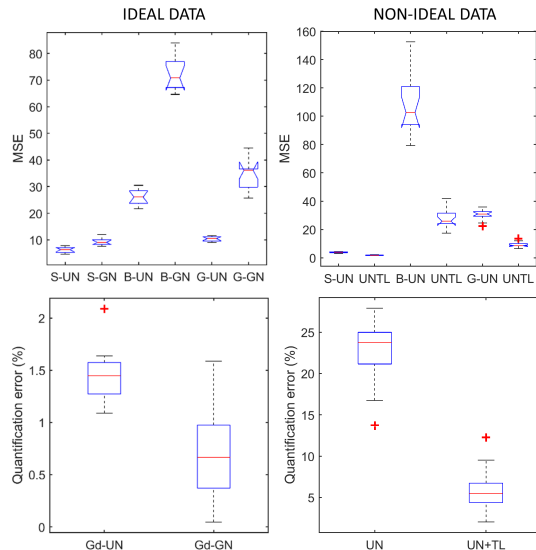


Fig. 4: MSE of reconstructed images, for soft tissue (S), bone (B) and Gd (G), and Gd quantification error. Left: U-Net (UN) vs RGN (GN). Right: UN vs UN with transfer learning (UNTL).

In practice, a first recalibration could require more data and computational training time yet subsequent recalibrations should need less resources.

There are different approaches to transfer learning. It is common routine during the fine tuning step to freeze some layers by fixing the weights to those values obtained in the pretraining step. However, we preferred to use the pretraining weights as an initial estimate but let them freedom to change [19]. One reason is that frozen layers are generally the first ones, which capture lower level features, but in our case it could be expected that the first layers are actually those that would overtake the recalibration. In any case, pretraining has been shown to have a regularization (prior) effect, making the solution to lie in a specific region of the solution space, with better generalization properties [19], so it seems ideal for spectral CT for which data are limited.

This work is subject to few limitations. We have assessed the effect of non-ideal data by perturbing the detector response by a shift in energy. While this model may not be completely exact in practice, it shows the effect of mismodelling the energy response of the scanner and that the proposed approach can learn these type of model deviations. With regard to the fine tuning step, it requires a new 'experimental' data set corresponding to non-ideal data. In practice, this new data set could be an experimental phantom, which properties are known. In this work, we consider a thorax phantom, which is not entirely realistic, but demonstrates the method's capabilities and shows that few data are needed for recalibration. Future work will also consider experimental data.

In conclusion, deep learning has a great potential for SCT as it allows for material decomposition without prior knowledge of the scanner energy response and provides superior image quality than previous variational approaches.

5. REFERENCES

- [1] E Roessl and R Proksa, "K-edge imaging in x-ray computed tomography using multi-bin photon counting detectors," *Phys. Med. Biol.*, vol. 52, no. 15, pp. 4679, 2007.
- [2] D P Cormode, E Roessl, A Thran, et al., "Atherosclerotic Plaque Composition: Analysis with Multicolor CT and Targeted Gold Nanoparticles1," *Radiology*, vol. 256, no. 3, pp. 774–782, 2010.
- [3] J F P J Abascal, N Ducros, and F Peyrin, "Nonlinear material decomposition using a regularized iterative scheme based on the Bregman distance," *Inverse Probl.*, vol. 34, no. 12, pp. 124003, 2018.
- [4] N Ducros, J F P-J Abascal, B Sixou, et al., "Regularization of nonlinear decomposition of spectral x-ray projection images," *Med. Phys.*, vol. 44, no. 9, pp. e174–e187, 2017.
- [5] Y LeCun, Y Bengio, and G Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
- [6] A Hauptmann, F Lucka, M Betcke, et al., "Model-Based Learning for Accelerated, Limited-View 3-D Photoacoustic Tomography," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1382–1393, 2018.
- [7] S Ravishankar, J Chul Ye, and J A Fessler, "Image Reconstruction: From Sparsity to Data-adaptive Methods and Machine Learning," Tech. Rep., 2019.
- [8] S Arridge, P Maass, et al., "Solving inverse problems using data-driven models," *Acta Numerica*, vol. 28, pp. 1–174, 2019.
- [9] M Touch, D P Clark, et al., "A neural network-based method for spectral distortion correction in photon counting x-ray CT," *Phys. Med. Biol.*, vol. 61, no. 16, pp. 6132, 2016.
- [10] A Krizhevsky, I Sutskever, and G E Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. of NIPS'12*, 2012, pp. 1097–1105.
- [11] H-C Shin, H R Roth, et al., "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285–1298, 2016.
- [12] O Ronneberger, P Fischer, and T Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," pp. 234–241. Springer, Cham, 2015.
- [13] J P Schlomka, E Roessl, et al., "Experimental feasibility of multi-energy photon-counting K-edge imaging in pre-clinical computed tomography," *Phys. Med. Biol.*, vol. 53, no. 15, pp. 4031, 2008.
- [14] N Ducros, "SPRAY - Spectral X-ray image reconstruction toolbox," 2017.
- [15] K H Jin, M T McCann, et al., "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, 2017.
- [16] E Kang, J Min, and J C Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. e360–e375, 2017.
- [17] N Heller, N Sathianathan, et al., "The KiTS19 Challenge Data: 300 Kidney Tumor Cases with Clinical Context, CT Semantic Segmentations, and Surgical Outcomes," Tech. Rep., 2019.
- [18] J Adler and O Oktem, "Deep Bayesian Inversion Computational uncertainty quantification for large scale inverse problems," Tech. Rep., 2018.
- [19] D Erhan, Y Bengio, et al., "Why does Unsupervised Pre-training Help Deep Learning?," *J. Mach. Learn. Res.*, pp. 625–660, 2010.