Multi-material Decomposition with Triple Layer Flat-Panel Detector CBCT using Model-based and Deep Learning Approaches

Xiao Jiang[1], Xiaoxuan Zhang[2], J. Webster Stayman[1], Grace J. Gang[2]

[1]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, USA

[2]Department of Radiology, University of Pennsylvania, Philadelphia, USA

Abstract Spectral CT has been investigated widely for a range of diagnostic applications with increasing potential interest for cone-beam CT (CBCT) applications. Current CBCT technology has largely focused on flat-panel detectors due to their relatively small form factor and ease of integration within a compact gantry that fits well in an interventional suite. The recent commercial availability of triple-layer flat panel detectors has provided a new avenue for spectral CBCT. In particular, while many spectral systems are limited to two channels with different energy sensitivity (e.g. dual layer detectors, kV-switching systems, etc.), a triple-layer system has the potential to be able to perform three material decomposition without additional constraints. Unfortunately, the spectral separation of a triple-layer panel is modest leading to a relatively ill-conditioned material decomposition problem (which consequently can be highly noise magnifying). In this work, we explore the possibility of three material decomposition and CBCT using a triple-layer panel and two sophisticated processing approaches: 1) model-based projection-domain material decomposition and 2) deep-learning-based projection-domain decomposition. Both approaches use simple filtered-backprojection of material line integral estimates to form 3D material maps. A simulation study with realistic measurement models is conducted using anthropomorphic phantoms and three material bases (water, calcium, and exogenous gadolinium contrast agent). A preliminary performance evaluation of reconstructed phantom data is provided to illustrate the potential of spectral CBCT using triple-layer detectors.

1 Introduction

Spectral x-ray imaging has the potential to enable a number of novel clinical diagnostics and CT image quality improvements [1] including non-contrast-enhanced image synthesis, virtual monoenergetic images, beam hardening and metal artifact reduction. The capability of spectral method to provide material separation similarly has the potential to enhance various clinical tasks like artery calcification detection and visualization, uric acid characterization/quantification, etc. Such applications are increasingly being developed and translated using diagnostic CT scanners. Research into the application of spectral imaging is also being investigated in cone-beam CT[1]. Cone-beam CT (CBCT) applications often target specific interventional procedures and their associated diagnostics. Applications include hemorrhage detection (including contrast agent extravasation), revascularization assessments, and vascular lumen characterization[2].

Tube voltage switching[3] and multi-layer flat-panel detector[4] are two approaches that have been investigated for spectral cone-beam CT data acquisitions. Tube voltage switching can provide large spectral separation and is compatible with the conventional flat-panel detectors (FPDs). However, this strategy requires a more complex x-ray generator and may have increased sensitivity to patient

and gantry motion due to projection mismatches between different energy channels. In contrast, recently available multi-layer flat-panel detectors provide projection data where spectral channels are collected simultaneously, minimizing cross-channel geometry mismatches and motion-induced artifacts. Dual-layer flat-panel detector [4] has been explored for radiography and interventional imaging systems. Strictly speaking, such two channel systems only permit differentiation between two materials. While a volume constraint could be added to permit a third material estimate, such a constraint does not hold universally (e.g. in the lungs).

Three-material decomposition is particularly important for separation of exogenous contrast agents and from anatomy[5]. Applications include angiography [6], perfusion studies, and lesion enhancement.

Recently available triple-layer flat panel detectors provide one potential avenue to produce three-material decompositions without an explicit volume constraint. However, such technology is challenged by the relatively poor spectral separability of these detectors (and consequent ill-conditioning of the decomposition problem). In this work, we conducted a study to investigate the potential of creating accurate water, calcium, and contrast density images using triple-layer CBCT. We developed and evaluated two processing schemes to produce 3D material estimates: 1) a model-based iterative approach; and 2) a deep-learning-based decomposition. Both approaches consider a projection domain decomposition followed by FDK reconstruction. Preliminary evaluations comparing the two approaches are provided.

2 Materials and Methods

The two material decomposition approaches are introduced in the following sections.

2.1. Model-based projection-domain decomposition

The general forward model for a multi-layer flat-panel detector can be written in matrix notation as [7]:

$$\bar{y}(q) = \mathbf{BS} \exp(-\mathbf{QA}\rho) = \mathbf{BS} \exp(-\mathbf{Q}l) \quad (1)$$

where $l \in \mathbb{R}^{jk}$ is the vectorized material density projection (i.e., the physical density ρ of each basis material forward projected by operator $\mathbf{A}$) with j pixels and k basis materials, $\mathbf{Q}$ stacks the mass attenuation coefficients of each basis material, and $\exp(-\mathbf{Q}l)$ represents the total attenuation in each energy bin. The matrix $\mathbf{S}$ characterizes

Poster Session 1

D2-P1-18

17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine   16 - 21 July 2023, Stony Brook, NY, USA

the overall spectral sensitivities of the system and $\mathbf{B}$ models the (potentially) layer-dependent blur kernels.

Assuming measurements y follow a multivariate Gaussian distribution, $y \sim \mathcal{N}(\bar{y}, \Sigma)$, the regularized likelihood objective function for material decomposition can be written as:

$$\hat{l} = \text{argmax}_l(\Phi(l; y)) \qquad (2)$$
$$\text{where } \Phi(l) = (y - \bar{y}(l))^T \Sigma^{-1}(y - \bar{y}(l)) + \beta l^T \mathbf{R} l$$

A quadratic regularization functin ($l^T \mathbf{R} l$) is used in this work that penalizes the differences between the 4-nearest neighboring pixels within each material map. We applied Newton's method to solve the objective. Within the nth iteration, the material maps are updated according to:

$$l_{n+1} = l_n - \alpha(\nabla_l^2 \Phi)^{-1} \nabla_l \Phi \qquad (3)$$

where the gradient, $\nabla_q \Phi$, and Hessian, $\nabla_q^2 \Phi$, of the objective are given by:

$$\nabla_q \Phi = \mathbf{Q}^T \mathbf{D_1} \mathbf{S}^T \mathbf{B}^T \Sigma^{-1}(\mathbf{BS}\exp(-\mathbf{Q}l) - y) + \beta \mathbf{R} l \qquad (4)$$
$$\nabla_q^2 \Phi \approx \mathbf{Q}^T \mathbf{D_1} \mathbf{S}^T \mathbf{B}^T \Sigma^{-1} \mathbf{BSD_1} \mathbf{Q} + \beta \mathbf{R} \qquad (5)$$

where $\mathbf{D_1} = \text{diag}\{\exp(-\mathbf{Q}l)\}$. The update step size, α, is empirically chosen to be 0.5.

We additionally adopted the following strategies to model additional physical effects, improve performance, and accelerate convergence: 1) The middle and bottom layer projections were registered to that of the top layer using an affine transformation to account for geometry mismatch and pixel grid misalignment among the layers; 2) The projections were pre-processed by deconvolving the blurs $\mathbf{B}$ using the Richardson–Lucy method [8], thus making the pixels approximately separable; 3) An initial material decomposition was first performed on 8x downsampled projection data, upsampled to the full resolution, and then used as initialization for the Newton update. The latter strategy increased robustness against local minima induced by image noise.

The application in this work focuses on contrast-enhanced studies using gadolinium as the contrast media. The three basis materials were therefore chosen to be water, calcium, and gadolinium. The material-specific regularization strength parameters, β, were chosen to minimize the error in the gadolinium images. The values are set to $2 \times$ $10^{-6}, 2 \times 10^{-5}$, and 2×10^{-4} for water, calcium, and gadolinium, respectively.

2.2. Deep-learning projection-domain decomposition

We have previously developed a deep learning network capable of performing three material decomposition using simulated data generated from a realistic forward model of the triple layer detector (Eq.1). The network architecture is shown in Fig. 1 The input to the network consists of a three-channel input formed by concatenating the three projections from the triple-layer detector, while the output consists of three basis material maps. Material decomposition is performed by a network following the ResUnet architecture with 9 residual blocks (ResBlock) [9]. Each block consists of two 3×3 convolution layers separated by a LeakyReLU activation layer, one residual connection adding the block input to output, and a final LeakyReLU activation layer. Each ResBlock in the encoder portion of the network reduces the dimension of the input by half and doubles the number of feature channels, while each ResBlock in the decoder portion does the opposite. A final 1×1 convolution layer is applied to reduce the number of channels to three - the number of basis materials.

For training, we adopted the following loss function:

$$\mathcal{L} = \mathcal{L}_p + \lambda_1 \mathcal{L}_{edge} + \lambda_2 \mathcal{L}_{consistency} \qquad (6)$$

The first term, $\mathcal{L}_p(l, \hat{l}) = \|l - \hat{l}\|_2^2$, is a typical MSE loss which quantifies the difference between the predicted material density projections, $\hat{l}$, and the ground truth, l. We additionally included a gradient-based loss term, $\mathcal{L}_{edge}(l, \hat{l}) = \|\nabla l - \nabla \hat{l}\|_1$, which has been explored to preserve spatial resolution[10]. The third term is a data consistency loss that penalized differences in the measurement (y) domain rather than the material density line integral (l) domain, i.e., $\mathcal{L}_{consistency}(y, \hat{l}) = \|\mathbf{BS}\exp(-\mathbf{Q}\hat{l}) - y\|_2^2$. Such loss functions have been investigated in previous work to incorporate a physics-driven constraint to the network output[11]. The scalars, λ_1, λ_2, control the relative weight of each term. Different combinations of λ_1, λ_2 were investigated for their impact on imaging performance. The optimal weighting was selected to minimize the MSE over all three material maps ($\lambda_1 = 10$ and $\lambda_2 = 1$). The network was implemented in PyTorch. We used the Adam optimizer with a batch size of 4 and
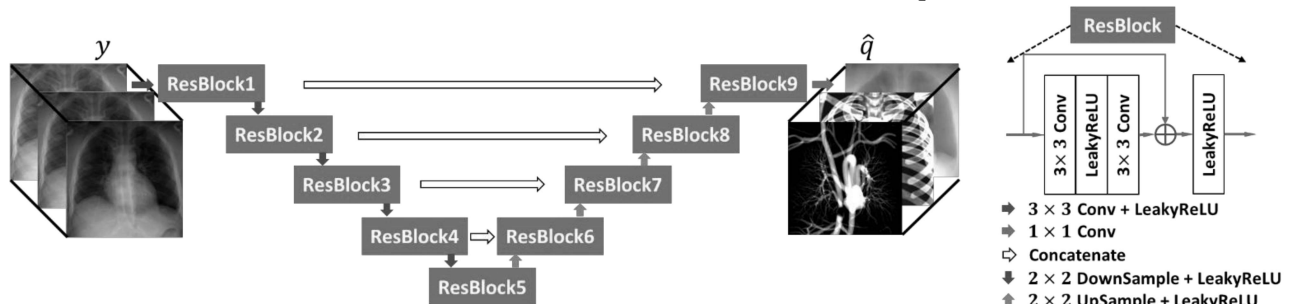

Figure 1: ResUnet network architecture with triple layer projection data (y) as an input, and predicted material projections ($\hat{q}$) as the output.

Poster Session 1   **D2-P1-18**

17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine   **16 - 21 July 2023, Stony Brook, NY, USA**
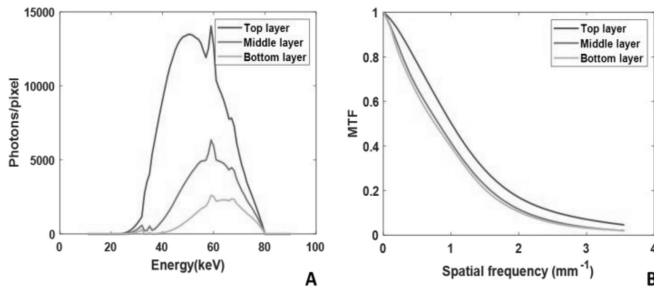
Figure 2: Spectral sensitivities (A) and MTFs for each channel (B).

terminated after 400 epochs based on empirical observation. The loss function typically decreased by <0.1% in the last 100 iterations.

2.3 Phantoms and data simulation

Imaigng phantoms in this work were generated using the XCAT package[12]. High resolution volumetric chest phantoms were genereated at a voxel size of 0.4 mm×0.4 mm×0.4 mm. The ground truth density of water and calcium are obtained using an image-domain decomposition method applied to two attenuation images generated at 60 keV and 100 keV. Vessels are uniformly assigned to 20 mg/ml of gadolinium. The ventricles and atria in the heart were intentionally not enhanced for better visualization of overalying vessels.

To generate simulated projection data for both the model- and learning-based methods, we used the detector blurs, **B**, and spectral sensitivities, **S**, shown in Fig.2. These characteristics are intended to model a realistic triple-layer panel comprised of three stacked indirect flat-panel detectors with ~250 µm of CsI in each layer. Projection data were simulated according to Eq.1 using a pixel size of 0.28mm×0.28mm. Noisy projections was obtained by adding independent Poisson noise before applying **B**.

To generate training data for the learning-based decomposition algorithm, we used 24 XCAT patient models to obtain different realizations of chest phantoms. We simulated 100 projections evenly distributed over 360º for each phantom to obtain 2400 projections, of which 2000 were used for training while the remaining 400 were used for validation. We randomly extracted 32 patches (256 pixels×256 pixels) from each projection, and randomly applied horizontal or vertical flips for data augmentation.

2.4 Reconstruction

The model-based and learning-based algorithms were applied to 360 projections uniformly distributed over 360º. The resulting estimated material density line integrals $\hat{l}$ were reconstructed using an in-house FDK algorithm to obtain the 3D density distribution (ρ).

2.5. Evaluation

We evelated the structural similarity index measure (SSIM) and root-mean-square error (RMSE) of the estimated 3D density distribution with ground truth. Note that the ground truth here is the FDK reconstruction of the ground truth mateial density line integrals from the XCAT phantom. We further compared line profiles through anatomical structures for spatial resolution comparison.

3 Results

Figure 3 shows the 3D density distribution of water, calcium, and gadolinium following the model- and learning-based decomposition methods.
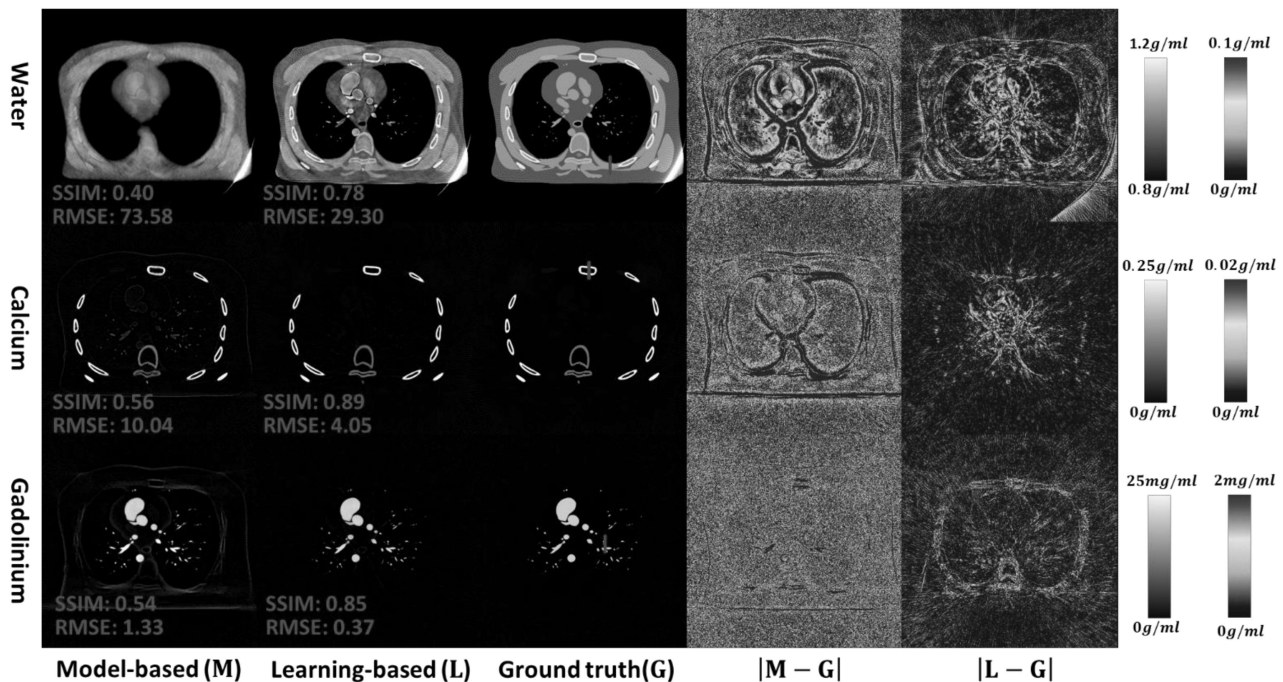

Figure 3: FDK reconstruction of material projections obtained by model-based decomposition, learning-based decomposition, and ground truth projections. The color images display the pixel-wise absolute error. RMSE unit: mg/ml.

Poster Session 1

D2-P1-18

17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine   16 - 21 July 2023, Stony Brook, NY, USA
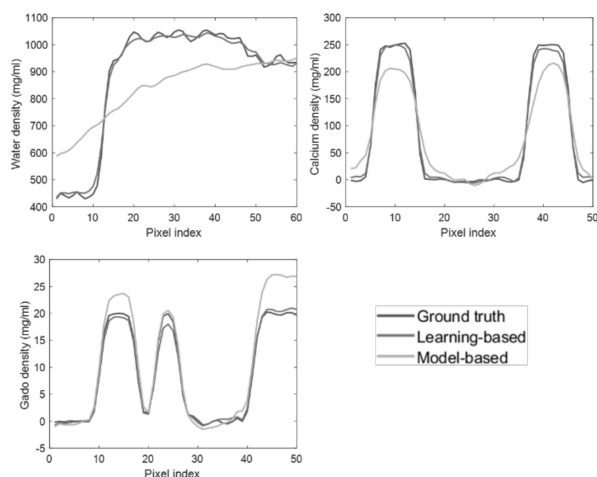
Figure 4: 1D profiles of the line segments in Fig. 3

From visual observation, the model-based method exhibits significant bias in all three material images – the water image appears overly smooth; lung parenchyma is absent in the water images but present on the calcium image; chest wall and patient boundary are present in the gadolium image. Larger errors are observed around tissue boundaries, consistent with previous observations of cross talk effects amongst material maps. Careful tuning and design of regularization can potentially mitigate such biases[13].

The learning-based method, on the other hand, is able to achieve low bias in the calcium and gadolinium images. The water image presents significant artifacts, likely from the propagation of different estimation bias from individual projections. The advantage of the learning-based method over model-based method is also reflected in the higher SSIM and lower RMSE in all three material images.

To further evaluate algorithm performance, we plotted the line profiles across anatomical structures in each material image. Consistent with visual observation, the model-based method produced blurred water and calcium images. The gadolinium image has comparable spatial resolution with the ground truth but exhibits an over-estimation of concentration by ~25%. The learning-based method shows better performance, with line profiles closely resembles those from the ground truth in both spatial resolution and density estimates.

4 Discussion and Conclusion

We have presented a simulation study to explore the potential of three material decomposition using a triple-layer panel. Such a system would fit within the current interventional imaging paradigm that largely uses flat-panel detectors for imaging. In our preliminary studies, we find that both model-based and deep-learning methods can provide three material estimates in a water-calcium-gadolinium decomposition problem, though imaging performance is significantly improved with the deep learning approach. We note that the deep learning approach

has a significantly faster processing speed making it more appropriate for the workflow of an interventional suite.

A number of limitations of this preliminary work are noted including the focus on gadolinium contrast (as opposed to the more popular iodine contrast agent) due to the greater potential for separation from calcium; a limited investigation of regularization strategies for the model-based approach; and a limited phantom evaluation. Despite these limitations, the technology appears to hold promise and we seek to continue to address these limitations in ongoing simulation studies and physical experiments.

References

[1] X. Jiang, H. Cui, Z. Liu, and L. Zhu, "Residual W-shape network (ResWnet) for dual-energy cone-beam CT imaging," in *7th International Conference on Image Formation in X-Ray Computed Tomography*, 2022, vol. 12304, pp. 488–492.

[2] K. Müller *et al.*, "Interventional dual-energy imaging—Feasibility of rapid kV-switching on a C-arm CT system," *Med Phys*, vol. 43, no. 10, pp. 5537–5546, 2016.

[3] R. Cassetta *et al.*, "Fast-switching dual energy cone beam computed tomography using the on-board imager of a commercial linear accelerator," *Phys Med Biol*, vol. 65, no. 1, p. 015013, 2020.

[4] L. Shi *et al.*, "Characterization and potential applications of a dual-layer flat-panel detector," *Med Phys*, vol. 47, no. 8, pp. 3332–3343, 2020.

[5] F. Schwarz, J. W. Nance Jr, B. Ruzsics, G. Bastarrika, A. Sterzik, and U. J. Schoepf, "Quantification of coronary artery calcium on the basis of dual-energy coronary CT angiography," *Radiology*, vol. 264, no. 3, pp. 700–707, 2012.

[6] C. Sandoval-Garcia *et al.*, "Comparison of the diagnostic utility of 4D-DSA with conventional 2D-and 3D-DSA in the diagnosis of cerebrovascular abnormalities," *American Journal of Neuroradiology*, vol. 38, no. 4, pp. 729–734, 2017.

[7] W. Wang *et al.*, "High-resolution model-based material decomposition in dual-layer flat-panel CBCT," *Med Phys*, vol. 48, no. 10, pp. 6375–6387, 2021.

[8] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *Astron J*, vol. 79, p. 745, 1974.

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[10] G. Wang and X. Hu, "Low-dose CT denoising using a Progressive Wasserstein generative adversarial network," *Comput Biol Med*, vol. 135, p. 104625, 2021.

[11] B. Yaman, S. A. H. Hosseini, S. Moeller, J. Ellermann, K. Uğurbil, and M. Akçakaya, "Self-supervised physics-based deep learning MRI reconstruction without fully-sampled data," in *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, 2020, pp. 921–925.

[12] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Med Phys*, vol. 37, no. 9, pp. 4902–4915, 2010.

[13] W. Wang, M. Tivnan, G. J. Gang, and J. W. Stayman, "Prospective prediction and control of image properties in model-based material decomposition for spectral CT," in *Medical Imaging 2020: Physics of Medical Imaging*, 2020, vol. 11312, pp. 475–480.