Solutions Spotlight



NEW! PlanCHECK[™] Automated plan quality verification and reporting – available as part of the SunCHECK[™] Platform





Varian Halcyon[™] System Support SNC Dosimetry[™] v3.6 and 3D SCANNER[™] for beam model verification and Halcyon Package for acceptance and verification



NEW! MultiPHAN™

An easy-to-use tool for comprehensive TG-142 daily alignment checks – verifying lasers to EPIDs to CBCT systems





Stereotactic Solutions NEW SRS MapCHECK[™] and StereoPHAN[™] for end-to-end testing and Reference Detector for interference-free dosimetry scanning





MR-Based Radiotherapy

Custom patient and machine QA solutions, including ArcCHECK®-MR, IC PROFILER™-MR, MICRO+™ MR, and Solid Water® HE

See more. Learn more. sunnuclear.com



A learning-based material decomposition pipeline for multi-energy x-ray imaging

Yanye Lu^{a)}

Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nuremberg, 91058 Erlangen, Germany

Advanced Therapies Siemens Healthineers, 91301 Forchheim, Germany

Markus Kowarschik

Advanced Therapies Siemens Healthineers, 91301 Forchheim, Germany

Xiaolin Huang

Institute of Image Processing and Pattern Recognition, Shanghai Jiao Tong University, 200240 Shanghai, P.R. China

Yan Xia

Radiological Sciences Lab, Stanford University, 94305 CA, USA

Jang-Hwan Choi

Division of Mechanical and Biomedical Engineering, Ewha Womans University, 03760 Seoul, Korea

Shuqing Chen and Shiyang Hu

Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nuremberg, 91058 Erlangen, Germany

Qiushi Ren

Department of Biomedical Engineering, Peking University, 100871 Beijing, China

Rebecca Fahrig

Advanced Therapies Siemens Healthineers, 91301 Forchheim, Germany Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nuremberg, 91058 Erlangen, Germany

Joachim Hornegger and Andreas Maier

Pattern Recognition Lab, Department of Computer Science, Friedrich-Alexander-University Erlangen-Nuremberg, 91058 Erlangen, Germany

(Received 16 February 2018; revised 18 November 2018; accepted for publication 22 November 2018; published 24 December 2018)

Purpose: Benefiting from multi-energy x-ray imaging technology, material decomposition facilitates the characterization of different materials in x-ray imaging. However, the performance of material decomposition is limited by the accuracy of the decomposition model. Due to the presence of non-ideal effects in x-ray imaging systems, it is difficult to explicitly build the imaging system models for material decomposition. As an alternative, this paper explores the feasibility of using machine learning approaches for material decomposition tasks.

Methods: In this work, we propose a learning-based pipeline to perform material decomposition. In this pipeline, the step of feature extraction is implemented to integrate more informative features, such as neighboring information, to facilitate material decomposition tasks, and the step of hold-out validation with continuous interleaved sampling is employed to perform model evaluation and selection. We demonstrate the material decomposition capability of our proposed pipeline with promising machine learning algorithms in both simulation and experimentation, the algorithms of which are artificial neural network (ANN), Random Tree, REPTree and Random Forest. The performance was quantitatively evaluated using a simulated XCAT phantom and an anthropomorphic torso phantom. In order to evaluate the proposed method, two measurement-based material decomposition methods were used as the reference methods for comparison studies. In addition, deep learning-based solutions were also investigated to complete this work as a comprehensive comparison of machine learning solution for material decomposition.

Results: In both the simulation study and the experimental study, the introduced machine learning algorithms are able to train models for the material decomposition tasks. With the application of neighboring information, the performance of each machine learning algorithm is strongly improved. Compared to the state-of-the-art method, the performance of ANN in the simulation study is an improvement of over 24% in the noiseless scenarios and over 169% in the noisy scenario, while the performance of the Random Forest is an improvement of over 40% and 165%, respectively. Similarly, the performance of ANN in the experimental study is an improvement of over 42% in the denoised

scenario and over 45% in the original scenario, while the performance of Random Forest is an improvement by over 33% and 40%, respectively.

Conclusions: The proposed pipeline is able to build generic material decomposition models for different scenarios, and it was validated by quantitative evaluation in both simulation and experimentation. Compared to the reference methods, appropriate features and machine learning algorithms can significantly improve material decomposition performance. The results indicate that it is feasible and promising to perform material decomposition using machine learning methods, and our study will facilitate future efforts toward clinical applications. © 2018 American Association of Physicists in Medicine [https://doi.org/10.1002/mp.13317]

Key words: deep learning, feature extraction, machine learning, material decomposition, model selection, multi-energy, spectral x-ray imaging

1. INTRODUCTION

Material decomposition can be performed either before computed tomography (CT) reconstruction in the projection domain or after CT reconstruction in the image domain.¹ The former was pioneered by Alvarez et al.,² and the latter by Brooks.3 In general, image-based material decomposition is inferior to projection-based material decomposition because the polychromatic properties of the spectra are lost during image reconstruction, which leads to material decomposition being less accurate.^{4,5} Projection-based material decomposition employs its corresponding decomposition models to numerically decompose the measured multi-energy projections into material-specific projections, which relies on the spectral information of each energy bin. However, the spectral information of projections is typically affected by the spectral distortions caused by x-ray detectors,^{1,6} which limits the accuracy of material decomposition. Since it is difficult to measure the polychromatic properties of the spectra in real scans, various methods have been proposed to estimate the spectral information for material decomposition. One family of these methods proposes explicitly modeling the distortion of detectors for the material decomposition models.^{7,8} As an alternative, measurementbased methods that utilize measured projections to estimate empirical material decomposition estimators have been proposed.^{5, 9–13} Such estimators can be estimated without exactly knowing the spectral information using the measurement-based methods, which usually consist of empirical calibrations and model estimations. The former measures the projections using known phantoms with the corresponding analytic descriptions, and the latter uses hypothetical transmission models that estimate the system-specific behavior of the scanner. This approximation is generally susceptible to the nonlinearity of measurement systems. To solve these difficulties, researchers have looked into sophisticated machine learning solutions.

Machine learning, which has had a large impact on many aspects of science and industry, concerns the construction and study of systems that can learn from data. As an alternative, supervised machine learning can be utilized to model the relationship between the aforementioned spectral measurements and the relevant measurement system. This relationship is essential for measurement-based material decomposition, whose modeling can be processed as estimating hypothesis models using machine learning algorithms. Several related approaches have been reported in the literature. Lee et al.¹⁴ applied a feed-forward neural network to decompose multienergy images into material path-length images based on the simulated images of a three-dimensional head phantom. Zimmerman et al.¹⁵ used a feed-forward neural network to approximate the functional relationship between the measured projections and the basis material thickness. Touch et al.¹⁶ proposed an artificial neural network (ANN)-based spectral distortion correction mechanism, which trains an ANN to calibrate the spectral distortion, improving the accuracy of material decomposition. However, these studies used ANN only to estimate material composition. Other machine learning algorithms such as decision trees are also appropriate for modeling the relationship. Notably, only spectral measurements were used for training models in these previous studies, and these models are susceptible to noise. Also, the choices of the hyperparameters of machine learning algorithms are essential for the model learning, and should be an important consideration. In addition, state-of-the-art neural networks such as deep learning approaches should be investigated as well.

In our previous works,^{17,18} we demonstrated the preliminary results of performing material decomposition using machine learning. A registration-based method of acquiring labels and the potential of several image-based features have been investigated in the two papers. Building on this foundation, more comprehensive studies have been investigated, and we propose here a learning-based pipeline to perform material decomposition. The aim of this pipeline is to address material decomposition issues in accordance with the concept of machine learning systems. Compared to the conventional material decomposition approaches, the pipeline is able to perform material decomposition without explicitly knowing the transmission model of imaging systems. In addition, the pipeline employs not only spectral measurements, but also informative features such as neighboring information. We demonstrate the proposed pipeline in both the simulation and experimentation, and the decomposition results are quantitatively evaluated to investigate the feasibility and performance. Two measurement-based material decomposition methods (a representative one and a state-of-the-art one) were implemented as the reference methods for comparison studies. The results demonstrate that the proposed pipeline, using appropriate informative features, is able to suppress noise and strongly improve material decomposition. In addition, two state-of-the-art deep learning approaches were also investigated as a pilot study for a comprehensive comparison of machine learning solutions for material decomposition tasks.

2. MATERIALS AND METHODS

2.A. X-ray physics

According to the Beer–Lambert law, if the input x-ray photons are monoenergetic, the intensity of non-attenuated x-ray photons passing through a material can be computed as $I = I^0 e^{-\mu(E)l}$, where I^0 represents the intensity of x-ray photons before entering the material, *E* describes the photon energy, and $\mu(E)$ and *l* represent the spectral attenuation coefficient of the material at *E* and the path length of the material, respectively. After applying the minus logarithmic operation, the projection *p* can be expressed as $p = -\ln(\frac{I}{I^0}) = \mu(E)l$. Normally the scanning object contains various materials $i = 1 \cdots M$, resulting in $p = \sum_{i=1}^{M} p^i$ based on the mixture rule, where the material-specific projection p^i can be expressed as $p^i = -\ln(\frac{I_i}{I_i^0}) = \mu_i(E)l_i$, in which I_i^0 denotes the material-dependent I^0 of *i*. The measurements acquired from an x-ray image system yield:

$$q_j = -\ln \int [S_j(E)e^{-\sum_{i=1}^M \mu_i(E)l_i} + X(E)] dE_j$$

where $S_j(E)$ represents the distribution function of the x-ray spectrum used in the scan with the index j = 1...B. Note that j counts the independent spectral bins in multi-energy x-ray imaging, and B is the total number of kV settings in multiplekV detection, or the total number of energy bins in energyresolved spectral detection. X(E) represents the error function of the measurement, which includes several phenomena, such as photon scattering, detector distortion and signal noise. In practice, X(E) is difficult to model due to nonlinearity. As shown in the functions, we can obtain multi-energy measurements q_1, \dots, q_B from the same object from a spectral X-ray imaging system.

2.B. Material decomposition model estimation

The aim of this work is to build material-specific decomposition models to decompose the effective material-specific projection \hat{p}^i from the multiple energy measurements q_1, \dots, q_B . A target function of the material decomposition model is defined as $\hat{p}^i = D_i(\mathbf{q})$, where $D_i(\mathbf{q})$ is an as-yetunknown general function estimator of the basis material p^i with the multi-energy measurements $\mathbf{q} = (q_1, \dots, q_B)$. The goal of the estimator is to learn a mapping from the measurements to an objective quantity. To learn this mapping, a hypothesis model $H_i(\mathbf{x})$ is defined to approximate $D_i(\mathbf{q})$, where \mathbf{x} is a feature vector that numerically represents the measurements \mathbf{q} . Each element in \mathbf{x} , which is a so-called feature, represents a distinct description of \mathbf{q} . The concept of general function estimation is in accordance with machine learning systems. Thus, a material decomposition task can be transferred to a machine learning problem. In machine learning, the input measurements \mathbf{q} is also termed as an instance, and typically needs to be labeled with an output y^i in supervised learning. Facilitated by these instance-label pairs, a supervised learning procedure generates a machine learning system that can predict correct outcomes when applied to new input measurements.

On this basis, we first extract the informative feature vector \mathbf{x} from the observation \mathbf{q} using a feature extraction step of $\mathbf{x} = F(\mathbf{q})$, where F represents the corresponding feature extraction that can be implemented to build informative values from q using different approaches. Then, material decomposition models are trained from a large number of training examples of labeled feature vectors $(\mathbf{x}_w, \mathbf{y}_w^i)$ using machine learning algorithms, where $w = 1, \dots, W$. The model constructions rely on various machine learning algorithms, where the hypothesis model $H_i(\mathbf{x})$ is optimized to find the most optimal choice $H_i(\mathbf{x})$ that minimizes the expected value of a loss function \mathcal{L} on the training samples $H_i(\mathbf{x}) =$ $\arg\min_{\boldsymbol{\mu}} \mathcal{L}(H_i^T(\mathbf{x}_w, \mathbf{y}_w^i)), \text{ where } H_i^T(\mathbf{x}_w, \mathbf{y}_w^i) \text{ denotes the } H_i(\mathbf{x})$ using labeled feature vectors. Once the optimal hypothesis model $H_i(\mathbf{x})$ is determined, we can obtain an estimated material decomposition model $D_i(\mathbf{q})$ to perform material decomposition.

2.C. Leaning-based material decomposition pipeline

In light of the aforementioned theories, we propose a learning-based pipeline for material decomposition, which addresses material decomposition tasks in the form of solving a regression problem. The pipeline trains material-specific decomposition models to reconstruct material-specific images. Each model is equivalent to a pattern recognition system, whose input and output are the multi-energy measurements **q** from a multi-energy x-ray imaging equipment and the effective material-specific image p^i , respectively. The proposed pipeline is illustrated in Fig. 1, which contains two phases. The models are trained in the model learning phase using a learning dataset and can be used for material decomposition tasks afterwards in the material decomposition phase. In this work, a learning set of the instance-label pairs $\{(q_g, p_g^i)\}_{g=1}^G$ of multi-energy measurements $\{q_g\}_{g=1}^G$ and their corresponding material-specific image $\{p_g^i\}_{g=1}^G$ are required to train material-specific decomposition models, where g is the index of the pair of q_g and p_g^i . First, the learning set $\{(\boldsymbol{q}_g, \boldsymbol{p}_g^i)\}_{g=1}^G$ is split into two parts of a training set $\{(\boldsymbol{q}_n, \boldsymbol{p}_n^i)\}_{n=1}^N$ and a validation set $\{(\boldsymbol{q}_m, \boldsymbol{p}_m^i)\}_{m=1}^M$ by an evaluation strategy of machine learning, where n and m are the index, respectively. Then, the two sets are processed to obtain feature vectors by a feature extractor, whose purpose is to extract the feature information from the datasets based on relevant theories. These feature vectors from the training set are employed to learn material decomposition models for recognizing the input with a target output using machine learning algorithms. Since the models vary with the hyperparameters of the machine learning algorithms, once the models have been learned, the performance is quantitatively evaluated on the validation set, facilitating parameter optimization and model selection. Eventually, selected decomposition models can be applied to the objective multi-energy measurements for material decomposition tasks in the material decomposition phase, where the input multi-energy measurements also need to be processed by the same feature extractor with the same parameters that were used in the learning phase. We will elaborate these steps in detail in the following subsections.

2.C.1 Feature extraction

In Section 2. B, we introduced the feature extraction step, which aims at extracting the relevant information that characterizes each attribute of processing data. Such a step in the pipeline can extract more informative features from raw spectral measurements for the material decomposition tasks. We demonstrate two feature extraction strategies in this paper. The feature extractors extract feature information from the multi-energy measurements pixel-by-pixel to build the feature vectors. For instance in the training set, an instance-label pair $(\mathbf{q}_n, \mathbf{p}_n^i)$ is formed into $\{(\mathbf{V}_s^{q_n}, \mathbf{V}_s^{p_n^i})\}_{s=1}^S$, where $\mathbf{V}_s^{q_n}$ represents a vector of the respective information at the *s*-th pixel in each energy bin of $\mathbf{q}_n, \mathbf{V}_s^{p_n^i}$ is the label of the *s*-th pixel in p_n^i , and *S* is the total number of pixels of p_n^i . Each element in the vector $\mathbf{V}_s^{q_n}$ represents a feature, and the total number of the features is related to the total number of energy channels *B*.

Afterwards, the training set $\{\{(\mathbf{V}_{s}^{\mathbf{q}_{n}}, V_{s}^{p_{n}^{i}})\}_{s=1}^{S}\}_{n=1}^{N}$ is used for learning models by machine learning algorithms.

Because X-ray attenuation coefficients are material- and energy dependent, their behaviors are adequate features for classifying different materials. Intuitively, the pixel values of a multi-energy projection are informative features because they represent the x-ray attenuation behaviors in the projection domain. Therefore, a raw-pixel-value feature extractor was employed to extract the pixel values of each energy channel from the multi-energy measurements, and these values are used to represent the spectral measurement information. Thus, it yields $\mathbf{V}_{s}^{\mathbf{q}_{n}} = (q_{n,1}^{s}, \cdots, q_{n,B}^{s})$ and $V_{s}^{p_{n}^{i}} = p_{n,s}^{i}$. In order to integrate more informative features such as

In order to integrate more informative features such as neighboring information for training models, a mean-variation-median (MVM) feature extractor was also employed, which extracts not only the value of the target pixel at each energy channel, but also the mean, the variation, and the median of the pixels within a radius of σ pixels from the target pixel. Thus, it yields $\mathbf{V}_{s}^{\mathbf{q}_{n}} = (q_{n,1}^{s}, \cdots, q_{n,B}^{s}, v_{\sigma,mean}^{q_{n}^{s},1}, \cdots, v_{\sigma,wariation}^{q_{n}^{s},B}, v_{\sigma,meain}^{q_{n}^{s},1}, \cdots, v_{\sigma,wariation}^{q_{n}^{s},B}, v_{\sigma,meain}^{q_{n}^{s},1}, \cdots, v_{\sigma,meain}^{q_{n}^{s},1} = p_{n,s}^{i}$. This operation creates additional neighboring information features of the target pixel based on image patches. In this work, σ was set to 1, 2, 4, and 8 for each instance, yielding $B \times 13$ features in total.

2.C.2 Machine learning algorithms

Many machine learning algorithms are applicable to the proposed pipeline. In this work, we investigate some promising ones and evaluate their performance quantitatively to



Fig. 1. Schematic description of the learning-based material decomposition pipeline. [Color figure can be viewed at wileyonlinelibrary.com]

demonstrate the feasibility of the learning-based material decomposition pipeline. All the machine learning algorithms were implemented using the Waikato Environment for Knowledge Analysis (Weka)¹⁹ application programming interface (API).

Artificial neural network: Artificial neural networks (ANN) are massively parallel interconnected networks of simple, adaptive elements, which are intended to interact with objects of the real world in the same way as biological nervous systems do. The ANN employed in this work is a feed-forward multilayer perceptron²⁰ that fully connects many simple perceptron-like neurons in a hierarchical structure, training the network using the stochastic gradient descent (SGD) method²¹ and the backpropagation method.²² The idea of the SGD is to estimate the global gradient of all training inputs by averaging over the gradients that are iteratively computed from a randomly chosen mini-batch of training inputs. The picking out of training inputs lasts until the training inputs are exhausted, which is termed an epoch of training. The batch size and epochs were both set to 100. Learning rate and momentum are two hyperparameters in the backpropagation method. The former adjusts the step size of the SGD and the later is a method that helps accelerate the SGD in the relevant direction and dampens oscillations, leading to faster convergence. The learning rate and momentum was set to 0.3 and 0.2, respectively. The architecture of the ANN is composed of one input layer, one output layer and multiple hidden layers. The number of neurons of the input layer corresponds to the feature numbers, resulting in B input neurons for the raw-pixel-value features and $B \times 13$ input neurons for the MVM features. The output layer contains one neuron, which corresponds to the effective material-specific images. Since we used three energy channels (B = 3) in this work, there were 4 hidden layers that each contained 3 neurons while using the raw-pixel-value features, and 2 hidden layers that contained 39 neurons while using the MVM features. Each neuron represents a sigmoid activation function to process the output of neuron except the neuron of the output layer being a linear unit.

Decision trees: The purpose of decision tree learning, which uses recursive partitioning of repeatedly splitting on the values of features under divide and conquer strategy, is the construction of a decision tree model that has high accuracy of predicting new instances. A decision tree model is composed of a root node that is the topmost node in a tree structure, several internal nodes that correspond to feature tests, and several leaf nodes that hold the decision outcomes of the labels. In prediction, a series of feature tests are performed starting from the root node, and the predicted result is obtained at a leaf node. We used three decision trees: Random Tree,²³ REPTree,²⁴ and Random Forest.²³ All three use information gain²⁵ for the splitting strategy. REPTree considers all features to split at each node and performs reduced-error pruning (REP), whereas Random Tree

randomly picks out a number of features instead of all features for the tree induction process, but performs no pruning. However, the Random Tree in this work uses all features in order to be in accordance with the REPTree. Random Forest generates multiple Random Trees in parallel by using bootstrap sampling of the training data, and then combines these trees by averaging their outputs. The picking-out features parameter of these Random Trees in the Random Forest was set to the logarithm of the number of features, rounded up, in order to incorporate randomized feature selection. The sampling bag size was set to be 100% and the number of iterations was set to be 20. All algorithms set 1 as the minimum number of instances per leaf, 0.001 as the minimum variance proportion of train variance for the split, and the maximum depth of the tree to be unlimited.

2.C.3 Model evaluation and selection

The aim of the model selection is to find the appropriate hyperparameters of the pipeline, such as the parameters of the feature extractor and the machine learning algorithms, to generate models with superior generalization capabilities on the validation sets. Since parameter optimization is essential in the pipeline, the concept of machine learning is incorporated into the pipeline for facilitating model evaluation and parameter selection. We used the hold-out validation strategy to split the learning sets in this work. The strategy of sampling is that the training set was sampled with a fixed interval and then the remaining dataset in the learning set was used as the validation set. Ordinary quantitative measurements were not utilized in this work, instead the generalization capabilities were quantified by the Pearson's correlation coefficient (R) and the structural similarity (SSIM) index²⁶ between the decomposed material-specific images \hat{p} and the labeled material-specific images p. The correlation coefficient R measures the linear dependence between two images, where 0 indicates no linear correlation and 1 indicates total positive linear correlation. The formula for R is:

$$R_{\hat{p},p} = \frac{cov(\hat{p},p)}{\sigma_{\hat{p}}\sigma_{p}},$$

where $cov(\hat{p}, p)$ is the covariance, and $\sigma_{\hat{p}}$ and σ_{p} is the standard deviation of \hat{p} and p, respectively. The SSIM index measures the similarity of structural information in two images, where 0 indicates no similarity and 1 indicates total positive similarity. The formula for SSIM is:

$$SSIM_{\hat{p},p} = \frac{(2\mu_{\hat{p}}\mu_p + C_1)(2cov(\hat{p},p) + C_2)}{(\mu_{\hat{p}}^2 + \mu_p^2 + C_1)(\sigma_{\hat{p}}^2 + \sigma_p^2 + C_2)}$$

where $\mu_{\hat{p}}$ and μ_p is the mean of \hat{p} and p, respectively. C_1 and C_2 are two variables to stabilize the division with a weaker denominator. However, both C_1 and C_2 were set to 0 in this work because the images are gray-scale. The mean value and standard deviation of all the quantitative results from each scenario were used to evaluate the overall performance of the

material decomposition models. Finally, the models with superior overall performance were selected to perform material decomposition tasks.

2.D. Experimental setup

A simulation study and an experimental study were carried out to evaluate the proposed pipeline for material decomposition tasks, and we focus on projection-based material decomposition in this work. All methods were implemented in the Java-based framework CONRAD.²⁷

2.D.1 Data generation

In the simulation study, we performed energy-resolved detection using a virtual research Artis zeego C-arm angiography system (Siemens Healthineers, Forchheim, Germany) implemented in CONRAD²⁷ to generate multi-energy projections and material-specific projections. The geometry and xray spectrum (12 kVp, 0.65 mAs) were simulated based on the system. A flat panel photon-counting energy-resolving detector of 620×480 pixels with a pixel size of 0.616×0.616 mm was implemented to detect three energy bins (10-40, 40-81, and 81-125 keV) with a cross-talk of 3 keV. The source-to-patient distance was 700 mm while the source-to-detector distance was 1200 mm. We simulated three short scan scenarios using an append buffer-based rendering procedure,²⁸ and acquired the multi-energy projections of a modified XCAT phantom²⁹ containing bone; Ultravist370 (an iodinated contrast agent, 370 mg Iodine per mL, Bayer, Germany); and soft tissues such as liver, heart, and bone marrow. The energy-dependent x-ray absorption coefficients for the compounds in the phantom were obtained from the NIST database.³⁰ All the simulated scenarios used



Fig. 2. Example images showing the two simulation scenarios of the **XCAT** phantom (top row and middle row) and the real scan scenario of the **Torso** phantom (bottom row) for model tests. The right images present the corresponding material decomposition objects in this work, which are from the central projections of each scenario.

short scans (an angular scanning range of π plus the fanangle, 198.4 degrees in total) with an average angular increment of 1.6 degrees around the torso. One scenario was used for learning models, while the other two scenarios were used for testing. As shown in Supporting Information Fig. S1, the FOV of the learning scenario is centered around the heart to focus on the costal arch and coronary arteries that were filled with Ultravist370; both of the FOVs of the two test scenarios focus on a different area of the phantom, which have 40 mm translation over the learning scenario in the horizontal direction. In addition, one test scenario includes the cardiac motion of heart beats and the respiratory motion from full exhale to full inhale. The ground truth of the material decomposition objects were also simulated as monochromatic material-specific projections (Fig. 2) for labeling and performance evaluation. For each scenario, both noiseless projections and noisy projections (Poisson noise) were created.

In the experimental study, we scanned an anthropomorphic torso phantom (Sawbones Europe AB, Malmo, Sweden) containing a SAWBONES spine with biopsy needles inserted at different x-ray tube voltage setting of 50 kVp (4.3 mAs), 81 kVp (2.5 mAs), and 125 kVp (0.65 mAs) using a Siemens Artis zeego angiography CT system to generate multienergy projections. The flat panel detector was operated with 4×4 binning that allows an effective resolution of 620×480 pixels with a pixel size of 0.616×0.616 mm. The source-to-isocenter distance was 700 mm while the source-to-detector distance was 1200 mm. Acquisition of short scans (an angular scanning range of π plus the fanangle, 198.4 degrees in total) with an average angular increment of 1.6 degrees was performed. In order to obtain the ground truth projections of the needle, we performed a fourth scan at 125 kVp (0.65 mAs) after carefully removing the needles. Subtraction of the two 125 kVp scans yielded material-specific projections of the needle for labeling and performance evaluation. As shown in Supporting Information Fig. S2, two experimental scenarios were acquired. The scenario with one biopsy needle inserted was used for learning models, whereas the scenario with two needles inserted at different positions was used for testing. The experimental

data were preprocessed by a joint bilateral filter $(JBF)^{31}$ for noise reduction.

2.D.2 Experimental protocol

The experiments were performed in accordance with the proposed learning-based material decomposition pipeline. We used a two-stage splitting strategy on the learning set to yield the training set and the validation set for learning decomposition models. A learning set from the learning scenarios was first split into two datasets using the hold-out strategy with a sampling interval of 1 projection, and then one of the datasets was used as the validation set for model evaluation. Then we applied the hold-out strategies with different sampling intervals of 0, 1, 2, 3, 5, 8, and 16 (corresponding hold-out rates relative to the original dataset of 50.00%, 25.00%, 16.94%, 12.90%, 8.87%, 5.65%, and 3.23%, respectively) to the temporary dataset to yield the training sets, permitting the analysis of the hold-out rate strategy. The splitting strategy is demonstrated in Fig. 3, illustrating that the models learned from different hold-out rates are evaluated using the same validation set. The hyperparameters, as elaborated in Sections 2.C.1 and 2.C.2, were chosen by grid search (i.e., parameter sweep) on the validation sets empirically. The model learning was done using the mixed training set of all noise levels (noiseless and denoised projections in the simulation study, original and denoised projections in the experimental study). Once the models were built, they were quantitatively evaluated by the individual validation set at different noise levels using the quantitative evaluation methods, as elaborated in Section 2.C.3, for model selection. Finally, the models with superior performance were selected to perform material decomposition on a test scenario in the decomposition phase and the results are presented in Section 3.

2.D.3 Comparison study

In order to evaluate the proposed method, we used two measurement-based material decomposition methods as the reference methods for comparison studies. One is an



Fig. 3. Schematic description of the splitting strategy for generating the training sets in the learning phase. [Color figure can be viewed at wileyonlinelibrary. com]

estimator with polynomial fitting (namely Ref-Poly),⁵ which uses polynomial approximation to estimate the nonlinear behavior of the beam polychromacity. In this work, the degree of the polynomial was set to 3 (B = 3) and a standard least-squares multiple linear regression was used to estimate coefficients. Another one is the A-Table method (namely Ref-A-Table),^{11,13} which is a state-of-the-art decomposition method that uses a first-order Taylor series expansion and maximum likelihood estimation to approximate the relationship between spectral measurements and basis materials.

3. RESULTS

3.A. Simulation study

Table I shows the performance on R and SSIM (average \pm standard deviation) across all quantitative measurements in each scenario at the hold-out rate of 50%. All the machine learning algorithms were able to train models for the material decomposition tasks using the spectral measurements. The ANN, the REPTree, and the Random Forest demonstrated superior performance compared to the reference methods. However, these models are susceptible to noise and motion phenomena in this study, as indicated by the limited performance in the noisy scenarios and the motion scenarios, respectively. Thus, the neighboring information was employed to address this problem. As shown in Table I, the performance of the ANN in the motion scenario is improved by over 11% in the noiseless scenarios and by over 91% in the noisy scenario, while the performance of the Random Forest is improved by over 20% and 127%, respectively. Compared to the A-Table method, the performance of the ANN is improved by over 24% in the noiseless scenarios and by over 169% in the noisy scenario, while the performance of the Random Forest is improved by over 40% and 165%, respectively.

Figure 4 and Supporting Information Fig. S3 demonstrate the decomposed material-specific projections of the coronary arteries filled with Ultravist370 in the two test simulation scenarios at the hold-out rate of 50%. The contrast of the images was also enhanced by a histogram stretching with 0.3% saturated pixels and normalized to [0, 1]. There are overlaps among different materials (Fig. 2) in the central projection of each scenario, resulting in challenges for material decomposition tasks. Therefore, the results from such projections were selected to demonstrate the performance of each method. As shown in the images, the Random Forest was able to preserve more details of the vessels using the raw-pixel-value features in the noiseless scenario, but the performance also deteriorated in the noisy scenario, just as for the other algorithms. The application of the neighboring information has improved the performance of the machine learning algorithms, especially in the noisy scenarios. In general, despite a little residue of the bone, the ANN and the Random Forest have identified more vessels using the MVM features, even in the noisy scenarios.

Results of the sampling interval analysis are presented in Supporting Information Figs. S4 and S5. It can be seen that

ല	
Ę	
2	
2	
Σ	
-	
μ	
÷	
þ.	
~	
P	
aı	
Ś	
<u>e</u>	
Ξ.	
at	
Ъ	
e)	
[g]	
2	
G	
<u> </u>	
d.	
ż	
a	
1	
Ĕ	
Ŧ	
ਿ	
\sim	
문	
-5	
2	
SC	
٠Ĕ	
a	
er	
ũ	
50 E	
5	
N.	
¥.	
st	
te	
5	
Ă	
1	
.Ħ	
ŝ	
.е	
er	
r.	
a	
S.	
a	
E .	
ĭ	
8	
0	
ř	
<u> </u>	
-	
0	
S O	
o suo	
ions o	
ctions o	
jections o	
rojections o	
projections o	
Il projections o	
all projections o	
ss all projections o	
oss all projections o	
cross all projections o	
across all projections o	
n across all projections o	
ion across all projections o	
ation across all projections o	
viation across all projections o	
eviation across all projections or	
deviation across all projections or	
d deviation across all projections or	
ard deviation across all projections or	
ndard deviation across all projections o	
andard deviation across all projections or	
standard deviation across all projections or	
- standard deviation across all projections or	
\pm standard deviation across all projections or	
$1 \pm $ standard deviation across all projections o	
$IM \pm standard deviation across all projections or$	
SIM \pm standard deviation across all projections o	
SSIM \pm standard deviation across all projections o	
n SSIM \pm standard deviation across all projections o	
can SSIM \pm standard deviation across all projections o	
nean SSIM \pm standard deviation across all projections o	
I mean SSIM \pm standard deviation across all projections o	
nd mean SSIM \pm standard deviation across all projections o	
and mean SSIM \pm standard deviation across all projections o	
n and mean SSIM \pm standard deviation across all projections o	
ion and mean SSIM \pm standard deviation across all projections o	
ation and mean SSIM \pm standard deviation across all projections o	
viation and mean SSIM \pm standard deviation across all projections o	
eviation and mean SSIM \pm standard deviation across all projections o	
deviation and mean SSIM \pm standard deviation across all projections o	
d deviation and mean SSIM \pm standard deviation across all projections o	
ard deviation and mean SSIM \pm standard deviation across all projections o	
dard deviation and mean SSIM \pm standard deviation across all projections o	
andard deviation and mean SSIM \pm standard deviation across all projections o	
standard deviation and mean SSIM \pm standard deviation across all projections o	
: standard deviation and mean SSIM \pm standard deviation across all projections o	
\pm standard deviation and mean SSIM \pm standard deviation across all projections o	
$R \pm$ standard deviation and mean SSIM \pm standard deviation across all projections o	
$1\mathrm{R}\pm\mathrm{standard}$ deviation and mean SSIM $\pm\mathrm{standard}$ deviation across all projections o	
an R \pm standard deviation and mean SSIM \pm standard deviation across all projections o	
fean R \pm standard deviation and mean SSIM \pm standard deviation across all projections o	
Mean R \pm standard deviation and mean SSIM \pm standard deviation across all projections o	
. Mean R \pm standard deviation and mean SSIM \pm standard deviation across all projections o	
I. Mean R \pm standard deviation and mean SSIM \pm standard deviation across all projections o	
$_{ m JE}$ I. Mean R \pm standard deviation and mean SSIM \pm standard deviation across all projections o	
BLE I. Mean R \pm standard deviation and mean SSIM \pm standard deviation across all projections o	
TABLE I. Mean R \pm standard deviation and mean SSIM \pm standard deviation across all projections o	

LABLE L. IMEAL tures.	$\mathbf{R} \pm \mathbf{Standarc}$	l devlation and l	nean John II S	ianuaru ueviauo	JII across all pro	jecnons of the o	oronary artene	s III IIIe lest AC		un (a) une raw-p	ixei-value leatu	res and (b) the l	4 V IVI 16a-
		Ref-Poly		Ref-A-Table		ANN		Random tree		REPTree		Random forest	
Scenario	Noise level	$R \pm SD$	$\text{SSIM}\pm\text{SD}$	$\textbf{R} \pm \textbf{SD}$	$\text{SSIM}\pm\text{SD}$	$R\pm SD$	$\text{SSIM}\pm\text{SD}$	$R \pm SD$	$\text{SSIM}\pm\text{SD}$	$R\pm SD$	$SSIM \pm SD$	$\textbf{R} \pm \textbf{SD}$	SSIM \pm SD
(a)													
Non-motion	Noisy	0.33 ± 0.15	0.24 ± 0.13	0.37 ± 0.23	0.31 ± 0.15	0.48 ± 0.19	0.43 ± 0.17	0.30 ± 0.18	0.28 ± 0.18	0.47 ± 0.23	$\textbf{0.41}\pm\textbf{0.22}$	0.43 ± 0.20	0.38 ± 0.18
	Noiseless	0.67 ± 0.22	0.29 ± 0.25	0.70 ± 0.18	0.42 ± 0.16	0.83 ± 0.09	0.55 ± 0.23	0.70 ± 0.18	0.69 ± 0.18	0.82 ± 0.19	$\textbf{0.79}\pm\textbf{0.21}$	0.79 ± 0.22	0.73 ± 0.22
Motion	Noisy	0.28 ± 0.16	0.22 ± 0.13	0.30 ± 0.22	0.25 ± 0.09	0.39 ± 0.21	$\textbf{0.34}\pm\textbf{0.17}$	0.21 ± 0.18	0.20 ± 0.16	0.34 ± 0.23	0.30 ± 0.21	0.31 ± 0.21	0.27 ± 0.20
	Noiseless	0.66 ± 0.22	0.27 ± 0.25	0.68 ± 0.22	0.42 ± 0.12	$\textbf{0.75}\pm\textbf{0.16}$	0.46 ± 0.25	0.57 ± 0.19	0.55 ± 0.21	0.70 ± 0.21	$\textbf{0.68} \pm \textbf{0.23}$	0.65 ± 0.23	0.60 ± 0.23
(q)													
Non-motion	Noisy	0.33 ± 0.15	0.24 ± 0.13	0.37 ± 0.23	0.31 ± 0.15	0.84 ± 0.10	0.82 ± 0.10	0.68 ± 0.15	0.67 ± 0.15	0.75 ± 0.13	0.73 ± 0.14	0.87 ± 0.11	$\textbf{0.83}\pm\textbf{0.13}$
	Noiseless	0.67 ± 0.22	0.29 ± 0.25	0.70 ± 0.18	0.42 ± 0.16	$\textbf{0.91}\pm\textbf{0.05}$	0.61 ± 0.05	0.72 ± 0.12	0.72 ± 0.12	0.83 ± 0.11	0.82 ± 0.12	0.90 ± 0.11	$\textbf{0.89}\pm\textbf{0.12}$
Motion	Noisy	0.28 ± 0.16	0.22 ± 0.13	0.30 ± 0.22	0.25 ± 0.09	$\textbf{0.75}\pm\textbf{0.14}$	0.73 ± 0.15	0.55 ± 0.16	0.55 ± 0.16	0.64 ± 0.17	0.62 ± 0.17	0.75 ± 0.16	$\textbf{0.71}\pm\textbf{0.16}$
	Noiseless	0.66 ± 0.22	0.27 ± 0.25	0.68 ± 0.22	0.42 ± 0.12	$\textbf{0.83}\pm\textbf{0.12}$	0.54 ± 0.12	0.59 ± 0.16	0.59 ± 0.17	0.72 ± 0.16	0.71 ± 0.16	0.80 ± 0.15	$\textbf{0.75}\pm\textbf{0.16}$
The bold results	indicate superio	or performance.											

the performance of the machine learning algorithms tend to improve as hold-out rate increases. The shift of the curves between using the raw-pixel-value features and the MVM features also shows that it is beneficial to use neighboring information for material decomposition.

3.B. Experimental study

Figure 5 presents the material decomposition results for the two needles in the SAWBONES phantom scenario at the hold-out rate of 50%. The contrast of the images was enhanced by the same histogram stretching. The machine learning algorithms demonstrate better performance compared to the reference methods. Similar to the simulation study, the neighboring information is also beneficial to the performance, with improved contrast of the needles. The improvement is also reflected in the quantitative measurements. However, some residue of the other materials has appeared. Table II summarizes the performance using R and SSIM (average \pm standard deviation) across all quantitative measurements at the hold-out rate of 50%. Compared to the A-Table method, the performance of the ANN is improved by over 42% in the noiseless scenarios and by over 45% in the noisy scenario, while the performance of the Random Forest is improved by over 33% and 40%, respectively. The improvement is not as much as in the simulation study

because the experimental scenario has a lower noise level (as show in Fig. 2). Supporting Information Fig. S6 shows the sampling interval analysis and demonstrates a pattern of performance curves similar to those seen in the simulation study.

4. DISCUSSION

4.A. Model evaluation

In this work, only hold-out validation with continuous interleaved sampling was employed in the pipeline because the datasets of projections were generated across various angles with a fixed angular increment. A fixed interval sampling brings an effect of down sampling to the datasets, making subsets retain the properties of the original datasets depending on the sampling rate. The performance as a function of the hold-out rates was also investigated. Ignoring the computation load, the hold-out validation with continuous interleaved sampling strategy should be appropriate, because this would minimize the divergence of properties between the training set and the validation set.

R and SSIM were employed to quantify the performance of the models for quantitative evaluations instead of ordinary methods such as root mean squared error (RMSE). RMSE measures the average of the deviations between two images, treating all the pixels equally, which is somehow not in



FIG. 4. Decomposed material-specific projections of the coronary arteries in the test **XCAT** scenarios with cardiac and respiratory **motion** using (a) the raw-pixelvalue features and (b) the MVM features. The central projections with corresponding quantitative measurements are presented to demonstrate the performance.



Fig. 5. Decomposed material-specific projections of the two biopsy **needles** in the test **Torso** scenarios using (a) the raw-pixel-value features and (b) the MVM features. The central projections with the corresponding R and SSIM are presented.

accordance with human visual perception. For instance, RMSE cannot differentiate the divergence between two images where the first has a small number of pixels with large deviations and the second has a large number of pixels with small deviations. The quantitative evaluation in this work is therefore based on R and SSIM, which are popular choices for subjective quality assessment.

4.B. Feature extraction and machine learning algorithms

Two feature extractions have been demonstrated in this work. The raw-pixel-value feature extractor was implemented to extract the raw spectral measurements of multi-energy projections, which does not explicitly fit the polychromatic spectra in spectral x-ray imaging. Nevertheless, as shown in the results, the introduced machine learning algorithms with appropriate hyperparameters are able to learn successful models for material decomposition tasks, and have demonstrated superior performance compared to the reference methods. However, the performance deteriorates in the noisy scenarios because the training was based on pixel-wise spectral measurements. The pixels of noise may be treated as false positive instances, which reduces the accuracy of the models. The MVM feature extractor creates additional features of neighboring information to the spectral measurements. In this

Medical Physics, 46 (2), February 2019

work, we used multiple radial parameters to yield different levels of neighboring information for optimizing performance. The results show that it is feasible and beneficial for the material decomposition tasks. Especially, such information significantly improves the performance in the noisy scenarios. We speculate that this is due to the fact that the neighboring information of the target pixel has reduced the false positive probability, as well as provided highly relevant features for the tasks. Thus, adequate feature extraction approaches may prove valuable.

Machine learning algorithms have been employed to model the relationship between the multi-energy projections and the effective material-specific projections in the pipeline, largely conditioning the success of the material decomposition endeavor. As shown in the results, the performance of the respective algorithms is different. Since the sigmoid activation function was used to process the output of the neurons, the performance of the ANN appears more noisy than that of the decision tree algorithms. This is due to the fact that the models trained by the ANN may predict small negative values being displayed as residues of bone. Furthermore, this also affected the performance quantified by SSIM, especially when there was only a small amount of training data (representing low hold-out rate, as shown in Supporting Information Figs. S4–S6). Note that the neighboring information has significantly improved the performance of the ANN,

	Ref-poly		Ref-A-Table		ANN		Random tree		REPTree		Random forest	
Noise level	$R\pmSD$	$\text{SSIM}\pm\text{SD}$	$R \pm SD$	$\text{SSIM}\pm\text{SD}$	$\textbf{R} \pm \textbf{SD}$	$\text{SSIM}\pm\text{SD}$	$R\pm SD$	$SSIM \pm SD$	$R \pm SD$	$\text{SSIM} \pm \text{SD}$	$\textbf{R}\pm\textbf{SD}$	$\text{SSIM} \pm \text{SD}$
(a)												
Noisy	0.57 ± 0.09	0.50 ± 0.09	0.63 ± 0.07	0.55 ± 0.09	$\textbf{0.68}\pm\textbf{0.06}$	0.62 ± 0.09	0.51 ± 0.08	0.49 ± 0.08	0.64 ± 0.06	0.59 ± 0.07	$\textbf{0.66}\pm \textbf{0.06}$	$\textbf{0.62}\pm\textbf{0.08}$
JBF-denoised	0.65 ± 0.12	0.45 ± 0.09	0.67 ± 0.07	0.52 ± 0.12	$\textbf{0.78}\pm\textbf{0.07}$	$\textbf{0.65}\pm\textbf{0.12}$	0.58 ± 0.08	0.52 ± 0.08	0.73 ± 0.07	0.61 ± 0.08	0.75 ± 0.07	$\textbf{0.65}\pm\textbf{0.09}$
(p)												
Noisy	0.57 ± 0.09	0.50 ± 0.09	0.63 ± 0.07	0.55 ± 0.09	0.88 ± 0.05	0.84 ± 0.09	0.71 ± 0.09	0.69 ± 0.09	0.78 ± 0.08	0.76 ± 0.09	0.84 ± 0.06	0.82 ± 0.07
JBF-denoised	0.65 ± 0.12	0.45 ± 0.09	0.67 ± 0.07	0.52 ± 0.12	0.87 ± 0.05	0.82 ± 0.10	0.70 ± 0.09	0.66 ± 0.11	0.76 ± 0.09	0.72 ± 0.11	0.82 ± 0.07	0.77 ± 0.10

indicating good prospects for using deep learning approaches such as deep convolutional neural network (CNN). In order to complete this work as a comprehensive comparison of machine learning solutions for material decomposition tasks, we have also performed a pilot investigation on material decomposition using deep learning approaches. Figure 6 and Table III present the material decomposition results using two state-of-the-art deep learning-based solutions, DnCNNbased³² and ResNet-based.³³ Both of the two solutions use the three-channel projections as the inputs, and were implemented by PyTorch on a Titan Xp GPU.

The structure of the DnCNN-based network is shown in Fig. 7(a), and the number of the depth layers was 20. For the first layer, 64 filters of size $3 \times 3 \times 3$ were used to generate 64 feature maps with rectified linear units (ReLU) activations. For the 20-depth layers, 64 filters of $3 \times 3 \times 64$ were used, and batch normalization was added between convolution and ReLU. For the last layer, 1 filter of size $3 \times 3 \times 64$ was used to generate the output. The network training was performed by minimizing the MSE loss between the generated decomposed projection and the respective label. Adam gradient-based optimization with a momentum of 0.8 was used. The convolution kernel weights were initialized using random Gaussian distributions with a weight decay of 0.001 and a mini-batch size of 128. The learning rate and epochs were set to 0.0005 and 50, respectively. The image patch size was 40×40 .

The structure of the ResNet-based network is shown in Fig. 7(b), and we used nine residual blocks in this study. Adam with a momentum of 0.5 was used for minimizing the loss function. The generators were trained with a MSE loss function. The learning rate and epochs were set to 0.0002 and 100, respectively. The image patch size was 40×40 .

As show in Fig. 6 and Table III, the two deep learningbased solutions demonstrate comparable performance to ANN and Random Forest with MVM features, the performance is especially better in challenging projections. Counterintuitively, the simulation scenarios in this work are very challenging for segmentation tasks. The shape of vessels are similar to the shape of bones, and the overlap among materials is very challenging to segment. Although the performance of the two deep learning-based solutions using low hold-out rate are limited (Fig. 8) due to lacking sufficient training data, it is evident that while the performance of other methods seems to slowly taper off as the sampling rate increases, the deep learning methods still maintain a raising slope. Moreover, in the deep-learning methods, no specific choice of what goes in to the feature maps has been made, which could be considered as a major advantage as no restrictions are being imposed on the learning model with regards to what features should be learned. Further studies such as redesign concerning loss function, network architecture and sophisticated parameter tuning are required to develop a full-fledged deep learning method for material decomposition tasks.

As an alternative to ANN, decision trees are also appropriate for material decomposition tasks. Three decision tree algorithms have been demonstrated in this work. The Random Tree chose all features to construct decision tree models,



Motion and Noisy XCAT Scenario

Fig. 6. Decomposed material-specific projections of the Motion-Noisy XCAT scenario using two deep learning-based solutions.

TABLE III. Material decomposition results of two deep learning-based solutions in the motion-noisy XCAT scenario.

	DnCNN		ResNet		MVM + ANN		MVM + rando	m forest
	$R \pm SD$	$\rm SSIM\pmSD$	$R \pm SD$	$\rm SSIM \pm SD$	$R \pm SD$	$\rm SSIM \pm SD$	$R \pm SD$	$\rm SSIM \pm SD$
Noisy	0.76 ± 0.12	0.70 ± 0.15	0.77 ± 0.13	0.74 ± 0.14	0.75 ± 0.14	0.73 ± 0.15	0.75 ± 0.16	0.71 ± 0.16



FIG. 7. The structure of the two deep learning-based networks. [Color figure can be viewed at wileyonlinelibrary.com]



Fig. 8. Comparison study of different hold-out rates using DnCNN, ResNet, ANN + MVM and RandomForest + MVM. [Color figure can be viewed at wileyonlinelibrary.com]

which was susceptible to overfitting. As two methods of improving the generalization performance, the REPTree performs reduced-error by pruning the trees whereas the Random Forest combines the outputs of multiple Random Trees. According to the hold-out rate analysis in the results, the decision tree algorithms demonstrated superior performance on SSIM while using low hold-out rates, especially the Random Forest. Furthermore, there would be no negative value issue because they use tree construction to learn the models, resulting in better visual impression. Similar to the ANN, the performance of the Random Forest can be greatly improved by the neighboring information. This indicates that the Random Forest has better modeling capability and is more appropriate for further investigations compared to the other two decision trees.

As shown in the results, the appropriate machine learning algorithms demonstrated superior performance compared to the reference methods. Aside from the beneficial neighboring information, we speculate that the superior performance is due to the following reasons:

- (1) These machine learning algorithms can model the nonlinear and nonparametric models more accurately.
- (2) The per-pixel strategy employed in the feature extractions creates large amount of training data for learning the models.
- (3) The hold-out strategy with continuous interleaved sampling can minimize the divergences between the training set and the validation set, providing precise feedback to the parameter tunings to generate successful models.

4.C. Toward clinical practice

As shown in the Supporting Information Fig. S7, the performance does not vary a lot across different FOVs with the same imaging system parameters. This indicates the potential

TABLE IV. Computation time of each algorithms in the simulation study.

	С	omputation t	ime	
	L	earning ^a		Decom. ^a
Ref-Poly Ref-A-Table	0	0 h11 m34 s) h09 m22 s		4.122 s 0.324 s
	Raw-pixel-value	:	MVM	
	Learning ^a	Decom. ^b	Learning ^a	Decom. ^b
ANN	10 h50 m56 s	0.367 s	56 h57 m27 s	8.454 s
Random tree	00 h16 m13 s	0.301 s	00 h43 m18 s	1.482 s
REPTree	00 h12 m04 s	0.945 s	02 h03 m24 s	1.023 s
Random forest	01 h39 m10 s	3.654 s	03 h54 m01 s	12.654 s

^aLearning phase: the computation times of the learning phase are given for building a model from **all** pixels of each training set. (two noise levels, 36 902 400 pixels in total).

^bMaterial decomposition phase: the computation times of the material decomposition phase are given for decomposing **one** projection (average time across all projections) in each test set. (297 600 pixels for each projection). *h: hour *m: minute *s: second.

for clinical practice. In order to apply this pipeline to practice, the training needs to be general. The generalization performance of the models depend on the training data. In this work, we generated various imaging scenarios with different noise levels for investigation. However, if the models were trained by the training data from an individual scenario, the performance would deteriorate when applied to another individual scenario. For this reason, we combined the feature vectors from different scenarios to train generic models for each individual scenario. Since the parameter of image scenarios would vary in practice, sufficient amount of training data from various scenarios are necessary to train more generic models. However, the computation needs more time with the increasing number of features and amount of training data. This means expensive computation would be one of the bottlenecks of application toward clinical practice. Table IV presents the computation time of each algorithm using our workstation (Intel Xeon CPU E5-2650 v2 2.60 GHz, 64 GB DDR4 RAM), which is calculated based on single-thread implementation on CPU. As shown in the table, two powerful algorithms, the ANN and the Random Forest, are very timeconsuming in the learning phase, especially so when MVM features were included in the training. The deep learningbased solutions are also very computational expensively. Hardware acceleration can speed up the learning, but this requires high demands on hardware.

Material decomposition is not a simple segmentation problem and each objective material needs to be quantified. The proposed pipeline is applicable to provide quantitative information depending on its labels. The labels can be either material-specific line integrals (as demonstrated in this work) or objective quantifies (e.g., material thicknesses). However, another challenge in applying the pipeline to clinical practice is providing the labels. Since supervised learning was employed in the pipeline, creating labels becomes essential. The labels are not only used for learning models but also quantitative evaluations for model selection, representing the desired outputs of the pipeline. The ground truth can be used as labels directly in simulation studies. However, creating labels becomes more difficult in experimental studies. In this work, we demonstrated that the datasets can be labeled by experimental methods (as elaborated in Section 2.D.1). Beyond that, registration-based methods elaborated in the literature^{16,18} could also be considered. Nevertheless, creating labels would need further study, especially while applying the pipeline toward potential clinical applications. With multiple labels, the pipeline is able to perform multi-material decomposition as well. The decomposition results of the bone in the respective simulated scenarios are presented in Supporting Information Fig. S8 and Table I. In general, the results demonstrate the potential of applying the proposed pipeline to clinical practice. We will explore the feasibility of the pipeline toward clinical application in future studies.

5. CONCLUSION

In this work, we proposed a novel learning-based material decomposition pipeline. The material decomposition capability of the pipeline has been demonstrated in various scenarios. In general, the proposed pipeline is able to learn generic material decomposition models from input spectral measurements without explicitly modeling the imaging system. This was validated by quantitative evaluation in both the simulation study and the experimental study. Beyond that, the step of feature extraction in the pipeline has potential to integrate more informative features, such as neighboring information to facilitate material decomposition tasks. Deep learningbased solutions were also investigated in this study, and they demonstrated promising results. Compared to the reference methods, the performance of material decomposition was improved upon using appropriate features and machine learning algorithms. The results indicate that it is feasible and promising to perform material decomposition using machine learning methods, and the study will facilitate future efforts toward clinical applications.

ACKNOWLEDGMENT

The authors gratefully acknowledge the funding support from the Research Training Group 1773 Heterogeneous Image Systems by the German Research Foundation (DFG), the German Academic Exchange Service (DAAD), and the Siemens Healthineers Advanced Therapies. The authors acknowledge Charlie Zhou for his help of proof-reading the manuscript. Furthermore, we acknowledge the support of the NVIDIA Corporation with the donation of the Titan Xp GPU used for this research.

CONFLICT OF INTEREST

The concepts and information presented in this paper are based on research and are not commercially available. The authors have no conflicts to disclose. ^{a)}Authors to whom correspondence should be addressed. Electronic mail: yanye.lu@fau.de.

REFERENCES

- McCollough CH, Leng S, Yu L, Fletcher JG. Dual- and multi-energy CT: principles, technical approaches, and clinical applications. *Radiol*ogy. 2015;276:637–653.
- Alvarez RE, Macovski A. Energy-selective reconstructions in x-ray computerised tomography. *Phys Med Biol.* 1976;21:733–744.
- Brooks RA. A quantitative theory of the Hounsfield unit and its application to dual energy scanning. J Comput Assist Tomogr. 1977;1:487–493.
- Maaß C, Baer M, Kachelrieß M. Image-based dual energy CT using optimized precorrection functions: a practical new approach of material decomposition in image domain. *Med Phys.* 2009;36:3818–3829.
- Maaß C, Meyer E, Kachelrieß M. Exact dual energy material decomposition from inconsistent rays (MDIR). *Med Phys.* 2011;38:691–700.
- Taguchi K, Iwanczyk JS. Vision 20/20: single photon counting x-ray detectors in medical imaging. *Med Phys.* 2013;40:100901.
- Taguchi K, Zhang M, Frey EC, et al. Modeling the performance of a photon counting x-ray detector for CT: energy response and pulse pileup effects. *Med Phys.* 2011;38:1089–1102.
- Cammin J, Xu J, Barber WC, Iwanczyk JS, Hartsough NE, Taguchi K. A cascaded model of spectral distortions due to spectral response effects and pulse pileup effects in a photon-counting x-ray detector for CT. *Med Phys.* 2014;41:041905.
- Hassler U, Garnero L, Rizo P. X-ray dual-energy calibration based on estimated spectral properties of the experimental system. *IEEE Trans Nucl Sci.* 1998;45:1699–1712.
- Stenner P, Berkus T, Kachelriess M. Empirical dual energy calibration (EDEC) for cone-beam computed tomography. *Med Phys.* 2007;34:3630–3641.
- Alvarez RE. Estimator for photon counting energy selective x-ray imaging with multibin pulse height analysis. *Med Phys.* 2011;38:2324–2334.
- Wu D, Zhang L, Zhu X, Xu X, Wang S. A weighted polynomial based material decomposition method for spectral x-ray CT imaging. *Phys Med Biol.* 2016;61:3749–3783.
- Alvarez RE. Efficient, non-iterative estimator for imaging contrast agents with spectral x-ray detectors. *IEEE Trans Med Imaging*. 2016;35:1138–1146.
- Lee WJ, Kim DS, Kang SW, Yi WJ. Material depth reconstruction method of multi-energy x-ray images using neural network. In: 2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society; 2012:1514–1517.
- Zimmerman KC, Schmidt TG. Experimental comparison of empirical material decomposition methods for spectral CT. *Phys Med Biol*. 2015;60:3175–3191.
- Touch M, Clark DP, Barber W, Badea CT. A neural network-based method for spectral distortion correction in photon counting x-ray CT. *Phys Med Biol.* 2016;61:6132–6153.
- Lu Y, Geret J, Unberath M, et al. Projection-based material decomposition by machine learning using image-based features for computed tomography. In: Proceedings of the Fully 3D 2015; 2015:448–451.
- Lu Y, Berger M, Manhart M, et al. Bridge to real data: Empirical multiple material calibration for learning-based material decomposition. In: Proceedings of the 2016 IEEE International Symposium on Biomedical Imaging; 2016:457–460.
- Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH. The WEKA data mining software. *SIGKDD Explorat*. 2009;11:1–10.
- Kohonen T. An introduction to neural computing. *Neural Netw.* 1988;1:3–16.
- Duchi JC, Hazan E, Singer Y. Adaptive subgradient methods for online learning and stochastic optimization. J Mach Learn Res. 2010;12:257.
- Qian N. On the momentum term in gradient descent learning algorithms. *Neural Netw.* 1999;12:145–151.
- 23. Breiman L. Random forests. Mach Learn. 2001;45:5-32.
- 24. Quinlan J. Simplifying decision trees. Int J Man Mach Stud. 1987;27:221–234.
- 25. Quinlan J. Induction of decision trees. Mach Learn. 1986;1:81-106.
- Wang Z, Bovik A, Sheikh H, Simoncelli E. Image quality assessment: from error visibility to structural similarity. *IEEE Trans Image Process*. 2004;13:600–612.

- Maier A, Hofmann HG, Berger M, et al. CONRAD-A software framework for cone-beam imaging in radiology. *Med Phys.* 2013;40:111914.
- Maier A, Hofmann HG, Schwemmer C, Hornegger J, Keil A, Fahrig R. Fast simulation of x-ray projections of spline-based surfaces using an append buffer. *Phys Med Biol.* 2012;57:6193–6210.
- Rohkohl C, Lauritsch G, Keil A, Hornegger J. CAVAREV–an open platform for evaluating 3D and 4D cardiac vasculature reconstruction. *Phys Med Biol.* 2010;55:2905–2915.
- Hubbell JH, Seltzer SM. *Tables of X-ray mass attenuation coefficients* and mass energy-absorption coefficients. Technical Report (National Inst. of Standards and Technology-PL, Gaithersburg, MD (United States). Ionizing Radiation Div.) 1995.
- Lu Y, Manhart M, Taubmann O. Projection-based denoising method for photoncounting energy-resolving detectors. *Bildverarbeitung für die Medizin 2015*. Lübeck, Germany: Springer; 2015:137–142.
- Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising. *IEEE Trans Image Process*. 2017;26:3142–3155.
- He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*; 2016.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Fig. S1. Three noiseless simulated scenarios of **XCAT** short scan in the simulation study. The top row presents the first projections and the bottom row presents the central projections. Differences in FOV can be seen between the learning scenario and test scenarios. Motion phenomenon can be observed between the two test scenarios.

Fig. S2. Two experimental scenarios of **Torso** scan. The oneneedle scenario was used for learning and the two-needle scenario was used for test.

Fig. S3. Decomposed material-specific projections of the coronary arteries in the testing XCAT scenarios with (a) the

raw-pixel-value features and (b) the MVM features. The central projections with corresponding quantitative measurements are presented to demonstrate the performance.

Fig. S4. Quantitative measurements plotted against the holdout rate derived from the material decomposition results in the test **XCAT** scenario with cardiac and respiratory **motion** using (a) the raw-pixel-value features and (b) the MVM features. The y axis corresponds to either mean R or mean SSIM depends on the type of curves.

Fig. S5. Quantitative measurements plotted against the holdout rate derived from the material decomposition results in the testing scenario of **XCAT** using (a) the raw-pixel-value features and (b) the MVM features. The y axis corresponds to either mean R or mean SSIM depends on the type of curves.

Fig. S6.Quantitative measurements plotted against the holdout rate derived from the material decomposition results in the **Torso** scenarios using (a) the raw-pixel-value features and (b) the MVM features. The y axis corresponds to either mean R or mean SSIM depends on the type of curves.

Fig. S7. Quantitative measurements plotted against the FOV offsets from the FOV of the learning scenario using the MVM features in the noisy XCAT testing scenario.

Figs. 8. Decomposed material-specific projections of the **bone** in the testing **XCAT** scenario with motion using (a) the raw-pixel-value features and (b) the MVM features. The central projections with corresponding quantitative measurements are presented to demonstrate the performance.

Table SI. Mean $R \pm$ standard deviation and mean SSIM \pm standard deviation across all projections of the **bone** with (a) the raw-pixel-value features and (b) the MVM features.