

Unsupervised denoising for sparse multi-spectral computed tomography

Satu I. Inkinen, Mikael A. K. Brix, Miika T. Nieminen, Simon Arridge, Andreas Hauptmann,

Abstract—Multi-energy computed tomography (CT) with photon counting detectors (PCDs) enables spectral imaging as PCDs can assign the incoming photons to specific energy channels. However, PCDs with many spectral channels drastically increase the computational complexity of the CT reconstruction, and bespoke reconstruction algorithms need fine-tuning to varying noise statistics. Especially if many projections are taken, a large amount of data has to be collected and stored. Sparse view CT is one solution for data reduction. However, these issues are especially exacerbated when sparse imaging scenarios are encountered due to a significant reduction in photon counts. In this work, we investigate the suitability of learning-based improvements to the challenging task of obtaining high-quality reconstructions from sparse measurements for a 64-channel PCD-CT. In particular, to overcome missing reference data for the training procedure, we propose an unsupervised denoising and artefact removal approach by exploiting different filter functions in the reconstruction and an explicit coupling of spectral channels with the nuclear norm. Performance is assessed on both simulated synthetic data and the openly available experimental Multi-Spectral Imaging via Computed Tomography (MUSIC) dataset. We compared the quality of our unsupervised method to iterative total nuclear variation regularized reconstructions and a supervised denoiser trained with reference data. We show that improved reconstruction quality can be achieved with flexibility on noise statistics and effective suppression of streaking artefacts when using unsupervised denoising with spectral coupling.

Index Terms—Computed tomography, deep learning, unsupervised methods, image reconstruction, sparse angles, spectral imaging

I. INTRODUCTION

SPECTRAL computed tomography (CT) has recently emerged as a rapidly growing imaging modality as photon counting detector (PCD) technology, in particular, has made significant advances in past years [1], [2]. Spectral CT provides several advantages over conventional single energy CT such as enhanced material discrimination ability, quantitative imaging, and energy selective imaging [3], [4]. In the medical imaging domain, spectral CT with PCDs has been successfully applied, for example, in the characterization of gout [5] and contrast-enhanced imaging of the kidneys [6]. In the security

industry, spectral imaging has enabled automated detection and characterization of hazardous materials e.g. explosives and weapons [7], [8]. Furthermore, multi-bin PCDs having more than 64 energy channels have enabled spectroscopic CT imaging, which can be exploited in material classification tasks [8], [9]. For example, a material classification method was applied in estimating relative electron density and effective atomic number maps based on the corrected linear attenuation coefficient (LAC) data from spectral channel reconstructions [8], [9].

If the object is imaged with thousands of projections in multi-channel spectral CT, the amount of acquired data may become problematic. Especially, if the projection images are acquired in 3D. With sparse spectral CT, the number of projections is reduced. The sparse-view CT has been mainly studied for single-energy CT [10], [11], however, due to the increased data with the additional energy channel dimension in spectral CT, it is a useful method for data reduction. For example in airport security applications where rapid data acquisition and processing is needed, effective spectral sparse CT could be useful.

As only a limited number of photons arrive in the spectral channels, the acquired projection data is inherently noisy. Therefore, dedicated noise-robust spectral reconstruction methods have been developed for spectral PCD-CT [12]–[14]. These methods make use of iterative reconstruction algorithms and cover joint regularization methods [12], tensor-based nuclear norm regularization [13], [15], and dictionary learning [14] approaches. However, the above studies focus on PCDs with a maximum of eight energy channels, and as the number of energy channels increases, the computation time of the iterative methods will become a challenge – especially when the spectral projection data is acquired in 3-D. Therefore, alternative approaches have to be sought and developed to speed up the channel-wise reconstructions of spectral multi-bin CT. Prior studies utilizing multi-bin (over 64 energy channels) data have mainly used filtered back projection or iterative reconstruction methods such as ART or SIRT after reduction of the energy channels [8], [16], where the focus has mainly been on the material differentiation aspect. Kehl *et al.* [16] compared different unsupervised segmentation approaches for the openly available spectral CT dataset (MUSIC), also used in this study, collected from a 128-bin PCD utilizing iterative ART-TV reconstruction.

Recently, data-driven approaches have been developed for single energy CT reconstruction [17]–[22] to improve image quality with a considerable speed-up in reconstruction times compared to iterative approaches after the training is per-

This work was supported by Academy of Finland (project: 316899, 336796, 334817, 338408) and the CMIC-EPSRC platform grant (EP/M020533/1).

Satu I. Inkinen, Mikael A. K. Brix and Miika T. Nieminen are with the Research Unit of Medical Imaging, Physics and Technology, University of Oulu, Aapistie 5A, 90220, Oulu, Finland (e-mail: satu.inkinen@oulu.fi; mikael.juntunen@oulu.fi; miika.nieminen@oulu.fi).

Simon Arridge is with the Department of Computer Science, University College London, WC1E 6BT London, U.K. (e-mail: s.arridge@cs.ucl.ac.uk).

Andreas Hauptmann is with the Research Unit of Mathematical Sciences, University of Oulu, Pentti Kaiteran katu 1, 90570, Oulu, Finland and with the Department of Computer Science, University College London, WC1E 6BT London, U.K (e-mail: andreas.hauptmann@oulu.fi).

formed. For single energy CT reconstruction, several different learning approaches have been developed. The most common can be roughly divided into two categories: post-processing [17], [18], [21], [23] and learned iterative networks [19], [20], [22], [24]. Post-processing networks primarily focus on removing noise and artefacts from an initial reconstruction, obtained, for instance, by filtered back projection (FBP), and are especially time-efficient. Learned iterative approaches, on the other hand, are more time-consuming due to the repeated application of the projection operator. Nevertheless, regardless of the chosen network architecture, both approaches provide significant improvements in image quality over conventional iterative variational methods for single energy CT. For multi-spectral CT, most data-driven approaches address the material decomposition problem, by training a network to perform the decomposition either in data space prior to reconstruction [25]–[28] or afterwards [29]. In a recent study, Mustafa et al. [30] have investigated the use of a post-processing network for the MUSIC data trained in a supervised manner. Their post-processing network improved the image quality of the FBP images and achieved higher image quality when compared to the iterative ART-TV reconstruction [30].

In particular, the majority of learning-based reconstructions are trained supervised, which means that a relatively large set of ground-truth reconstructions is needed. This is especially problematic for applications where no densely sampled measurement is available, and hence one would need to resort to either synthetically generated simulated data, as in [30], to train the networks or use gold-standard reference reconstructions from the low-quality measurement data. Consequently, unsupervised approaches that do not need such reference data have gained increased attention recently [31]–[33]. Most notably, and with relevance to this study, are denoising approaches that can be trained with noisy reference data of either paired [31], [32] or unpaired images [33]. If paired images are used, both instantiations must have uncorrelated noise. In tomographic reconstruction, this has been realized as *Noise2Inverse* [31], where by angular binning non-overlapping subsets of the measured sinogram are used to produce reconstructions of the same target with uncorrelated noise. Nevertheless, this approach is expected to yield a reconstruction equivalent to the filtered backprojection from noise free data and hence is limitedly applicable to the sparse angle scenario, as it retains streaking artefacts as well. To overcome this problem, we adopt a different strategy tailored to our application of sparse multi-spectral CT. First, instead of taking subsets of the sinogram, we produce two reconstructions from the full measurement with different filter functions in the filtered backprojection (FBP), which are used as training pairs. Second, we add a total nuclear variation loss to the training to couple spectral channels and suppress remaining streaking artefacts.

Let us shortly discuss other unsupervised methods that could be adapted with some modifications. Generative adversarial networks still need (unpaired) reference data to train the discriminator [34], [35], which are not readily available in our case. SURE (Stein’s unbiased risk estimate) based approaches make use of the noise statistics in the image [36]–[38]. Here, we have strongly varying noise between spectral channels and

even within the same image (air and metallic) and hence necessitates careful adjustment of the method. Lastly, the Deep Image Prior (DIP) [39] could be used with a suitable regularizer to suppress the streaking artefacts [40], but is known to be computationally demanding and would need an involved pretraining procedure for 3D spectral CT [41].

This paper is structured as follows. The proposed unsupervised denoiser for flexible sparse spectral multi-bin PCD-CT will be discussed in Section II-B. We will also compare the proposed method to iterative total nuclear variation (TNV) reconstructions [15] and a supervised denoiser using a reference reconstruction, following [18], [30]. We will apply the learned methods on synthetic 2D multi-channel (64 energy channels) sparse PCD projection data, described in Section III. Finally, we utilize the openly available MUSIC experimental dataset [16] to show how the developed method performs for real experimental data. We present results, assess the computation time and reconstruction quality in Section IV. We then discuss the proposed method in Section V and present final conclusions in Section VI.

II. UNSUPERVISED TRAINING FOR SPARSE MULTISPECTRAL TOMOGRAPHIC DATA

There is recent rising interest in unsupervised learned reconstruction methods, as in many applications there is only limited ground-truth information available, which either needs to be simulated or computed from high accuracy scans. Obtaining reference scans can be especially difficult for multi-spectral CT, where strong noise occurs in very low and high energy channels due to lower photon counts, and emphasized streaking artefacts appear for strongly attenuating objects in the low energy channels. Additionally, depending on the photon energies we observe different noise statistics, which makes fine-tuning of reference reconstructions a tedious and difficult task.

Consequently, we investigate in the following the suitability of a class of learned denoisers that can operate without ground-truth data and can learn from the same dataset by exploiting uncorrelated noise. As various studies have recently reported, a denoising network can be trained with pairs of images with different noise instantiations [31], [32], [42]. The observation by [32], introduced as *Noise2Noise*, is that when given two realisations of a noisy image $x^{\delta_1} = x + \delta_1$ and $x^{\delta_2} = x + \delta_2$ with independent draws of $\delta_i \sim \pi_{\text{noise}}$ one can reformulate the supervised training setting using the two noise corrupted images. For zero-mean noise $\mathbb{E}[\delta y] = 0$ the expected prediction error

$$\theta^* = \arg \min_{\theta} \mathbb{E}_{x, \delta_1, \delta_2} [\|\Lambda_{\theta}(x + \delta_1) - (x + \delta_2)\|_2^2] \quad (1)$$

for a class of regression networks Λ_{θ} with parameters θ is then minimized by the same (regression) network Λ_{θ^*} as in the supervised case.

Based on the same observation, the *Noise2Inverse* [31], [42] method has been introduced for single energy CT setting using two, or more, reconstructions with uncorrelated noise. Such reconstructions can be easily produced by angular binning of the projection data into non-overlapping subsets and by

reconstruction with FBP. This has been shown to work well for dense angular sampling, when no streaking artefacts are present. Nevertheless, for sparse imaging scenarios, streaking artefacts will remain. Additionally, varying noise statistics in the multi-spectral case can be problematic for the approach. Thus, we need to take a different angle on the unsupervised training of a denoiser that also works in the case of sparse multi-spectral data. Here, the trained denoiser also needs to act as an artefact removal network. In the following, we will extend the unsupervised denoising approach based on the Noise2Inverse by producing two reconstructions from the same set of projections using different filters in the backprojection, instead of angular binning.

In the following we will denote the ground-truth tomographic image by x and the forward model describing the projection geometry is given by A , the precise imaging environment will be described in Section III-A. Then the measurement data y is obtained as

$$Ax + \delta = y, \quad (2)$$

where δ denotes noise in the measurement process. A reconstruction x_{FBP} can be obtained from y by filtered backprojection (FBP), where visible features in the obtained reconstructions will depend on the type of filtering employed. That means, we can decide on whether to emphasize low or high frequencies, influencing the image characteristics, as will be discussed below in Section II-B.

The aim in the following is to train a network Λ_θ , with parameters θ , that can improve the reconstruction quality of the obtained filtered backprojection, such that

$$x_{\text{rec}} = \Lambda_\theta(x_{\text{FBP}}). \quad (3)$$

The primary question is how training of the denoising network can be performed, especially considering one has no or little ground-truth data available.

A. Supervised training

Let us first discuss the supervised approach when a known ground-truth or high-quality reconstruction is available [17], [18], as it is the most established way to train the network in (3) and it will also serve as the basis for our unsupervised setting. Here, a reference high-quality reconstruction should be ideally given from dense angular sampling under high dose. If no such data is available, one would need to create a reference reconstruction that is considered the gold-standard for the given scenario. The network is then trained to reproduce a similar result but considerably faster than the reference method. For the application of spectral CT, such a reference reconstruction could be given by an iterative reconstruction with a total nuclear variation penalty term [15], which introduces a gradient-coupling effect between energy channels by promoting rank-sparsity of the coupled derivatives.

If given such training pairs of noisy reconstruction and a reference for training $\{x_{\text{FBP}}^{(n,b)}, x_{\text{ref}}^{(n,b)}\}$, where $n \in [1, N]$ denotes a sample from the training pairs and b the spectral channels. One can then train the supervised denoising network

by minimizing a suitable loss function. In this study we will use the squared ℓ^2 -norm

$$L(x_{\text{FBP}}^{(n,b)}, x_{\text{ref}}^{(n,b)}) = \|\Lambda_\theta(x_{\text{FBP}}^{(n,b)}) - x_{\text{ref}}^{(n,b)}\|_2^2. \quad (4)$$

The primary advantage of the supervised denoiser setup lies in its simplicity and its computational speed in the reconstruction phase, as it only requires a single reconstruction by filtered backprojection for each spectral channel and a computationally efficient evaluation of the network, after the training has been performed. On the downside, the performance of the network is governed by the quality of the reference reconstructions in the training data. As such, unsupervised approaches are desirable that can extract relevant information from only the noisy data and are not limited by the quality of the available reference. In our experiments, the FBP reconstruction for the supervised training had Hann filtering with a frequency scaling of 0.6.

B. Unsupervised training

We now aim to train a denoising and artefact removal network with only noisy pairs following the Noise2Inverse approach. Such a denoising network can be efficiently trained when two noisy instances of the same tomographic image are available. Here, we propose a strategy to produce two reconstructions from the same sinogram. To achieve this we will utilize different filter functions in the filtered backprojection.

We first realize, that given a noisy reconstruction $\tilde{x} = x + \delta$, we ideally would like to have access to the noise component δ . Then for training one could separate it into two components δ_1 and δ_2 , such as low and high frequency components by separation in Fourier space using high-pass and low-pass filtering. Similarly, a separation of frequencies could be achieved by employing different filters in the filtered backprojection. Specifically, a reconstruction can be expressed by

$$x_{\text{FBP}} = A^*(h * y) = A^*(h * (Ax + \delta)) \quad (5)$$

where the obtained data y is filtered by a convolutional filter before backprojected with the corresponding adjoint operator A^* , also called the backprojection, for the given measurement geometry (including measurement noise). In learning methods for low-count data h is often chosen to be a low-pass filter, such that primarily high-frequency noise is removed from the data before backprojection is performed. Our aim is to utilize two different filters to produce two reconstructions, where one has an emphasis on low-frequency features and the other leaves high-frequency features intact.

Unfortunately, we can not separate the data into high and low frequencies as this would result in a loss of structural similarity of image features in the reconstruction. But, rather we can obtain one smoothed (almost) noise and streaking artefact free reconstruction from the low-pass filter and one reconstruction including the high frequency noise and sharp image features. The resulting filters used in this study are shown in Fig. 1 with their normalized representation in frequency space. For this study we utilize a Ram-Lak filter for the

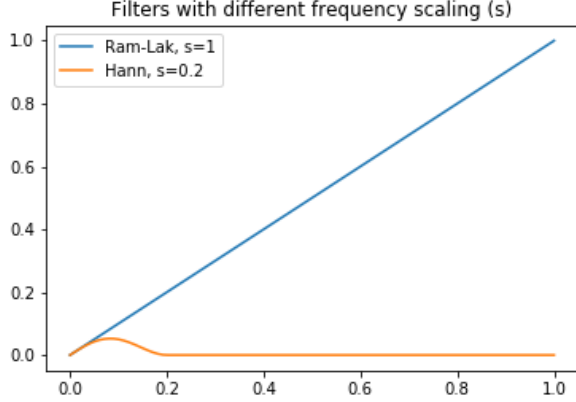


Fig. 1: Used filter functions in frequency space to create the two training pairs for the Low2High method.

high frequency features, given by the identity with frequency scaling $s = 1$,

$$h_{\text{Hi}}(k) = k, \text{ for } k \in [0, 1],$$

which, in fact, means we do not cut-off any noise components before we perform the filtered backprojection. For the low-pass reconstruction, we use a Hann filter with a frequency scaling s , that is

$$h_{\text{Lo}}(k) = k \left(\cos \left(\frac{k\pi}{2s} \right) \right)^2, \text{ for } k \in [0, s]$$

and $h_{\text{Lo}}(k) = 0$ otherwise. In this study, frequency scaling of $s = 0.2$ was applied for all experiments. We note that also other choices of frequency scaling s and the filter functions can be chosen, if needed. For the unsupervised training task we then obtain two reconstructions x_{Lo} and x_{Hi} that serve as training pairs. We can then employ the same loss function in Eq. (4) with the corresponding training pairs $\{x_{\text{Lo}}^{(n,b)}, x_{\text{Hi}}^{(n,b)}\}$. We note, that training over multiple samples or spectral channels is necessary for the varying noise realizations. We will call our approach *Low2High*. An illustrative result of this approach is shown in Fig. 2. As it can be observed for dense angular sampling the network nicely sharpens the low-pass filtered reconstruction without reproducing the noise. Whereas in the sparse case, clear streaking artefacts are reintroduced, which also result in a residual noise pattern in the image. Even though the low-pass filtered version suppresses streaking artefacts, these reappear partly during the training procedure of the denoiser. This can be explained by the prediction error in Eq. (1) and zero-mean condition on the noise, which is clearly not satisfied for streaking artefacts. Thus, in the following we will address this issue with an additional penalty term for the unsupervised training task.

1) *Spectral channel coupling with total nuclear variation:* In the case of sparse spectral computed tomography we can exploit that edges between spectral channels will be aligned. Thus, to avoid that the Low2High training procedure reintroduces streaking artefacts, while keeping edges of the target, we add a total nuclear variation penalty to the training procedure.

The use of a total nuclear variation (TNV) penalty is a highly successful technique to couple the gradients between spectral channels by requiring rank sparsity of the Jacobian [15]. Similar penalty terms based on the coupling of edges have been also successfully used in multi-modality imaging [43], [44]. Thus, in the following we will employ a TNV penalty in our training task to improve the results of the unsupervised training procedure.

Given the same paired training data $\{x_{\text{Lo}}^{(n,b)}, x_{\text{Hi}}^{(n,b)}\}$ as for the unsupervised training task above, we now add a total nuclear variation penalty to the loss function in Eq. (4), that is

$$L(x_{\text{Lo}}^{(n,b)}, x_{\text{Hi}}^{(n,b)}) = \|\Lambda_{\theta}(x_{\text{Lo}}^{(n,b)}) - x_{\text{Hi}}^{(n,b)}\|_2^2 + \lambda \|J(\Lambda_{\theta}(x_{\text{Lo}}^{(n,b)}))\|_{1,*}, \quad (6)$$

where λ is a weighting term for the TNV penalty. The TNV penalty can be efficiently computed as the ℓ^1 -norm of the singular values of the Jacobian J of the output from the network. Most importantly, this can be easily added to the training procedure as an additional loss function, where automatic differentiation can be performed to update the parameters θ .

Finally, we note that the low-pass filter herein has been chosen as a starting point such that streaking artefacts are suppressed in the input to the network, the employed TNV penalty in the training procedure then successfully prevents these to reappear. Nevertheless, we can see in Fig. 2 that for lower values of the weighting parameter λ we still see some residual artefacts in the background. For higher values we get a clean image, but edges between different regions start to be smoothed out. For the rest of this study, we will chose a value of $\lambda = 0.005$ to balance this behaviour, for an emphasis on retaining edges of the target.

III. SYNTHETIC AND EXPERIMENTAL DATA

Two datasets were adopted for this study: a measured, openly available, Multi-Spectral Imaging via Computed Tomography (MUSIC) dataset and synthetic images dataset using the TomoPhantom python software package [16], [45].

A. Open source MUSIC dataset

Detailed description of the openly available MUSIC dataset can be found from Kehl et al. [16] and the dataset can be downloaded from¹. Here, we briefly summarize the MUSIC dataset. The dataset was designed for security scanning assessment, and it consists of seven sample tubes containing different items from fluids, fruits, threat, and non-threat item categories (Table I). Samples were imaged with a 2-D table-top spectral CT device. The device had a MultiX ME-100 V2 cadmium telluride (CdTe) X-ray detector, which has 128 energy bins and an X-ray microfocus source with peak kilovoltage set to 160 kVp. The PCD consists of two modules of 128 pixels yielding a total of 256 pixels with 0.77 mm pixel size. The scan geometry was set such that the source-to-detector and source-to-object distances were 115.55 mm and 57.50 mm, respectively. Thirty-seven projections were collected during acquisition [16].

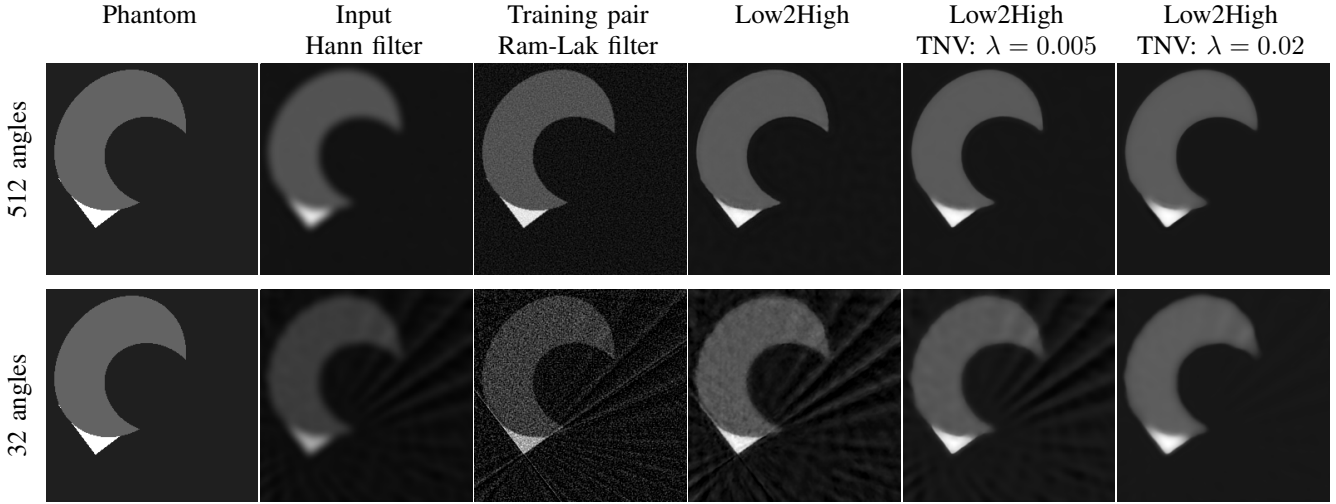


Fig. 2: Illustration of the Low2High training strategy. Training pairs of low frequency by a Hann filter ($s = 0.2$) and high frequency with a Ram-Lak filter ($s = 1$) are shown in the 2nd and 3rd column, respectively. Result by training only with the two pairs as Low2High are shown in the 4th column. The 5th and 6th column show denoising results with added TNV penalty and different weighting factors.

After the projection data was downloaded, it was processed as follows. The number of spectral channels was reduced from 128 energy bins to 64 channels from 55 keV to 124 keV, and projection data from each sample was manually cropped such that the empty tube ends containing no material were excluded.

For the application of spectral computerized tomography, we compute a computationally expensive iterative joint (spectrally coupled) reconstruction using total nuclear variation regularization (IR TNV) which introduces a gradient-coupling effect for energy channels as it encourages rank-sparsity. IR TNV is used as the reference ground truth (x_{ref}) for sparse MUSIC dataset. The reconstruction problem is then given as by

$$x_{\text{ref}} = \arg \min_x \frac{1}{2} \|Ax - y\|_2^2 + \alpha \| (Jx) \|_{1,*}, \quad (7)$$

¹<http://easi-cil.compute.dtu.dk/index.php/datasets/music/>

where y is the logarithm transformed flat-field corrected, i.e., linearized projection data (sinogram) and A denotes the projection matrix, Jx is the Jacobian, and $\| (Jx) \|_{1,*}$ is the total nuclear variation penalty, the parameter $\alpha > 0$ controls the regularization strength. Reconstructions were performed in 2D slices such that four adjacent energy slices were fed into the algorithm yielding 16 batches for all 64 energy channels.

Pre-processing and reconstructions were conducted using python and the operator discretization library (ODL) library [19]. ODL library's Douglas-Rachford primal-dual solver was applied for this problem with step sizes $\tau = 0.01$ for indicator box function and $\sigma_1 = 0.01$ for data fidelity and $\sigma_2 = 0.5$ for TNV. The regularization parameter α was fixed to 100 for each energy channel and the solver was run for 50 iterations using FBP as initial solution.

After computing the reference reconstructions the training, validation, and test set splitting was performed for the dataset

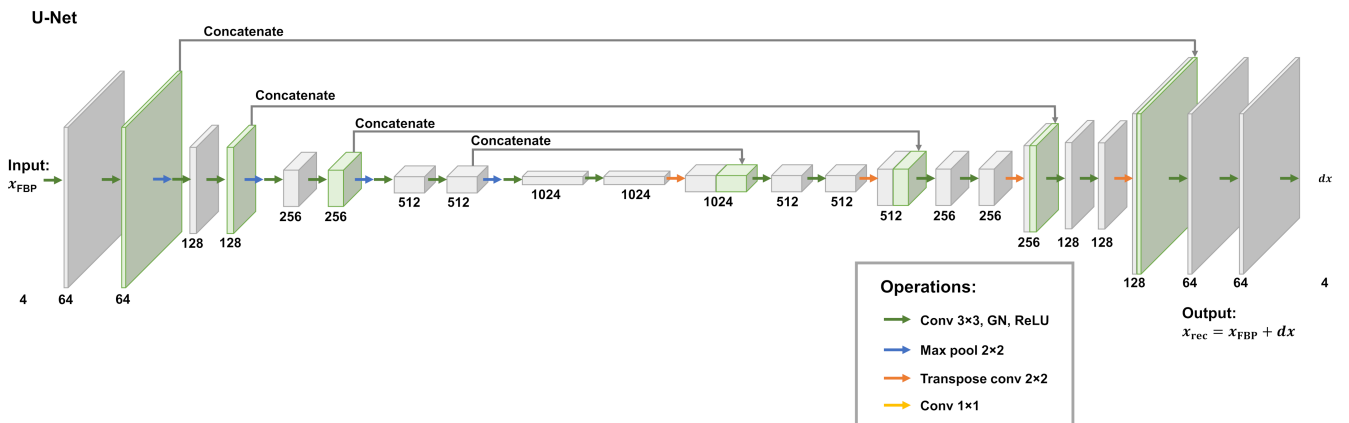


Fig. 3: Diagram of the residual UNet architecture used as post-processing network. The inputs reconstructions are given by a filtered backprojection (FBP) and the network computes a residual update as additive correction term to the initial reconstruction. We used group normalization (GN) here in the convolutional layers.

(Table I), where samples tubes were not split between sets. We note that each sample consists of 64 channels to make up the whole dataset size.

B. Synthetic multi-energy dataset

A simulated synthetic dataset was created with the TomoPhantom python package [45]. The simulation geometry was kept similar to the MUSIC data acquisition. The 2D synthetic dataset consisted of random sized, positioned and rotated ellipsoidal or rectangular objects. The number of structures varied from 0 to 5 per slice and the objects were enforced to be contained in the circular reconstruction domain. If two or more ellipses overlapped, the most recently generated ellipse was kept in the overlapping area. Finally, the objects extending outside the field-of-view of a 256-pixel detector, were cropped from the image.

A simple effective linear attenuation (LAC) model was applied to generate spectral information

$$\text{LAC} = Ae^{-BE}, \quad (8)$$

where $A \sim U(0, 0.03)$, $B \sim U(0, 10)$, E are the 64 energy channels from 55 keV to 124 keV. The ranges for uniform distributions were based on the MUSIC dataset as the maximum effective linear attenuation of 0.03 was estimated from aluminium bar region present in the reconstructed dataset.

After the spectral information was added to the synthetic dataset, we simulated 37 angular projections using ODL library with ASTRA backend extension [46], [47] with the MUSIC geometry. The number of projections was matched to the MUSIC dataset. 2% random Gaussian noise was added to the simulated projections. For training, validation and testing 1200, 100 and 100 samples were realized, respectively.

C. Network implementation

For our implementation we used an adaption of the popular UNet architecture [48] with a residual update similarly to [18], as illustrated in Fig. 3. Implementation differed from [18] such that group normalization (number of groups = 1) was used instead of batch normalization. In particular, we trained a 2-D UNet which operates on 1 and 4 adjacent energy channel input images for supervised and unsupervised learning, respectively. The energy channels were randomly selected from the spectral channels of a slice during training. For synthetic dataset the input and reference data was normalized prior feeding to network but for MUSIC dataset the input and reference was not normalized.

TABLE I: Summary of MUSIC dataset splits; Number of slices (N) and samples in the sets.

Set	N	Sample name
Train	929	Fluids, Fruits, Sample_23012018, Sample_31102016
Validation	382	NonThreat, Threat
Test	278	Sample_24012018

For supervised and unsupervised training of the UNet we applied approaches explained in the II-A and II-B, respectively. We note, that $x_{\text{ref}}^{(n,b)}$ here is either the ground-truth synthetic dataset or the reference reconstruction (IR TNV) for the MUSIC dataset obtained by (7). We note, that in the latter, the $x_{\text{ref}}^{(n,b)}$ might include a bias due to the chosen reconstruction algorithm to obtain a gold-standard reference. Here the randomization over spectral channels will be essential to counteract possible overfitting issues to the reference method as will be discussed later.

Networks were implemented under the PyTorch deep learning framework (v.1.2.0). For all networks and both datasets the training was performed with batch size of 1, using a default Adam optimizer for 50 epochs, with learning rate of 10^{-4} . A cosine annealing scheduler with default parameters was applied after each epoch to reduce the initial learning rate. The supervised UNet denoiser model was trained on training set and the unsupervised Low2High method was trained using test set only.

The neural networks were trained on a CSC IT center for science supercomputer with NVIDIA Volta V100 GPU (24 GBit memory) and 10 CPU threads. Training of UNet and UNet with TNV penalization network for the MUSIC dataset took 34 minutes and 2 hours and 4 minutes, respectively. The TNV had to be computed on CPU, as at the time of this study no GPU support for computing the singular value decomposition was available in PyTorch. After the neural network models were trained, the reconstruction were fast to compute compared to IR TNV (Table II).

TABLE II: Computational time of ten 64 spectral channel slice images on MUSIC dataset.

Method	Computation time (HH:MM:SS)
IR TNV	0:32:41
UNet	0:01:12
Low2High	0:01:09

D. Image quality assessment

Structural similarity index (SSIM) and peak signal-to-noise ratio (PSNR) were determined from the reconstructed test set images. In addition, channel-wise image quality was evaluated. For the synthetic dataset the ground truth images were available. For MUSIC dataset iterative total nuclear variation (IR TNV) was chosen as the reference reconstruction. To qualitatively assess the preservation of the effective linear attenuation coefficients (LACs), we determined LACs of different materials from example slices for MUSIC dataset. In addition, one representative slice image was chosen from the MUSIC test dataset containing both low (tube) and high (aluminium) contrast regions (Fig. 4). From those regions, contrast-to-noise ratio (CNR), and signal-to-noise ratio (SNR) were computed.

Finally, maximum intensity projection (MIP) images were computed from a sum of reconstructed energy channel images for the test set of the MUSIC dataset.

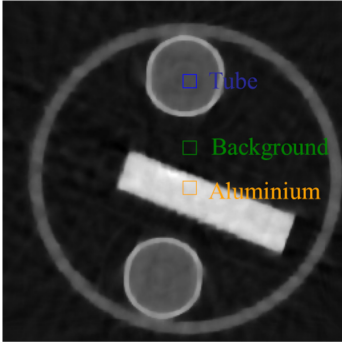


Fig. 4: Region of interest locations chosen for CNR and SNR computations of tube (low contrast) and aluminium (high contrast) regions. The illustration slice image presented here is the IR TNV average over all energy channels.

IV. RESULTS

A. Image quality assessment

1) *Synthetic spectral dataset*: The supervised UNet denoising model had the best SSIM when compared with FBP and unsupervised Low2High method (Table III). The Low2High method however, was able to suppress streaking artefacts of FBP reconstructions and clearly surpassed the FBP reconstruction in image quality (Fig. 5). Interestingly, in the PSNR assessment both supervised and unsupervised methods had similar performance (Table III and Fig. 6).

In our experiments, we found that model input normalization for synthetic dataset improved the overall image quality of different energy channels and the technical image quality was preserved for different reconstruction methods energy channel-wise (Fig. 6).

TABLE III: Structural similarity index (SSIM), and peak-signal-to-noise (PSNR) for different reconstruction methods for Spectral synthetic dataset (mean \pm SD).

Method	SSIM	PSNR (dB)
UNet (supervised)	0.916 \pm 0.088	28.6 \pm 5.7
Low2High (unsupervised)	0.717 \pm 0.081	29.5 \pm 2.9
FBP	0.067 \pm 0.041	13.2 \pm 1.7

2) *Spectral MUSIC dataset*: After training, the learning-based methods produced qualitatively similar reconstruction quality as the reference IR TNV method as shown in (Fig. 7). Interestingly, at the noisier high energy channels, the learning based methods were able to effectively suppress noise more than the reference IR TNV method when evaluated qualitatively (Fig. 7), even for the supervised approach that was trained on the reference IR TNV reconstructions. This can be interpreted as a regularizing effect of training with all spectral channels and effectively reduces the bias introduced by the reference reconstructions. The unsupervised reconstructions Low2High showed excellent noise suppression exceeding the supervised method in image quality. However, the FBP related streaking artefacts were not perfectly suppressed in the low energy channels of the unsupervised Low2High method (Fig. 7).

TABLE IV: Structural similarity index (SSIM), and peak-signal-to-noise (PSNR) for different reconstruction methods for MUSIC dataset (mean \pm SD). The ground truth reference image was the IR TNV.

Method	SSIM	PSNR (dB)
UNet (supervised)	0.643 \pm 0.048	30.2 \pm 1.5
Low2High (unsupervised)	0.454 \pm 0.080	27.6 \pm 1.8
FBP	0.315 \pm 0.044	25.2 \pm 1.4

When evaluated quantitatively, the supervised UNet denoiser had the most similar image quality performance to IR TNV (Table IV). However, the unsupervised Low2High had only slightly inferior performance in PSNR. The supervised UNet denoiser preserved the image quality similar to IR TNV over the spectral channels (Fig. 8). We note, that these values are limitedly useful as the reference reconstructions are not perfect, especially for the high noise channels. In particular, the decline in SSIM for the unsupervised Low2High approach is due to a better noise suppression compared to the supervised training. The performance of Low2High and supervised UNet were similar when the CNR and SNR were assessed, see Fig. 4 for the target and Table V for the results. Interestingly, both methods outperformed the reference IR TNV method, which again suggests a regularizing effect of training over all spectral channels. The Low2High method had a slightly higher CNR and SNR in low contrast region and UNet on the other hand had higher CNR and SNR in high contrast region.

The effective LAC values were also preserved with the measured MUSIC dataset over the spectral channels for all reconstruction methods (Fig. 9) and most notable the Low2High method produced the smoothest value, i.e., least noise present in the LAC values. Finally, MIP images illustrated that each reconstruction method had streaking artefacts at the lower end of the aluminium bar (Fig. 10). Both learning based methods had less streaking artefacts present in MIP images when compared with FBP.

V. DISCUSSION

To address the computational complexity of spectral multi-bin PCD-CT, we studied the applicability of learning-based reconstruction methods, i.e., denoising UNet in supervised and unsupervised learning domains, for fast sparse multi-bin spectral CT reconstruction. We were able to show with both simulated synthetic and measured data that the learned

TABLE V: Qualitative image measures presented as mean values over energy channels on one representative slice image from MUSIC dataset. Shown are: contrast-to-noise ratio (CNR), and signal-to-noise ratio (SNR).

Method	CNR _T	SNR _T	CNR _{Al}	SNR _{Al}
Low2High	31.3	32.2	84.0	84.9
FBP	5.0	5.2	14.0	14.2
UNet	30.7	31.2	91.2	91.7
IR TNV	20.6	20.8	61.3	61.5

T: tube (low contrast target), Al: Aluminium (high contrast target)

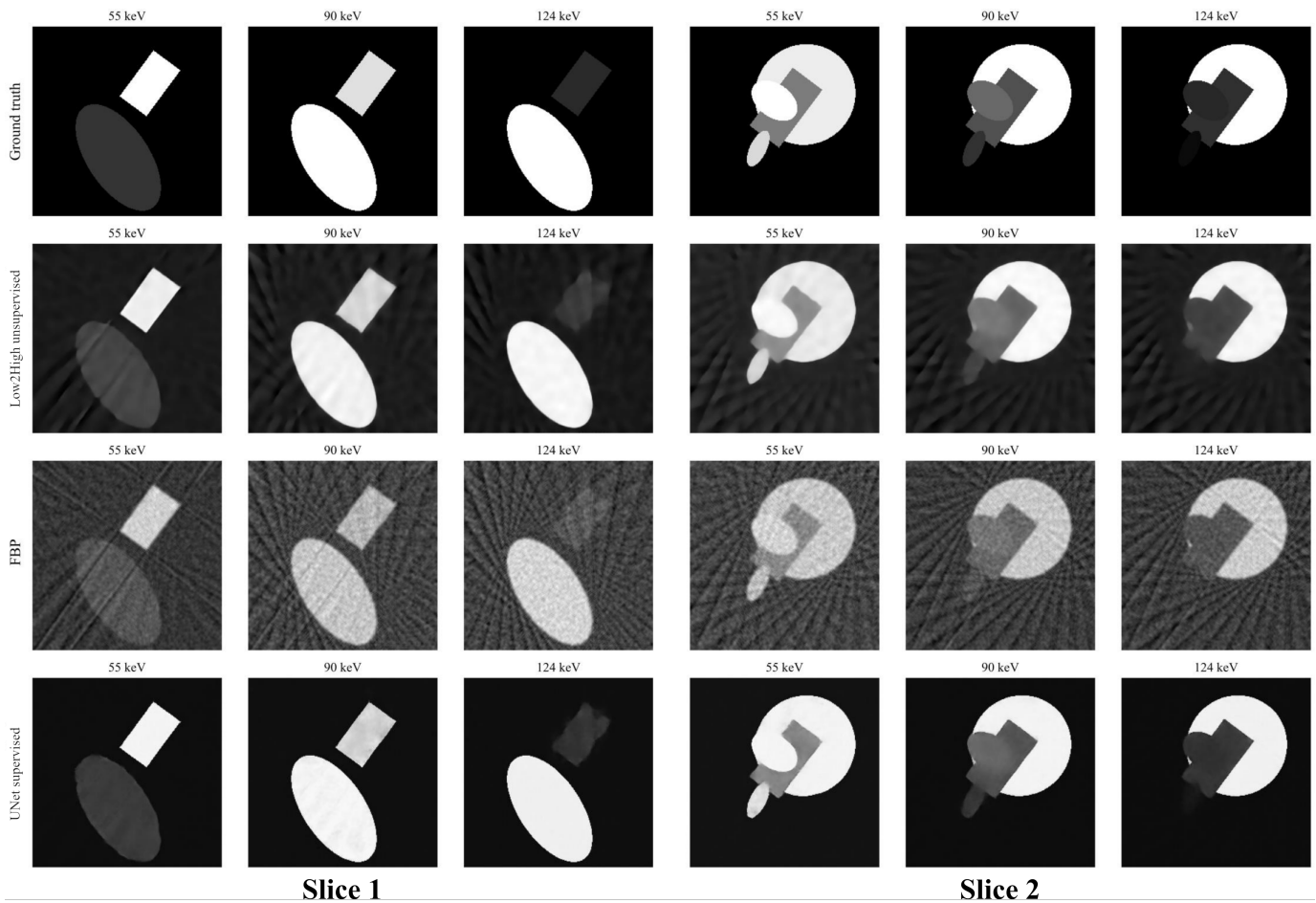


Fig. 5: Channel-wise reconstructions for sparse synthetic dataset of two randomly selected slices from test set. The unsupervised Low2High and supervised UNet methods were able to suppress the streaking artefacts. In the unsupervised approaches the streaking artefacts were not so effectively suppressed at low energies, however in the noisy high energy channel 124 keV results are improved when compared to FBP.

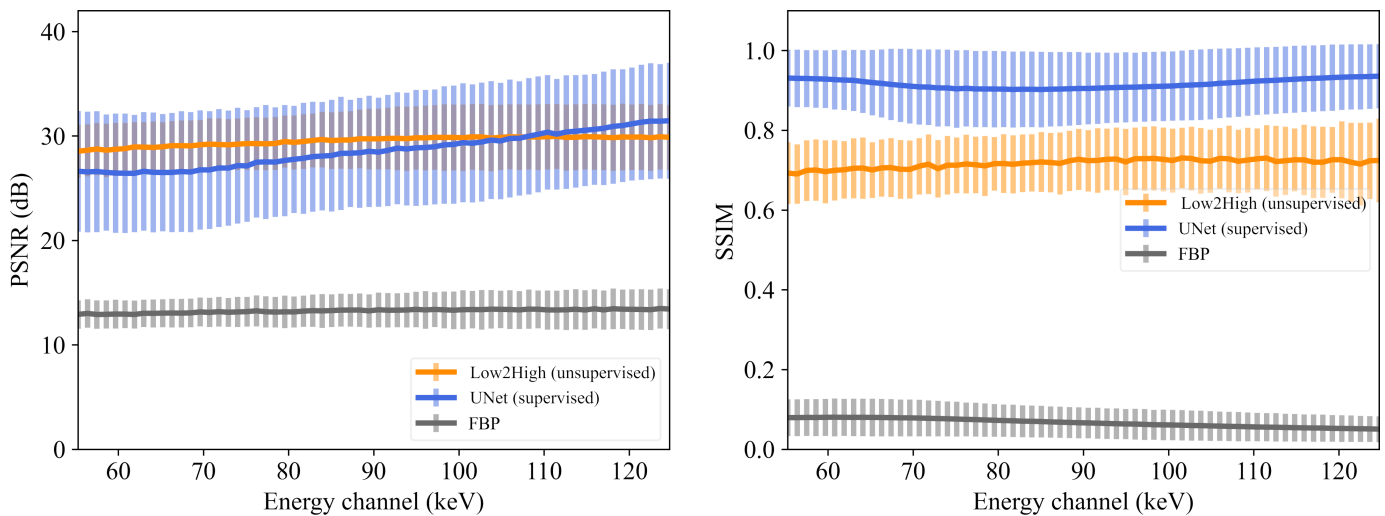


Fig. 6: Channel-wise PSNR and SSIM (mean \pm SD) for spectral synthetic dataset. The highest PSNR and SSIM was observed with UNet.

reconstruction methods preserve the image quality in the spectral channels while keeping the computation time feasible. Previous multi-bin spectral PCD-CT studies have overcome

computational challenges by reducing the number of channels or re-binning channels prior to applying iterative reconstruction

tion [8], [9]. Dimension reduction methods have been proposed as well such as principal component analysis or non-negative matrix factorization [49]. However, in spectroscopic X-ray CT applications, it is beneficial that energy channels are preserved as the multi-bin information may enable a material separation to materials with similar attenuation properties and improve the accuracy of material decomposition depending on the decomposition method. Furthermore, improved the material decomposition accuracy for high k-edge contrast agents gadolinium and iodine has been reported with increasing the number of energy channels [50].

In this study, we chose previously introduced UNet [18] as our backbone network architecture. The UNet denoises the input FBP reconstruction. UNet was robust to the varying noise levels of the different energy channels. Also, the application of iterative TNV as a reference reconstruction to replace a ground truth label did not deteriorate the learning of the networks for the measured MUSIC dataset even though IR TNV had varying noise characteristics in energy channels. In fact, by training over all spectral channels the learned reconstruction produced qualitatively better reconstructions than the reference method.

The primary limitation within this study is given by the quality of the training data, as this effectively limits the performance of the learning-based reconstruction methods. We note, that the quality of the measured MUSIC dataset was limited, since the projection data were acquired from 37 projection angles yielding a sparse-view problem, and data from only seven sample tubes were available. However, this resulted in 929 slice images with each 64 spectral channels for the training set for the supervised approach, which should be a feasible amount for the training of the model. The sparseness and noisiness of the MUSIC dataset led to the selection of iterative reconstruction with total nuclear variation as reference reconstruction to replace a ground truth label. For supervised training, one would preferably need separate measurements from densely sampled projection angle increments with higher exposure (i.e. less noise) to generate high-quality labels. Interestingly, the supervised method showed that when trained on a reference reconstruction it was capable to overcome a possible bias spectral bias by training over all channels simultaneously.

Given the limitations of the supervised approach, we have proposed an unsupervised training regime that is designed to work specifically for sparse spectral data. This is achieved by

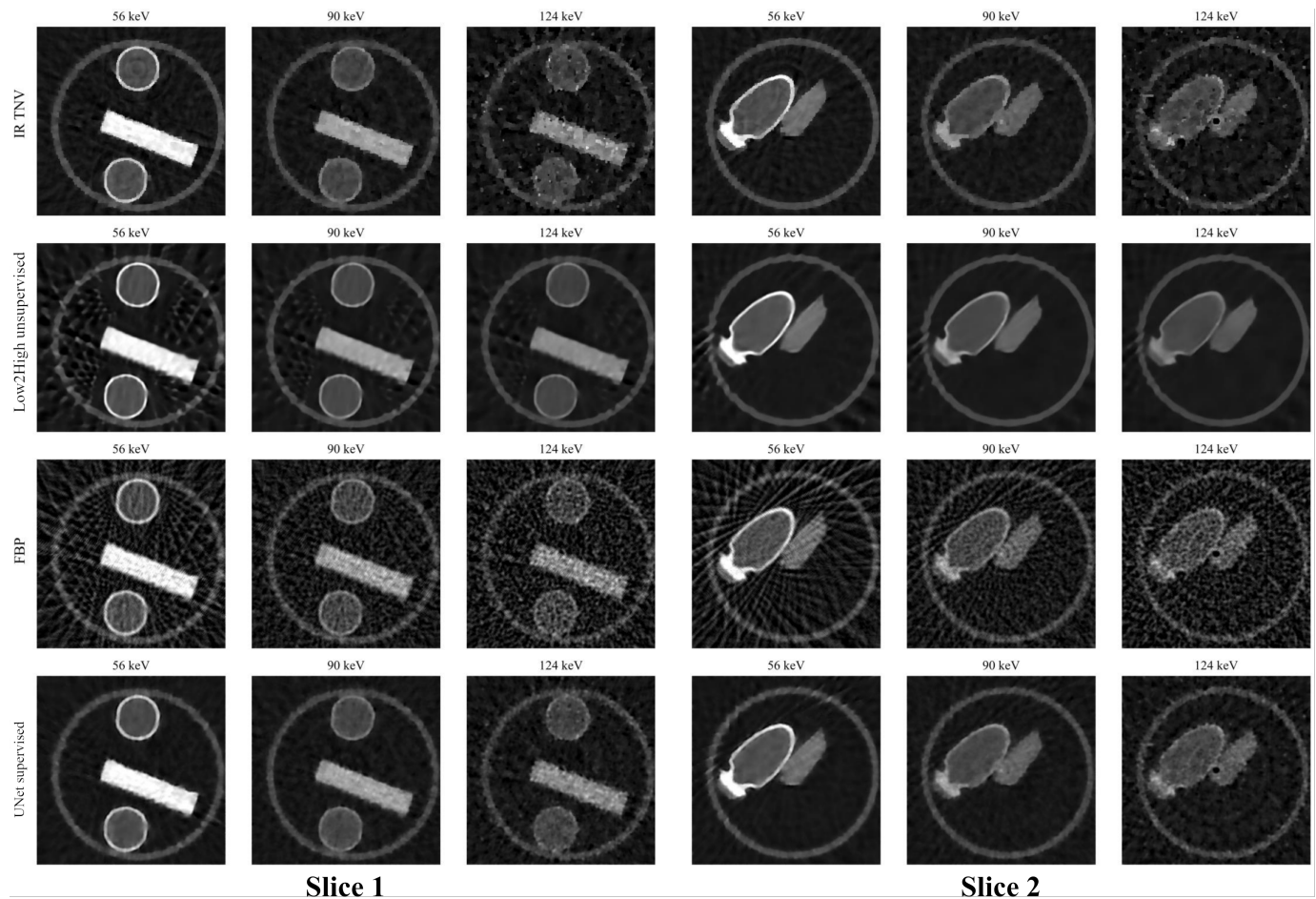


Fig. 7: Channel-wise reconstructions for MUSIC dataset of two slices. The unsupervised Low2High method for test set was able to suppress streaking artefacts most effectively. IR TNV, supervised and unsupervised methods were able to suppress the streaking artefacts arising from aluminium bar low energy channel 56 keV as shown in Slice 1. Windowing is set to $[-0.002, 0.03]$.

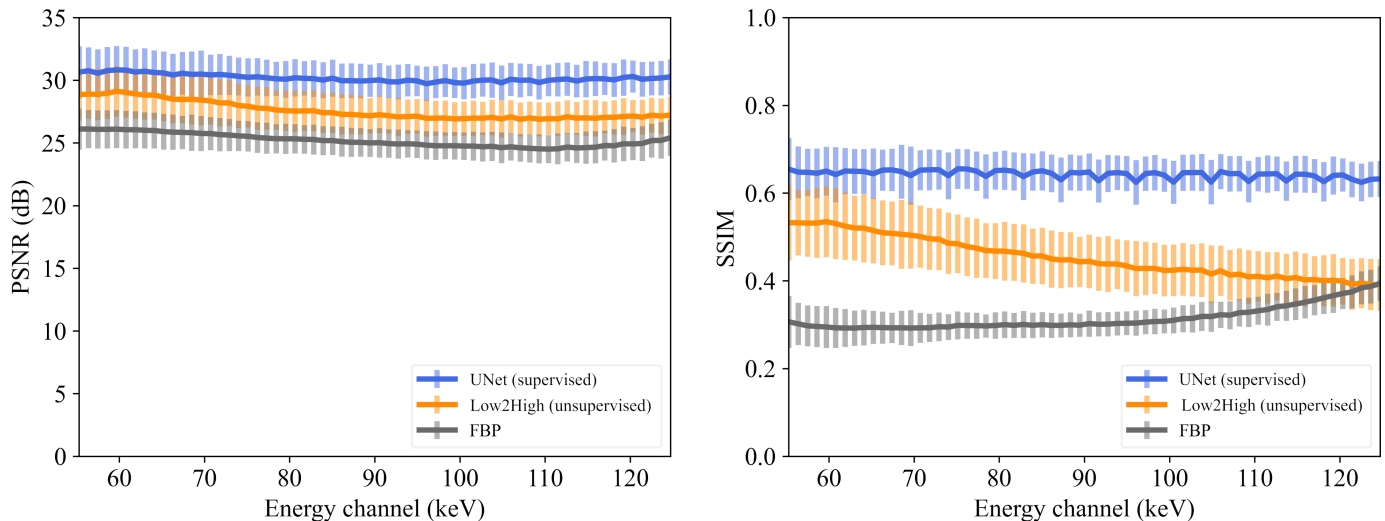


Fig. 8: Channel-wise PSNR and SSIM (mean \pm SD) for spectral MUSIC dataset. The highest PSNR and SSIM was observed with UNet.

creating training pairs from two differently filtered back-projections, Hann and Ram-Lak. We combined the training pairs with a TNV penalty to suppress streaking artefacts. While this Low2High approach produces excellent reconstructions for the noisy high energy channels with a successful suppression of streaking artefacts the performance for low energy channels was not as satisfactory, where stronger residual streaking artefacts were visible. Nevertheless, in comparison to the supervised method the proposed unsupervised Low2High training produced visually and quantitatively (by smoother LAC values) better reconstructions than the comparison methods for

the experimental MUSIC data. In the simulated case, streaking artefacts were still more dominant in comparison.

Future studies should focus on expanding the learned spectral reconstruction methods to the three-dimensional spatial domain, which constitutes still a major computational burden due to large memory requirements and computation times limiting applications where fast reconstruction is needed e.g. in the security domain. In addition, a study addressing the further applicability of the reconstructed spectral slice images from the learned reconstructions e.g., in the estimation of electron density and effective atomic numbers as shown in

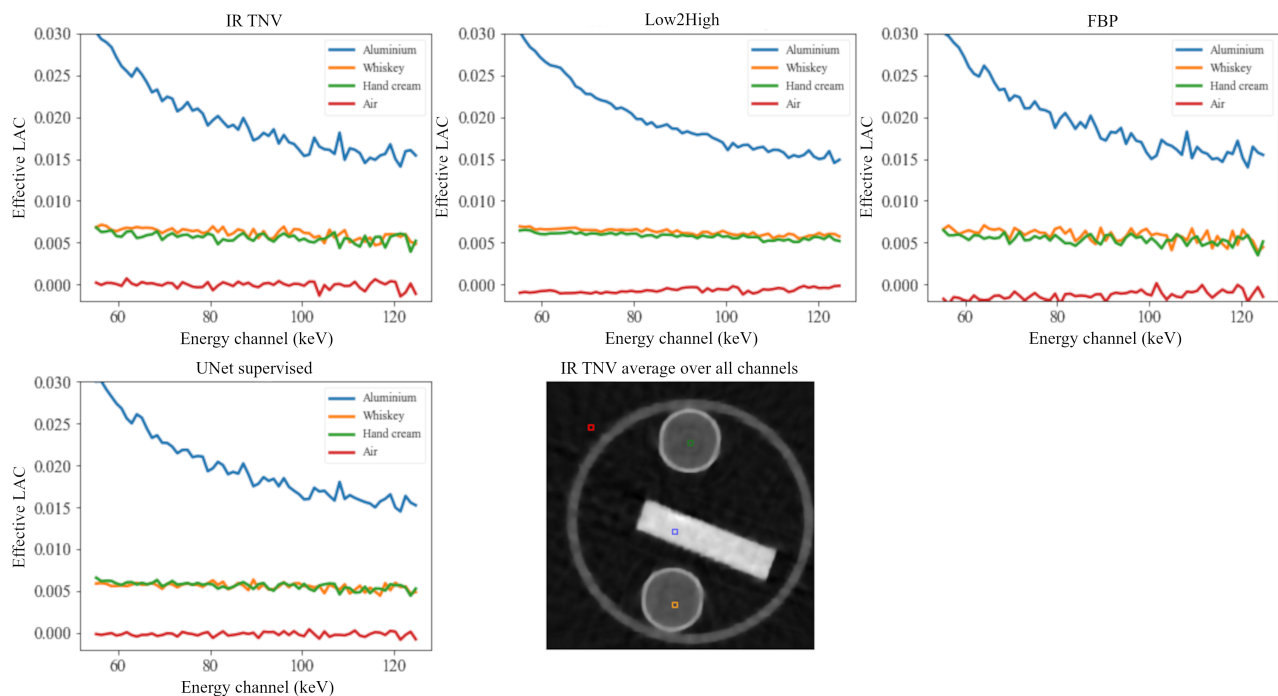


Fig. 9: Effective linear attenuation coefficients for different materials over energy channels for MUSIC dataset. Both, the leaning based methods and FBP were able to preserve the effective LAC values similarly as the reference IR TNV.

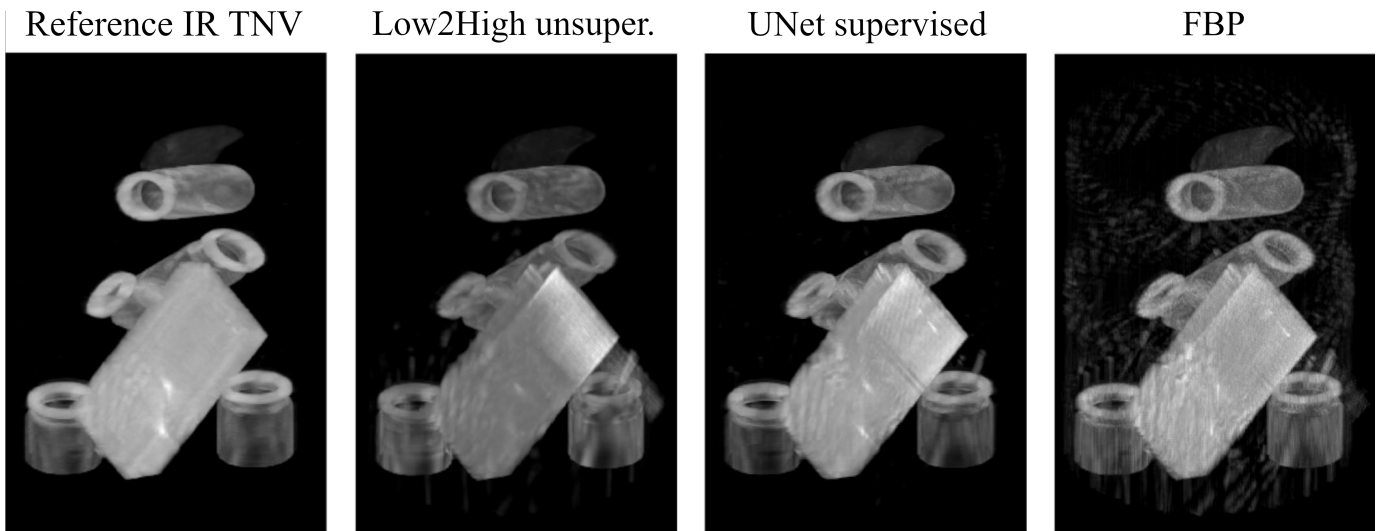


Fig. 10: Maximum intensity projection (MIP) images rendered from the test tube reconstructions (Sample 24012018) of MUSIC dataset. Striking artefacts were observed for all MIP images at the lower end of the sample tube where there were aluminium bar and two test tubes.

[8] warrants for investigation. Finally, a combination of the supervised and unsupervised method should be investigated to be able training on a designated training set, instead of the test set, and allow for rapid application of the learned reconstruction method.

VI. CONCLUSION

We have presented an unsupervised reconstruction approach with total nuclear norm for sparse-view spectral multi-bin PCD-CT in comparison to a fully supervised variant. Both unsupervised and supervised UNet networks preserved the image quality and effective LAC values over the spectral channels in reconstructions. To conclude, the computation time of sparse angle multi-bin spectral CT reconstruction can be successfully reduced with learned reconstructions in a supervised manner while producing comparable image quality with even better noise suppression than a reference iterative reconstruction. If trained unsupervised additional regularization by a spectral coupling is needed. Reconstruction times are governed by the training procedure, but excellent results can be obtained for experimental measurement data, especially for noisy high energy channels.

Finally, we believe that the employed TNV penalty in the training procedure will also prove useful for other supervised approaches, such as DIP [39], [40].

ACKNOWLEDGMENT

The authors would like to thank Allard Hendriksen for helpful discussions on the Noise2Inverse. The authors wish to acknowledge CSC – IT Center for Science, Finland, for computational resources.

REFERENCES

- [1] S. Leng, M. Bruesewitz, S. Tao, K. Rajendran, A. F. Halaweish, N. G. Campeau, J. G. Fletcher, and C. H. McCollough, "Photon-counting Detector CT: System Design and Clinical Applications of an Emerging Technology," *RadioGraphics*, vol. 39, no. 3, pp. 729–743, may 2019.
- [2] J. Dudak, "High-resolution X-ray imaging applications of hybrid-pixel photon counting detectors Timepix," *Radiation Measurements*, vol. 137, no. February 2019, p. 106409, sep 2020.
- [3] O. Lee, S. Kappler, C. Polster, and K. Taguchi, "Estimation of Basis Line-Integrals in a Spectral Distortion-Modeled Photon Counting Detector Using Low-Rank Approximation-Based X-Ray Transmittance Modeling: K-Edge Imaging Application," *IEEE Transactions on Medical Imaging*, vol. 36, no. 11, pp. 2389–2403, nov 2017.
- [4] M. A. K. Juntunen, S. I. Inkinen, J. H. Ketola, A. Kotiaho, M. Kauppinen, A. Winkler, and M. T. Nieminen, "Framework for Photon Counting Quantitative Material Decomposition," *IEEE Transactions on Medical Imaging*, vol. 39, no. 1, pp. 35–47, jan 2020.
- [5] F. A. Huber, S. Gkoumas, T. Thüring, F. Becce, and R. Guggenberger, "Detection and Characterization of Monosodium Urate and Calcium Hydroxyapatite Crystals Using Spectral Photon-Counting Radiography: A Proof-of-Concept Study," *European Journal of Radiology*, vol. 129, no. November 2019, p. 109080, aug 2020.
- [6] A. Pourmorteza, R. Symons, V. Sandfort, M. Mallek, M. K. Fuld, G. Henderson, E. C. Jones, A. A. Malayeri, L. R. Folio, and D. A. Bluemke, "Abdominal Imaging with Contrast-enhanced Photon-counting CT: First Human Experience," *Radiology*, vol. 279, no. 1, pp. 239–45, apr 2016.
- [7] A. Mouton and T. P. Breckon, "Materials-based 3D segmentation of unknown objects from dual-energy computed tomography imagery in baggage security screening," *Pattern Recognition*, vol. 48, no. 6, pp. 1961–1978, jun 2015.
- [8] D. Jumanazarov, J. Koo, M. Busi, H. F. Poulsen, U. L. Olsen, and M. Iovea, "System-independent material classification through X-ray attenuation decomposition from spectral X-ray CT," *NDT & E International*, vol. 116, no. June, p. 102336, dec 2020.
- [9] M. Busi, K. A. Mohan, A. A. Dooraghi, K. M. Champley, H. E. Martz, and U. L. Olsen, "Method for system-independent material characterization from spectral X-ray CT," *NDT & E International*, vol. 107, no. April, p. 102136, oct 2019.
- [10] S. J. Kisner, E. Haneda, C. A. Bouman, S. Skatter, M. Mourinny, and S. Bedford, "Model-based CT reconstruction from sparse views," *Proc. 2nd Intl. Mtg. on Image Formation in X-ray CT*, vol. 60, pp. 444–7, 2012.
- [11] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, "A Sparse-View CT Reconstruction Method Based on Combination of DenseNet and Deconvolution," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1407–1417, jun 2018.
- [12] J. Toivanen, A. Meaney, S. Siltanen, and V. Kolehmainen, "Joint reconstruction in low dose multi-energy CT," *Inverse Problems & Imaging*, vol. 14, no. 4, pp. 607–629, 2020.
- [13] O. Semerci, Ning Hao, M. E. Kilmer, and E. L. Miller, "Tensor-Based Formulation and Nuclear Norm Regularization for Multienergy Com-

- puted Tomography," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1678–1693, apr 2014.
- [14] Y. Zhang, X. Mou, G. Wang, and H. Yu, "Tensor-Based Dictionary Learning for Spectral CT Reconstruction," *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 142–154, 2017.
- [15] D. S. Rigie and P. J. La Rivière, "Joint reconstruction of multi-channel, spectral ct data via constrained total nuclear variation minimization," *Physics in Medicine & Biology*, vol. 60, no. 5, p. 1741, 2015.
- [16] C. Kehl, W. Mustafa, J. Kehres, A. B. Dahl, and U. L. Olsen, "Multi-Spectral Imaging via Computed Tomography (MUSIC) - Comparing Unsupervised Spectral Segmentations for Material Differentiation," oct 2018. [Online]. Available: <http://arxiv.org/abs/1810.11823>
- [17] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Medical Physics*, vol. 44, no. 10, 2017.
- [18] K. H. Jin, M. T. Mccann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Transactions on Image processing*, vol. 26, no. 9, pp. 4509–4522, 2017.
- [19] J. Adler and O. Öktem, "Solving ill-posed inverse problems using iterative deep neural networks," no. 1, pp. 1–24, apr 2017.
- [20] H. Chen, Y. Zhang, Y. Chen, J. Zhang, W. Zhang, H. Sun, Y. Lv, P. Liao, J. Zhou, and G. Wang, "LEARN: Learned Experts' Assessment-based Reconstruction Network for Sparse-data CT," *IEEE Transactions on Medical Imaging*, pp. 1–1, jul 2018.
- [21] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-D Convolutional Encoder-Decoder Network for Low-Dose CT via Transfer Learning From a 2-D Trained Network," *IEEE Transactions on Medical Imaging*, vol. 37, no. 6, pp. 1522–1534, 2018.
- [22] J. Adler and O. Öktem, "Learned primal-dual reconstruction," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1322–1332, 2018.
- [23] D. M. Pelt and J. A. Sethian, "A mixed-scale dense convolutional neural network for image analysis," *Proceedings of the National Academy of Sciences*, vol. 115, no. 2, pp. 254–259, 2018.
- [24] A. Hauptmann, J. Adler, S. R. Arridge, and O. Oktem, "Multi-scale learned iterative reconstruction," *IEEE Transactions on Computational Imaging*, 2020.
- [25] K. C. Zimmerman and T. G. Schmidt, "Experimental comparison of empirical material decomposition methods for spectral ct," *Physics in Medicine & Biology*, vol. 60, no. 8, p. 3175, 2015.
- [26] M. Touch, D. P. Clark, W. Barber, and C. T. Badea, "A neural network-based method for spectral distortion correction in photon counting x-ray ct," *Physics in Medicine & Biology*, vol. 61, no. 16, p. 6132, 2016.
- [27] J. Abascal, N. Ducros, V. Pronina, S. Bussod, A. Hauptmann, S. Arridge, P. Douek, and F. Peyrin, "Material decomposition problem in spectral ct: A transfer deep learning approach," in *2020 IEEE 17th International Symposium on Biomedical Imaging Workshops (ISBI Workshops)*, 2020, pp. 1–4.
- [28] J. F. Abascal, N. Ducros, V. Pronina, S. Rit, P.-A. Rodesch, T. Broussaud, S. Bussod, P. C. Douek, A. Hauptmann, S. Arridge *et al.*, "Material decomposition in spectral ct using deep learning: A sim2real transfer approach," *IEEE Access*, vol. 9, pp. 25 632–25 647, 2021.
- [29] D. P. Clark, M. Holbrook, and C. T. Badea, "Multi-energy ct decomposition using convolutional neural networks," in *Medical Imaging 2018: Physics of Medical Imaging*, vol. 10573. International Society for Optics and Photonics, 2018, p. 105731O.
- [30] W. Mustafa, C. Kehl, U. L. Olsen, S. K. S. Gregersen, D. Malmgren-Hansen, J. Kehres, and A. B. Dahl, "Sparse-view spectral ct reconstruction using deep learning," *arXiv preprint arXiv:2011.14842*, 2020.
- [31] A. A. Hendriksen, D. M. Pelt, and K. J. Batenburg, "Noise2inverse: Self-supervised deep convolutional denoising for tomography," *IEEE Transactions on Computational Imaging*, vol. 6, pp. 1320–1335, 2020.
- [32] J. Lehtinen, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, "Noise2Noise: Learning Image Restoration without Clean Data," *35th International Conference on Machine Learning, ICML 2018*, vol. 7, no. 3, pp. 4620–4631, mar 2018. [Online]. Available: <http://arxiv.org/abs/1803.04189>
- [33] B. Kim, H. Shim, and J. Baek, "Weakly-supervised progressive denoising with unpaired CT images," *Medical Image Analysis*, vol. 71, p. 102065, jul 2021. [Online]. Available: <https://doi.org/10.1016/j.media.2021.102065https://linkinghub.elsevier.com/retrieve/pii/S1361841521001110>
- [34] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.
- [35] Y. Ma, B. Wei, P. Feng, P. He, X. Guo, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with a hybrid loss function for noise learning," *IEEE Access*, vol. 8, pp. 67 519–67 529, 2020.
- [36] C. A. Metzler, A. Mousavi, R. Heckel, and R. G. Baraniuk, "Unsupervised learning with stein's unbiased risk estimator," *arXiv preprint arXiv:1805.10531*, 2018.
- [37] H. V. Nguyen, M. O. Ulfarsson, and J. R. Sveinsson, "Hyperspectral image denoising using sure-based unsupervised convolutional neural networks," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 59, no. 4, pp. 3369–3382, 2020.
- [38] K. Kim, S. Soltanayev, and S. Y. Chun, "Unsupervised training of denoisers for low-dose ct reconstruction without full-dose ground truth," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 6, p. 1112–1125, 2020.
- [39] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "Deep image prior," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 9446–9454.
- [40] D. O. Bague, J. Leuschner, and M. Schmidt, "Computed tomography reconstruction using deep image prior and learned reconstruction methods," *Inverse Problems*, vol. 36, no. 9, p. 094004, 2020.
- [41] R. Barbano, J. Leuschner, M. Schmidt, A. Denker, A. Hauptmann, P. Maaß, and B. Jin, "An educated warm start for deep image prior-based micro ct reconstruction," *arXiv preprint arXiv:2111.11926*, 2021.
- [42] A. A. Hendriksen, M. Bührer, L. Leone, M. Merlini, N. Vignano, D. M. Pelt, F. Marone, M. di Michiel, and K. J. Batenburg, "Deep denoising for multi-dimensional synchrotron x-ray tomography without high-quality reference data," *Scientific reports*, vol. 11, no. 1, pp. 1–13, 2021.
- [43] M. J. Ehrhardt, K. Thielemans, L. Pizarro, D. Atkinson, S. Ourselin, B. F. Hutton, and S. R. Arridge, "Joint reconstruction of pet-mri by exploiting structural similarity," *Inverse Problems*, vol. 31, no. 1, p. 015001, 2014.
- [44] F. Knoll, M. Holler, T. Koesters, R. Otazo, K. Bredies, and D. K. Sodickson, "Joint mr-pet reconstruction using a multi-channel image regularizer," *IEEE transactions on medical imaging*, vol. 36, no. 1, pp. 1–16, 2016.
- [45] D. Kazantsev, V. Pickalov, S. Nagella, E. Pasca, and P. J. Withers, "TomoPhantom, a software package to generate 2D–4D analytical phantoms for CT image reconstruction algorithm benchmarks," *SoftwareX*, vol. 7, pp. 150–155, jan 2018. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S2352711018300335>
- [46] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabrovolski, J. De Beenhouwer, K. Joost Batenburg, and J. Sijbers, "Fast and flexible X-ray tomography using the ASTRA toolbox," *Optics Express*, vol. 24, no. 22, p. 25129, oct 2016.
- [47] W. van Aarle, W. J. Palenstijn, J. De Beenhouwer, T. Altantzis, S. Bals, K. J. Batenburg, and J. Sijbers, "The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography," *Ultra-microscopy*, vol. 157, pp. 35–47, oct 2015.
- [48] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [49] M. Kheirabadi, A. Bjorholm Dahl, U. Lund Olsen, W. Mustafa, and M. Lyksborg, "Multispectral x-ray CT: multivariate statistical analysis for efficient reconstruction," in *Developments in X-Ray Tomography XI*, B. Müller and G. Wang, Eds., vol. 1039113, no. October 2017. SPIE, oct 2017, p. 38.
- [50] M. Yveborg, M. Danielsson, and H. Bornefalk, "Theoretical Comparison of a Dual Energy System and Photon Counting Silicon Detector Used for Material Quantification in Spectral CT," *IEEE Transactions on Medical Imaging*, vol. 34, no. 3, pp. 796–806, mar 2015.