Uconnect

Wang, Zhihan; Bousse, Alexandre; Vermet, Franck; Froment, Jacques; Vedel, Béatrice; Perelli, Alessandro

DOI:
10.1109/TRPMS.2023.3330045

Publication date:
2023

Uconnect: Synergistic Spectral CT Reconstruction with U-Nets Connecting the Energy bins

Zhihan Wang, Alexandre Bousse, *Member, IEEE*, Franck Vermet, Jacques Froment, Béatrice Vedel, Alessandro Perelli, Jean-Pierre Tasu, Dimitris Visvikis, *Fellow, IEEE*

***Abstract*—Spectral computed tomography (CT) offers the possibility to reconstruct attenuation images at different energy levels, which can be then used for material decomposition. However, traditional methods reconstruct each energy bin individually and are vulnerable to noise. In this paper, we propose a novel synergistic method for spectral CT reconstruction, namely Uconnect. It utilizes trained convolutional neural networks (CNNs) to connect the energy bins to a latent image so that the full binned data is used synergistically. We experiment on two types of low-dose data: simulated and real patient data. Qualitative and quantitative analysis show that our proposed Uconnect outperforms state-of-art model-based iterative reconstruction (MBIR) techniques as well as CNN-based denoising.**

***Index Terms*—Spectral CT, Deep learning, Synergistic Reconstruction, Regularization**

I. Introduction

SPECTRAL CT is fast becoming an important technology in medical imaging for its potential to differentiate between different types of tissues and structures based on their unique spectral properties [1]. Its applications include lesion detection [2], material decomposition [3], [4], automated bone removal [5], etc. As an emerging technology to implement photon-counting CT (PCCT) imaging, photon-counting detectors (PCDs) are capable of collecting multiple sets of measured data with different spectral information in a single exposure by distinguishing photon energies during data acquisition. On the other hand, with the increasing number of energy bins, the number of X-ray photons in each energy bin is limited, which reduces signal-to-noise ratio in spectral projections acquired by PCDs. At the same time efforts to reduce radiation dose, largely motivated by the introduction of potential clinical applications such as lung cancer screening, are associated with decreasing number of projection angles and/or reduced X-ray source flux. In this situation, traditional analytic reconstruction methods such as filtered backprojection will cause severe artifacts and noise. Therefore, many studies have focused on developing methods to improve the reconstruction quality.

Over the past few years, several sparsity-exploiting methods have been applied to low-dose CT reconstruction, such as total variation (TV) [6]–[8], dictionary learning (DiL) [9], and total generalized variation [10]. These methods can be extended to PCCT by applying the sparse regularization to each energy bin separately. However, this kind of strategy ignores dependencies between images acquired at different energy bins, which may lead to suboptimal results. Many algorithms, thus, have been proposed to take advantage of the similarities between different energy bins.

Synergistic regularization terms have been designed to promote structural similarities across channels. Examples include joint TV (JTV) [11], [12] which promotes joint sparsity of the gradients of images, total nuclear variation (TNV) [13], [14] which encourages gradient-sparse solutions in the same way as the conventional TV and also favors solutions that have common edge directions in all channels, and parallel level sets (PLS) [15] which promotes images with similar contours. In addition, the low-rank property is often used as a prior assumption to exploit the correlation among different energy bins. Specifically, the spectral CT data can be represented as a low-rank matrix, where each row corresponds to an energy bin and each column corresponds to a voxel in the image. In this way, it is possible to obtain more accurate and robust spectral CT images during reconstruction [16]–[19]. Another class of approach consists in enforcing structural similarities with a reference clean image [20]–[23].

Synergistic regularization can also be trained as opposed to the use of a fixed analytical model. Tensor DiL (TDiL) [24]–[26] has also emerged as a promising technique for spectral CT reconstruction. It is a powerful extension of traditional dictionary learning, where instead of representing the data using a single dictionary, a collection of dictionaries is learned, such that information can be conveyed across channels. However, they rely on the patch-based sparse representation that may be suboptimal as the trained atoms are often shifted version of each other. To remedy this problem, multichannel convolutional DiL techniques can be deployed [27]. More details can be found in a recent review [28].

Recently, there has been growing interest in applying deep learning (DL) techniques to CT reconstruction [29]–[31].

This work did not involve human subjects or animals in its research.

This work was supported by the French National Research Agency (ANR) under grant No ANR-20-CE45-0020, and by France 2030 framework program, Centre Henri Lebesgue, under grant No ANR-11-LABX-0020-01.

Z. Wang is with the LaTIM, INSERM, UMR 1101, Université de Bretagne Occidentale, Brest, France, and also with Univ Brest and Université Bretagne Sud, Vannes, CNRS, UMR 6205, Laboratoire de Mathematiques de Bretagne Atlantique, France.

A. Bousse, and D. Visvikis are with the LaTIM, INSERM, UMR 1101, Université de Bretagne Occidentale, Brest, France.

F. Vermet is with Univ Brest, CNRS, UMR 6205, Laboratoire de Mathematiques de Bretagne Atlantique, France.

J. Froment and B. Vedel are with Laboratoire de Mathématiques de Bretagne Atlantique, CNRS UMR 6205, Université Bretagne Sud, Vannes, France.

A. Perelli is with the School of Science and Engineering, University of Dundee, Scotland DD1 4HN, UK.

J.-P. Tasu is with the LaTIM, INSERM, UMR 1101, Université de Bretagne Occidentale, Brest, France, and also with the Department of Radiology, University Hospital Poitiers, Poitiers, France.

Corresponding authors: A. Bousse, bousse@univ-brest.fr

Several studies have shown promising results using various DL architectures, such as CNNs, generative adversarial networks, autoencoders, and unrolling architectures. However, few investigations have been carried out in the domain of spectral CT. For example, [32], [33] have proposed DL-based spectral CT multichannel post-processing techniques. Recently, an unrolling architecture, namely SOUL-Net [34], was proposed. However the training of such models is challenging, especially with a large number of channels.

In this paper, we propose a novel spectral CT reconstruction technique, namely Uconnect, where images at each energy bin are connected to a reference image by a collection of U-Nets. We considered the case of PCCT with 6 energy bins, although our method can also be applied to dual-energy CT. The training is performed on an image basis, and therefore does not require a computationally expensive supervised training for sinogram-to-image mappings such as unrolling architectures. The U-Nets are pretrained on a separate dataset to map the attenuation image at a reference energy to other energies and are incorporated into a plug-and-play penalty term which can be used in any MBIR, for any system geometry. To the best of our knowledge, this is the first method that employs a deep-learned penalty for synergistic MBIR. This work follows previous work we presented at the 2022 IEEE Medical Imaging Conference [35].

The rest of the paper is structured as follows. Section II briefly reviews the physical model and several existing approaches for spectral CT reconstruction, and presents the details on our proposed approach Uconnect. Section III shows the details of experiments on synthetic data and real patient data, as well as qualitative and quantitative comparison results with other representative approaches. Section IV discusses several technical issues, while Section V provides our concluding remarks.

II. Methods

A. Forward Model

In this work we consider a standard PCCT set-up. We consider 2-dimensional (2-D) images, although the proposed method can generalize to the 3-dimensional case.

The discrete anatomical image takes the form of an energy-dependent vector $\boldsymbol{x}(E) = [x_1(E), \ldots, x_J(E)]^\top \in \mathbb{R}^J$, where for all $j = 1, \ldots, J$, $x_j(E)$ is the linear attenuation at pixel j and energy E, and J is the total number of pixels (or voxels) in the image, and '$^\top$' denotes the matrix transposition.

The spectral CT system performs measurements along a collection of I rays (from the source to the detectors), with $I = N_\mathrm{d} \times N_\mathrm{s}$, N_d and N_s being respectively the number of detectors and the number of source positions. The measurements are regrouped into K energy bins, each bin k corresponding to an interval $[E_{k-1}, E_k]$ with $E_0 < E_1 < \cdots < E_K$. For each bin $k = 1, \ldots, K$ and each ray $i = 1, \ldots, I$, the expected number of detected photons is given by the Beer-Lambert law as

$$\bar{y}_{i,k}(\boldsymbol{x}) = \int_0^{+\infty} h_k(E) \cdot \mathrm{e}^{-[\boldsymbol{A}\boldsymbol{x}(E)]_i} \, \mathrm{d}E \tag{1}$$

where $h_k \colon \mathbb{R}^+ \to \mathbb{R}^+$ is the photon flux X-ray intensity associated to energy bin k. For simplicity we ignored background events such as scatter but they can be incorporated in the model without affecting the rational of the proposed approach. The matrix $\boldsymbol{A} \in \mathbb{R}^{I \times J}$ is the system matrix modeling line integrals along each beam in the discrete case, and for ray i and pixel j, $[\boldsymbol{A}]_{i,j}$ is the contribution of pixel j to beam i.

Instead of reconstructing $\boldsymbol{x}(E)$ as a function of E, we reconstruct a multichannel image $\{\boldsymbol{x}_k\} \in \left(\mathbb{R}_+^J\right)^K$ where for each $k = 1, \ldots, K$ the image $\boldsymbol{x}_k$ is an "average" attenuation image corresponding to energy bin k, using the simplified forward model

$$\bar{y}_{i,k}(\boldsymbol{x}_k) = \bar{h}_k \cdot \mathrm{e}^{-[\boldsymbol{A}\boldsymbol{x}_k]_i} \tag{2}$$

where $\bar{h}_k = \int h_k(E) \, \mathrm{d}E$ is the mean photon flux X-ray intensity for energy bin k. The number of detected photons at each bin k and ray i is modeled as a Poisson random variable denoted $y_{i,k}$, i.e.,

$$y_{i,k} \sim \mathrm{Poisson}\left(\bar{y}_{i,k}(\boldsymbol{x}_k)\right), \tag{3}$$

and for all i, i', k, k', $y_{i,k}$ and $y_{i',k'}$ are independent if $(i,k) \neq (i',k')$. We denote by $\boldsymbol{y}_k = [y_{1,k}, \ldots, y_{I,k}]^\top \in \mathbb{R}^I$ the vector of measurement at bin k.

B. Conventional Reconstruction

Each $\boldsymbol{x}_k$ can be reconstructed by minimizing the negative log-likelihood of the measurement $\boldsymbol{y}_k$. To simplify the reconstruction, we approximate the negative Poisson distribution with weighted least squares (WLS) loss function L_k as proposed in [36]:

$$L_k(\boldsymbol{x}_k) = \sum_{i=1}^{I} \frac{1}{2} y_{i,k} \left([\boldsymbol{A}\boldsymbol{x}_k]_i - b_{i,k}\right)^2 \tag{4}$$

$$= \frac{1}{2} \left\|\boldsymbol{A}\boldsymbol{x}_k - \boldsymbol{b}_k\right\|_{\boldsymbol{W}_k}^2 \tag{5}$$

with $\boldsymbol{b}_k = [b_{1,k}, \ldots, b_{I,k}]^\top$, $b_{i,k} = \log\left(\bar{h}_k / y_{i,k}\right)$, where we assumed $y_{i,k} > 0$ for all (i,k), and $\boldsymbol{W}_k = \mathrm{diag}\{\boldsymbol{y}_k\}$ is a diagonal matrix of statistical weights. Finally, each $\boldsymbol{x}_k$ can be reconstructed by penalized WLS with positivity constraint:

$$\hat{\boldsymbol{x}}_k = \underset{\boldsymbol{x}_k \in \mathbb{R}_+^J}{\arg\min} \; L_k(\boldsymbol{x}_k) + \beta R(\boldsymbol{x}_k) \tag{6}$$

where $\beta > 0$ is a weight and R is an edge-preserving regularization term. A popular choice is the Huber regularizer which enforces piecewise smoothness [37], defined as

$$R_\mathrm{H}(\boldsymbol{x}) = \sum_{j=1}^{J} \sum_{l \in \mathcal{N}_j} \omega_{j,l} \psi_\delta(x_j - x_l), \tag{7}$$

$$\psi_\delta(t) = \begin{cases} \frac{1}{2}t^2 & \text{if } |t| \leq \delta \\ \delta|t| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases}$$

where $\delta > 0$, $\mathcal{N}_j$ is the index set of the neighbor of pixel j, and $\omega_{j,l}$ is the inverse Euclidean distance between j and l.

C. Synergistic Penalties: Existing Approaches

Instead of reconstructing each $\boldsymbol{x}_k$ independently following (6), the entire collection of images $\{\boldsymbol{x}_k\}$ can be reconstructed simultaneously in a single optimization problem as

$$\{\hat{\boldsymbol{x}}_k\} = \underset{\{\boldsymbol{x}_k\}\in\left(\mathbb{R}_+^J\right)^K}{\arg\min} \sum_{k=1}^{K} L_k(\boldsymbol{x}_k) + \beta R_{\text{syn}}\left(\{\boldsymbol{x}_k\}\right) \quad (8)$$

where in this case R_{syn} is a synergistic regularization term for the multichannel image $\{\boldsymbol{x}_k\}$. Note that the sum of the L_ks can be weighted in order to equally smooth the $\boldsymbol{x}_k$s. In this section we describe existing categories of methods for spectral CT reconstruction and we highlight the methods we implemented for comparison with our proposed approach outlined in the following section II-D.

A first category of methods consists in using a regularizer that enforces structural similarities between channels. An example is JTV, whose regularizer is defined as

$$R_{\text{JTV}}\left(\{\boldsymbol{x}_k\}\right) = \sum_{j=1}^{J} \sqrt{\sum_{k=1}^{K} \left\| [\nabla \boldsymbol{x}_k]_j \right\|_2^2} \quad (9)$$

where $\nabla\colon \mathbb{R}^J \to \left(\mathbb{R}^J\right)^2$ denotes the discrete gradient. This promotes joint sparsity of the gradient of the $\boldsymbol{x}_k$s. Other examples include TNV [13], [14] and PLS [15].

A second category involves controlling the rank of the multichannel image to promote linear dependencies between the channels. It makes sense to assume some form of "low rankness" due to the fact that the energy dependency of human tissues can be expressed as a linear combination of only two materials. In [38], a low-rank and sparse (LRS) decomposition was proposed for the multichannel image matrix $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K] \in \mathbb{R}^{J\times K}$ such that $\boldsymbol{X} = \boldsymbol{L} + \boldsymbol{S}$, where the low-rank matrix $\boldsymbol{L}$ represents the underlying structure or common information between channels and the sparse matrix $\boldsymbol{S}$ reflects distinct spectral features across channels. A synergistic penalty is then defined as

$$R_{\text{LRS}}\left(\boldsymbol{X}\right) = \gamma \left\|\boldsymbol{L}\right\|_* + \left\|\boldsymbol{T}\boldsymbol{S}\right\|_1 \quad (10)$$

where $\boldsymbol{T}$ is a sparsifying transform (such as the gradient or a wavelet transform) with $\gamma > 0$ and the nuclear norm $\|\cdot\|_*$ is a convex relaxation of the rank for an efficient optimization. The method aided by self-similarity in image-spectral tensors (ASSIST) in [19] is a generalization of this approach that selects the M most similar d-dimensional patches within a search window of a clean reference (or prior) image $\boldsymbol{x}^{\text{prior}}$, to form a collection of $d \times M \times K$ tensor units of which the nuclear norm is computed.

A third category of methods consists in reconstructing each channel $\boldsymbol{x}_k$ independently by solving (6) with a single-channel penalty that enforces similarities with a reference low-noise prior image $\boldsymbol{x}^{\text{prior}} \in \mathbb{R}^J$ [20]–[23]. For example in [20] the directional TV (DTV) regularizer promotes common edge directions with the prior image, and is defined for all $\boldsymbol{x} \in \mathbb{R}^J$ as

$$R_{\text{DTV}}(\boldsymbol{x}) = \sum_{j=1}^{J} \left\| \boldsymbol{\Pi}_j [\nabla \boldsymbol{x}]_j \right\|_2 \quad (11)$$

where $\boldsymbol{\Pi}_j = \boldsymbol{I} - \boldsymbol{\xi}_j \otimes \boldsymbol{\xi}_j$, $\boldsymbol{I}$ being the 2×2 identity matrix and $\otimes$ being the outer product of vectors, and

$$\boldsymbol{\xi}_j = \eta \frac{[\nabla \boldsymbol{x}^{\text{prior}}]_j}{\sqrt{\|[\nabla \boldsymbol{x}^{\text{prior}}]_j\|_2^2 + \epsilon}} \quad (12)$$

with $\epsilon > 0$ avoiding singularities when $[\nabla \boldsymbol{x}^{\text{prior}}]_j = 0$, and $\eta < 1$ depending on how much influence $\boldsymbol{x}^{\text{prior}}$ has. In particular, we observe that [39]

$$\left(1 - \eta^2\right) R_{\text{TV}} \leq R_{\text{DTV}} \leq R_{\text{TV}} \quad (13)$$

where the TV regulariser R_{TV} corresponds to R_{DTV} with $\boldsymbol{\Pi}_j = \boldsymbol{I}_{\mathbb{R}^2}$ for all j, and the lower bound is reached when $[\nabla \boldsymbol{x}]_j$ and $\boldsymbol{\xi}_j$ are collinear for all j. $R_{\text{DTV}}(\boldsymbol{x})$ is minimized when the gradient vanishes and therefore potential artifacts in $\boldsymbol{x}^{\text{prior}}$ do not propagate in the reconstructed image. Similar approaches include [23], [40], [41] which use the gradient of the difference between each $\boldsymbol{x}_k$ and $\boldsymbol{x}^{\text{prior}}$. The choices of $\boldsymbol{x}^{\text{prior}}$ will be discussed in Section III.

D. Proposed Learned Synergistic Penalty: Uconnect

The traditional regularization techniques presented in Section II-C enforce a fixed handcrafted penalty on the reconstructed image due to certain assumptions about its structure, such as sparsity or smoothness. However, these assumptions may not always be valid in practice. By contrast, learned penalty functions adjust adaptively according to the data's specific characteristics, providing more accurate and adaptable reconstructions.

1) General principle: Learned synergistic penalties can be derived from multichannel DiL [24], which consists in training a collection of dictionaries $\boldsymbol{D}_1, \ldots, \boldsymbol{D}_K$, i.e., over-complete bases comprising S atoms, to approximately represent each image $\boldsymbol{x}_k$ with a fraction of their columns with a single sparse code $\boldsymbol{z} \in \mathbb{R}^S$, that is to say, $\boldsymbol{x}_k \approx \boldsymbol{D}_k \boldsymbol{z}$ for all k, in order to convey the information across channels. Unfortunately, a large number of atoms is required to accurately represent all possible spectral images, which can compromise training. Therefore, to reduce the complexity, the image is split into P smaller d-dimensional "patches" with $d \ll J$. The synergistic penalty is therefore defined as follows:

$$R_{\text{dict}}\left(\{\boldsymbol{x}_k\}\right) = \min_{\{\boldsymbol{z}_p\}\in(\mathbb{R}^S)^P} \sum_{k=1}^{K}\sum_{p=1}^{P} \frac{1}{2}\|\boldsymbol{P}_p \boldsymbol{x}_k - \boldsymbol{D}_k \boldsymbol{z}_p\|_2^2 + \alpha \|\boldsymbol{z}_p\|_q \quad (14)$$

where $\boldsymbol{P}_p \in \mathbb{R}^{d\times J}$ is the p^{th} patch extractor, $\boldsymbol{D}_k \in \mathbb{R}^{d\times S}$ is the k^{th} trained dictionary, $\boldsymbol{z}_p \in \mathbb{R}^S$ is the sparse coding (same for all channel k), $\|\cdot\|_q$ is either the ℓ_1-norm or the ℓ_0 semi-norm, and $\alpha > 0$. In [24] the authors used tensor dictionaries to represent the spectral image based on canonical polyadic decomposition, which is similar to (14). Similar coupled dictionary models have been used in multimodal imaging synergistic reconstruction, for example in positron emission tomography (PET)/magnetic resonance imaging (MRI) [42], [43] and multi-contrast MRI [44].

Despite the possibility to extract patches at random locations over the image, patch-based DiL may suffer from inefficiency

due to the shift-variant nature of the atoms that can generate duplicates during training. Furthermore, the utilization of many neighboring/overlapping patches across the training images is not optimal for sparse representation, as sparsification is performed independently on each patch.

In this work we propose an alternative to patch-based DiL. We assume that each image $\boldsymbol{x}_k$ is acquired from a single object $\boldsymbol{z} \in \mathbb{R}^N$ (non necessarily sparse), N being the dimension of the latent space (in DiL N corresponds to the number of atoms S), with a collection of trained CNNs $\{\boldsymbol{f}_k\}$, such that

$$\boldsymbol{f}_k(\boldsymbol{z}) \approx \boldsymbol{x}_k, \quad \forall k = 1, \dots, K\,. \tag{15}$$

The mappings $\{\boldsymbol{f}_k\}$ play the role of the "non-linear" dictionaries for the entire image (without patches), which connect all energy bins through $\boldsymbol{z}$. The proposed joint regularization term is therefore defined as

$$R_{\mathrm{NN}}\left(\{\boldsymbol{x}_k\}\right) = \min_{\boldsymbol{z}\in\mathbb{R}^N} \sum_{k=1}^{K} \frac{1}{2}\gamma_k \left\|\boldsymbol{f}_k(\boldsymbol{z}) - \boldsymbol{x}_k\right\|_2^2 + \alpha H(\boldsymbol{z}) \tag{16}$$

where H is a penalty term on $\boldsymbol{z}$. This representation is a generalization of (14) with non-linear mappings $\{\boldsymbol{f}_k\}$ instead of dictionaries, and without patch extraction. The regularizer $R_{\mathrm{NN}}\left(\{\boldsymbol{x}_k\}\right)$ is small if (i) each image $\boldsymbol{x}_k$ is close to $\boldsymbol{f}_k(\boldsymbol{z})$ for some $\boldsymbol{z}$ and (ii) $\boldsymbol{z}$ is piecewise smooth (in the sense of H). By doing so, the $\boldsymbol{x}_k$s are "connected" via a single latent variable $\boldsymbol{z}$, (cf. the schematic representation of the generative model in Fig. 1) such that each $\boldsymbol{x}_k$ is reconstructed using the entire measurement data at each energy bin.

The role of the γ_ks is to give more weight to less noisy images. A logical choice would be $\gamma_k = \|\boldsymbol{y}_k\|_1 / \sum_k \|\boldsymbol{y}_k\|_1$. However, most of the counts come from the rays that do not intersect the patient, and hence this approach does not account for the loss of counts due to higher attenuation. We therefore define $\gamma_k = \|\check{\boldsymbol{y}}_k\|_1 / \sum_k \|\check{\boldsymbol{y}}_k\|_1$ through manual experimentation, where $\check{\boldsymbol{y}}_k \in \mathbb{R}^{I'}$, $I' < I$, is the vector composed of the smallest first quarter of $\boldsymbol{y}_k$ in order to only account for rays intersecting the patient.

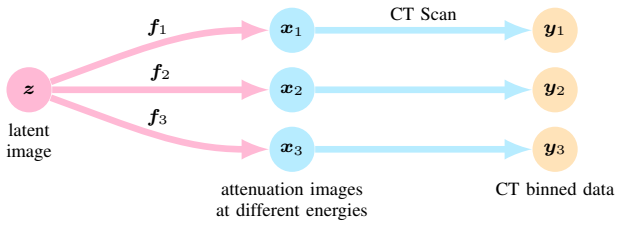


Fig. 1. Schematic representation of our generative model with $K = 3$ energy bins.

The training of the $\boldsymbol{f}_k$s can be unsupervised, for example using Wasserstein GANs (W-GANs) [45] or variational autoencoders (VAEs) [46] with some latent variable $\boldsymbol{z}$ that has no physical interpretation. However, this training can be challenging for a large number of channels K as the $\boldsymbol{f}_k$s must be jointly trained. We therefore propose the following alternative. Our method, namely *Uconnect*, consists of "connecting" the energies by the mean of image-to-image U-Nets [47] $\boldsymbol{f}_k : \mathbb{R}^J \to \mathbb{R}^J$ trained in a supervised fashion to map a CT image at energy bin 1 to energy bin k, that is to say[1],

$$\boldsymbol{f}_k(\boldsymbol{x}_1) \approx \boldsymbol{x}_k, \quad \forall k = 1, \dots, K\,, \tag{17}$$

In this model, the latent variable is an image, and the dimension of the latent space is $N = J$. This approach is somehow similar to the CNN representation proposed in [48] for PET, in the sense that they both assume that the images are outputs of some CNNs trained with clean images as target in order to reduce the noise. However, our method utilizes a single input $\boldsymbol{z}$ for all images, thus further reducing the noise as we will show in Section III-A2 and Section III-B2. In addition, the generative model is used within a penalty term, which is less constraining than the strict equality $\boldsymbol{x} = \boldsymbol{f}(\boldsymbol{z})$ used in [48].

The choice of the reference image $\boldsymbol{z} \approx \boldsymbol{x}_1$, i.e., the lowest-energy bin, is not arbitrary. It is generally accepted that CT images of the same anatomical part but acquired at different energies exhibit similar patterns despite different attenuation values. To take advantage of this similarity, we designated the CT image obtained from the lowest-energy bin as the reference image owing to the fact that low-energy attenuation images contain more features than high-energy ones, therefore high-energy to low-energy mappings would have to be trained to reveal features, which is more challenging than the other way around. By leveraging this approach, the shared characteristics across energy levels can enhance the accuracy and quality of spectral CT reconstructions. Note that the model (17) does not replicate the exact physics, as the attenuation at each energy cannot be predicted from a single image; the material decomposition should be used to generate the virtual mono-energetic images (VMIs), followed by an averaging over each energy bin. However, as we will show in Section III, the lowest-energy image can reasonably predict the images at other energies. Similar observation was made in [49].

To finish, $\boldsymbol{z}$ is an image and therefore the penalty function H can be any image regularizer such as R_{H} (7), or even R_{DTV} (11).

2) Training and Implementation: In this work, K mappings $\{\boldsymbol{f}_k\}$ need to be trained for each energy bin $k = 1, \dots, K$. The overall network architecture of the U-Net structure we consider for each $\boldsymbol{f}_k$ is shown in Fig. 2. This CNN has 23 convolutional layers in total. The number of trainable parameters is around 30 millions. The training is performed using a collection of L spectral CT images $\left\{\boldsymbol{x}_{k,l}^{\mathrm{train}}\right\}_{k,l=1}^{K,L}$, $\boldsymbol{x}_{k,l}^{\mathrm{train}}$ representing the CT image at bin k from the l^{th} dataset. Since the attenuation values vary across energy bins we introduced scaling factors $s_{k,l}$ in order to facilitate training, such that

$$\boldsymbol{f}_k\left(\boldsymbol{x}_{1,l}^{\mathrm{train}}\right) \approx s_{k,l}\boldsymbol{x}_{k,l}^{\mathrm{train}}, \quad \forall k = 1, \dots, K, \quad \forall l = 1, \dots, L \tag{18}$$

with $s_{k,l} = \dfrac{\|\boldsymbol{x}_{1,l}^{\mathrm{train}}\|_1}{\|\boldsymbol{x}_{k,l}^{\mathrm{train}}\|_1}$. We consider $\left\{\boldsymbol{x}_{1,l}^{\mathrm{train}}\right\}_{l=1}^{L}$ as input images and $\left\{\boldsymbol{x}_{k,l}^{\mathrm{train}}\right\}_{k,l=1}^{K,L}$ as their corresponding target images. L is

[1]An alternative to R_{NN} consists in dropping the latent variable $\boldsymbol{z}$, i.e., $R_{\mathrm{NN}}^{\mathrm{alt}}\left(\{\boldsymbol{x}_k\}\right) = \sum_{k=2}^{K} \frac{1}{2}\gamma_k \left\|\boldsymbol{f}_k(\boldsymbol{x}_1) - \boldsymbol{x}_k\right\|_2^2 + \alpha H(\boldsymbol{x}_1)$. Note that this penalty is limited image-to-image CNNs while R_{NN} in (16) can use any generative model defined on a latent space.

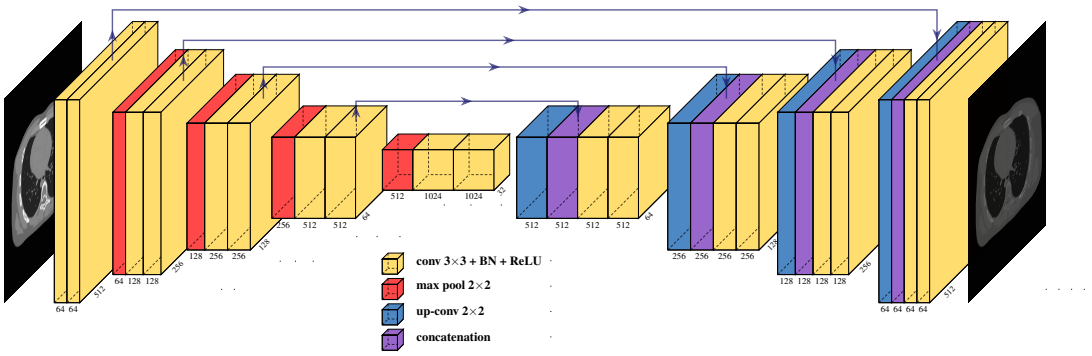


Fig. 2. U-Net architecture used for the $\boldsymbol{f}_k$.

the amount of labeled data for training. Note that we also train the network for $k = 1$, i.e., bin 1 to bin 1, in order to take advantage of the denoising properties of the $\boldsymbol{f}_k$. The training was achieved with mean squared error and Adam optimizer.

To account for the scaling factors, a modified version of the synergistic penalty (16) is used for our experiments as follows:

$$R_{\mathrm{NN}}\left(\{\boldsymbol{x}_k\}\right) = \underset{\boldsymbol{z}\in\mathbb{R}^N}{\arg\min} \sum_{k=1}^{K} \frac{1}{2}\gamma_k \left\|\boldsymbol{f}_k(\boldsymbol{z}) - \hat{s}_k\boldsymbol{x}_k\right\|_2^2 + \alpha H(\boldsymbol{z}) \tag{19}$$

where $\hat{s}_k$ is an estimated scaling factor, which can be obtained from scout reconstructions, for example by solving (6) for each bin independently. Note that this factor is object-dependent, although we treat it here as a constant.

3) Reconstruction Algorithm: Solving (8) using the synergistic penalty (19) is performed by alternating minimization with respect to $\boldsymbol{z}$ and $\{\boldsymbol{x}_k\}$, i.e., given the estimates $\left\{\boldsymbol{x}_k^{(n)}\right\}$ and $\boldsymbol{z}^{(n)}$ at iteration n, the new estimates at iteration $n+1$ are given by

$$\boldsymbol{z}^{(n+1)} = \underset{\boldsymbol{z}\in\mathbb{R}_+^J}{\arg\min} \sum_{k=1}^{K} \frac{1}{2}\gamma_k \left\|\boldsymbol{f}_k(\boldsymbol{z}) - \hat{s}_k\boldsymbol{x}_k^{(n)}\right\|_2^2 + \alpha H(\boldsymbol{z}) \tag{20}$$

$$\boldsymbol{x}_k^{(n+1)} = \underset{\boldsymbol{x}\in\mathbb{R}_+^J}{\arg\min}\; L_k(\boldsymbol{x}) + \beta\gamma_k \frac{1}{2}\left\|\boldsymbol{f}_k\left(\boldsymbol{z}^{(n+1)}\right) - \hat{s}_k\boldsymbol{x}\right\|_2^2, \quad \forall k = 1,\ldots,K \tag{21}$$

We treat each sub-problem with iterative algorithms.

When H is differentiable (e.g., $H = R_{\mathrm{H}}$), the $\boldsymbol{z}$-update (20) is achieved with a limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) algorithm [50] (initialized from $\boldsymbol{z}^{(n)}$) with non-negativity constraints, which is a quasi-Newton optimization algorithm that only requires the gradient of the objective function. We used the `GradientTape` function in TensorFlow to compute the gradient of a function with respect to its input variables. When H is non-differentiable (e.g., TV), the minimization can achieved with a Chambolle-Pock algorithm [51], combined with L-BFGS for the proximal of $\left\|\boldsymbol{f}_k(\cdot) - \hat{s}_k\boldsymbol{x}_k^{(n)}\right\|_2^2$.

For the $\boldsymbol{x}_k$-update (21), we used a separable quadratic surrogate (SQS) algorithm [36], i.e., using a diagonal matrix $\boldsymbol{H}_k$ that majorizes $\boldsymbol{A}^\top\boldsymbol{W}_k\boldsymbol{A}$ (e.g., $\boldsymbol{H}_k = \mathrm{diag}(\boldsymbol{A}^\top\boldsymbol{W}_k\boldsymbol{A}\mathbf{1})$ where $\mathbf{1}$ is a vector containing 1 only, see [52]). Let $\boldsymbol{x}_k^{(n,m)}$ be the current estimate at iteration n, sub-iteration m. The new image at sub-iteration $m+1$ is given by

$$\begin{aligned}\tilde{\boldsymbol{x}}_k &= \boldsymbol{x}_k^{(n,m)} - \boldsymbol{H}_k^{-1}\boldsymbol{A}^\top\boldsymbol{W}_k\left(\boldsymbol{A}\boldsymbol{x}_k^{(n,m)} - \boldsymbol{b}_k\right)\\ \boldsymbol{x}_k^{(n,m+1)} &= \left[\frac{\boldsymbol{H}_k\tilde{\boldsymbol{x}}_k + \beta\gamma_k\hat{s}_k\boldsymbol{f}_k\left(\boldsymbol{z}^{(n+1)}\right)}{\boldsymbol{H}_k + \beta\gamma_k\hat{s}_k^2}\right]_+\end{aligned} \tag{22}$$

where $[\cdot]_+$ denotes the positive part of a vector.

We summarize the reconstruction steps in Algorithm 1. Before the main loop that alternates minimization with respect to $\boldsymbol{z}$ and $\{\boldsymbol{x}_k\}$, we initialize $\left\{\boldsymbol{x}_k^{(0)}\right\}$ either by images of value 0 or by preliminary images $\{\hat{\boldsymbol{x}}_k\}$ reconstructed individually using WLS without penalty for faster convergence. Similarly, $\boldsymbol{z}^{(0)}$ can be initialized by $\boldsymbol{x}_1^{(0)}$. The convergence speed of the algorithm depends on β. Since R_{NN} is non-convex, proving the convergence is challenging. However, we observed in our experiments that the algorithm converges, even for large values of β.

Algorithm 1: Algorithm for multi-energy CT reconstruction with Uconnect

Input: System matrix $\boldsymbol{A}$; Measurements $\{\boldsymbol{y}_k\}$; Initial images $\left\{\boldsymbol{x}_k^{(0)}\right\}$; Initial latent image $\boldsymbol{z}^{(0)}$; Trained mappings: $\{\boldsymbol{f}_k\}$; Scaling factors: $\{\hat{s}_k\}$; Weights: $\{\gamma_k\}$, α, and β.
Output: Reconstructed images $\{\hat{\boldsymbol{x}}_k\}$.

$\boldsymbol{H}_k \leftarrow \left\{\mathrm{diag}\left(\boldsymbol{A}^\top\boldsymbol{W}_k\boldsymbol{A}\mathbf{1}\right)\right\}, \quad k = 1,\ldots,K$;
for $n = 0$ **to** $N_{\mathrm{outer}} - 1$ **do**
  $\boldsymbol{z}^{(n+1)} \leftarrow \arg\min_{\boldsymbol{z}} \sum_k \frac{1}{2}\gamma_k \left\|\boldsymbol{f}_k(\boldsymbol{z}) - \hat{s}_k\boldsymbol{x}_k^{(n)}\right\|_2^2 + \alpha H(\boldsymbol{z})$ (using L-BFGS initialized with $\boldsymbol{z}^{(n)}$) ;
  for $k \leftarrow 1$ **to** K **do**
    $\boldsymbol{x}_k^{(n,0)} \leftarrow \boldsymbol{x}_k^{(n)}$;
    for $m = 0$ **to** $N_{\mathrm{inner}} - 1$ **do**
      $\tilde{\boldsymbol{x}}_k \leftarrow \boldsymbol{x}_k^{(n,m)} - \boldsymbol{H}_k^{-1}\boldsymbol{A}^\top\boldsymbol{W}_k\left(\boldsymbol{A}\boldsymbol{x}_k^{(n,m)} - \boldsymbol{b}_k\right)$
      $\boldsymbol{x}_k^{(n,m+1)} \leftarrow \left[\frac{\boldsymbol{H}_k\tilde{\boldsymbol{x}}_k + \beta\gamma_k\hat{s}_k\boldsymbol{f}_k\left(\boldsymbol{z}^{(n+1)}\right)}{\boldsymbol{H}_k + \beta\gamma_k\hat{s}_k^2}\right]_+$;
    end
    $\boldsymbol{x}_k^{(n+1)} \leftarrow \boldsymbol{x}_k^{(n,N_{\mathrm{inner}})}$;
  end
end
$\hat{\boldsymbol{x}}_k \leftarrow \boldsymbol{x}_k^{(N_{\mathrm{outer}})}, \quad k = 1,\ldots,K$;

In addition to image reconstruction, our proposed penalty R_{NN} can also be applied for denoising by solving

$$\{\boldsymbol{x}^{\star}\} = \underset{\{\boldsymbol{x}_k\}}{\arg\min} \sum_{k=1}^{K} \frac{1}{2}\gamma_k \left\|\boldsymbol{x}_k - \boldsymbol{x}_k^{\mathrm{noisy}}\right\|_2^2 + \beta R_{\mathrm{NN}}(\{\boldsymbol{x}_k\}), \tag{23}$$

where the weights γ_k are used for equal smoothing across channels. The $\boldsymbol{z}$-update is the same as in (20) while the $\boldsymbol{x}_k$-update does not require an iterative algorithm (see Algorithm 2).

Algorithm 2: Algorithm for multi-energy CT denoising with Uconnect

Input: Noisy images $\left\{\boldsymbol{x}_k^{\mathrm{noisy}}\right\}$; Initial latent image $\boldsymbol{z}^{(0)}$; Trained mappings: $\{\boldsymbol{f}_k\}$; Scaling factors: $\{\hat{s}_k\}$; Weights: $\{\gamma_k\}$, α, and β.
Output: Denoised images $\left\{\boldsymbol{x}_k^{\mathrm{denoised}}\right\}$.

$\boldsymbol{x}_k^{(0)} \leftarrow \boldsymbol{x}_k^{\mathrm{noisy}}, \quad k = 1, \ldots, K$;
for $n = 0$ **to** $N-1$ **do**
  $\boldsymbol{z}^{(n+1)} \leftarrow \arg\min_{\boldsymbol{z}} \sum_k \frac{1}{2}\gamma_k \left\|\boldsymbol{f}_k(\boldsymbol{z}) - \hat{s}_k \boldsymbol{x}_k^{(n)}\right\|_2^2 + \alpha H(\boldsymbol{z})$ (using L-BFGS initialized with $\boldsymbol{z}^{(n)}$) ;
  $\boldsymbol{x}_k^{(n+1)} \leftarrow \left[\frac{\boldsymbol{x}_k^{\mathrm{noisy}} + \beta \boldsymbol{f}_k\left(\boldsymbol{z}^{(n)}\right)}{1+\beta}\right]_+, \quad k = 1, \ldots, K$;
end
$\boldsymbol{x}_k^{\mathrm{denoised}} \leftarrow \boldsymbol{x}_k^{(N)}, \quad k = 1, \ldots, K$;

III. Experiments

We conducted experiments on two types of data: synthetic data and real patient images. In both cases, we considered 2-D CT images at $K = 6$ energies: $E_1 = 40$ keV, $E_2 = 60$ keV, $E_3 = 80$ keV, $E_4 = 100$ keV, $E_5 = 120$ keV and $E_6 = 140$ keV.

U-Net architectures were built in Python 3.8 with Tensorflow 2.10 using NVIDIA GeForce GTX 1660 Ti (6 GB memory). In both experiments, the models were trained on one dataset, and were tested on a separate dataset of the same type.

We compared the approaches mentioned in Section II, i.e., Huber, JTV, DTV, ASSIST, as well as our proposed Uconnect method with $H = R_{\mathrm{H}}$ (Huber), for two different tasks: image reconstruction in Section III-A and image denoising in Section III-B. In addition, we also conducted a comparison with a DL approach called residual encoder-decoder convolutional neural network (RED-CNN) [53] in terms of image denoising task for patient data in Section III-B. We used a SQS algorithm [36] for the individual reconstructions with Huber penalty and the Chambolle-Pock algorithm [51], [54] for JTV and DTV. All algorithms were implemented in Python programming language.

The objective measures structural similarity index measure (SSIM) and peak signal-to-noise ratio (PSNR), with respect to a reference image $\boldsymbol{x}^{\mathrm{ref}} = \left[x_1^{\mathrm{ref}}, \ldots, x_J^{\mathrm{ref}}\right]^{\top}$, were employed for quantitative evaluation of the image quality. We utilized the functions `structural_similarity` and `peak_signal_noise_ratio` from the Python package `skimage.metrics`

A. Synthetic data

1) Data Preparation, Training, and Setup: We used the extended cardiac-torso (XCAT) software [55] to generate 2 phantoms $512 \times 512 \times 280$, one male and one female, at the different energies described above. We used the $L = 280$ slices of the male phantom to train the $\boldsymbol{f}_k$s throughout 700 epochs. The batch size was set to 5 and the learning rate was set to $2 \cdot 10^{-4}$. We considered the female phantom as the ground truth (GT) $\left\{\boldsymbol{x}_k^{\mathrm{gt}}\right\}$ for reconstruction experiments.

For each energy, we utilized a constant intensity $\bar{h}_k = 5000$ to simulate low-dose (low photon count) measurements $\{\boldsymbol{y}_k\}$ following (3) (with $\boldsymbol{x}_k = \boldsymbol{x}_k^{\mathrm{gt}}$) with $N_{\mathrm{s}} = 120$ projection angles and $N_{\mathrm{d}} = 750$ detectors using a fan beam projector within the ASTRA toolbox [56].

The prior image used for DTV and ASSIST was reconstructed using the unbinned data $\boldsymbol{y} = [y_1, \ldots, y_I]^{\top}$, $y_i = \sum_k y_{i,k}$ and the Huber penalty, i.e., as

$$\boldsymbol{x}^{\mathrm{prior}} = \underset{\boldsymbol{x} \in \mathbb{R}_+^J}{\arg\min} \frac{1}{2} \|\boldsymbol{A}\boldsymbol{x} - \boldsymbol{b}\|_{\boldsymbol{W}}^2 + \beta R_{\mathrm{H}}(\boldsymbol{x}) \tag{24}$$

with $\boldsymbol{b} = [b_i, \ldots, b_I]^{\top}$, $b_i = \log\left(\sum_{k=1}^K \bar{h}_k / y_i\right)$, $\boldsymbol{W} = \mathrm{diag}\{\boldsymbol{y}\}$ and $\beta = 300$, using a SQS-based iterative algorithm. For ASSIST, we searched 48 most similar overlapped patches of size 12×12 with a stride of 6 in a search window of 10×10 to construct tensor units. For DTV, we set $\eta = 0.7$ and $\epsilon = 10^{-5}$ in (12). For Uconnect, we manually set $\alpha = 1$ in (20), $N_{\mathrm{outer}} = 50$ and $N_{\mathrm{inner}} = 50$ in Algorithm 1.

2) Pre-evaluation: Uconnect is motivated by the hypothesis that combining the data from all energy bins may help in reducing noise. To verify this, we investigated the quality of the generated image $\boldsymbol{f}_k(\boldsymbol{z})$ used in R_{NN} (19), where the latent variable $\boldsymbol{z}$ is obtained using all energy bins, compared with the image obtained from each bin k individually, without the regularizer H (i.e., $\alpha = 0$), that is to say

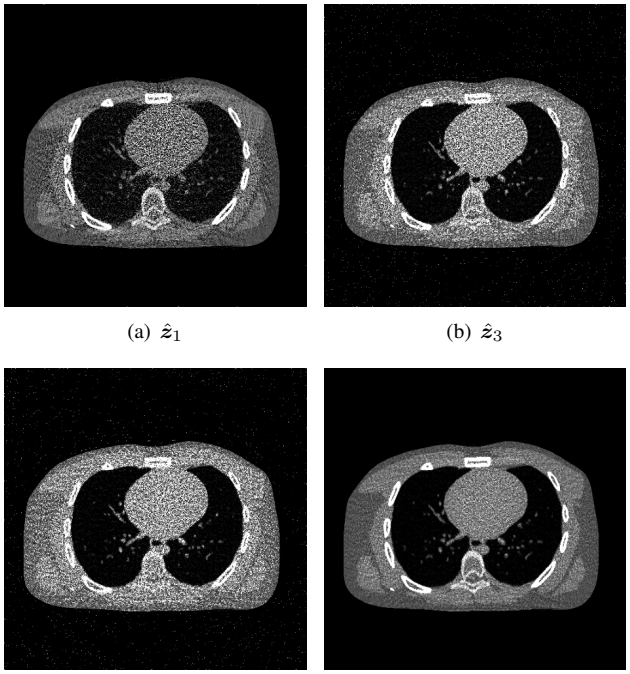
$$\hat{\boldsymbol{z}}_{\mathrm{all}} = \underset{\boldsymbol{z} \in \mathbb{R}_+^J}{\arg\min} \sum_{k=1}^{K} \frac{1}{2}\gamma_k \|\boldsymbol{f}_k(\boldsymbol{z}) - \hat{s}_k \hat{\boldsymbol{x}}_k\|_2^2 \tag{25}$$

and

$$\hat{\boldsymbol{z}}_k = \underset{\boldsymbol{z} \in \mathbb{R}_+^J}{\arg\min} \frac{1}{2} \|\boldsymbol{f}_k(\boldsymbol{z}) - \hat{s}_k \hat{\boldsymbol{x}}_k\|_2^2 \tag{26}$$

where the images $\hat{\boldsymbol{x}}_k$ are scout reconstructions obtained by WLS (i.e., solving (6) with $\beta = 0$) with SQS and 1000 iterations. Note that solving (26) corresponds to "inverting" $\boldsymbol{f}_k$. Fig. 3 shows the estimated latent variables $\hat{\boldsymbol{z}}_{\mathrm{all}}$ and $\hat{\boldsymbol{z}}_k$, $k = 1, 3, 6$, and we observe that using all channels reduces the noise. The generated images $\boldsymbol{f}_k(\hat{\boldsymbol{z}}_{\mathrm{all}})$ and $\boldsymbol{f}_k(\hat{\boldsymbol{z}}_k)$ are shown alongside their SSIM and PSNR (using $\boldsymbol{x}_k^{\mathrm{gt}}$ as the reference images) in Fig. 4. We observe that both SSIM and PSNR values of $\boldsymbol{f}_k(\hat{\boldsymbol{z}}_{\mathrm{all}})$ is higher than those of each $\boldsymbol{f}_k(\hat{\boldsymbol{z}}_k)$, which supports our initial hypothesis.

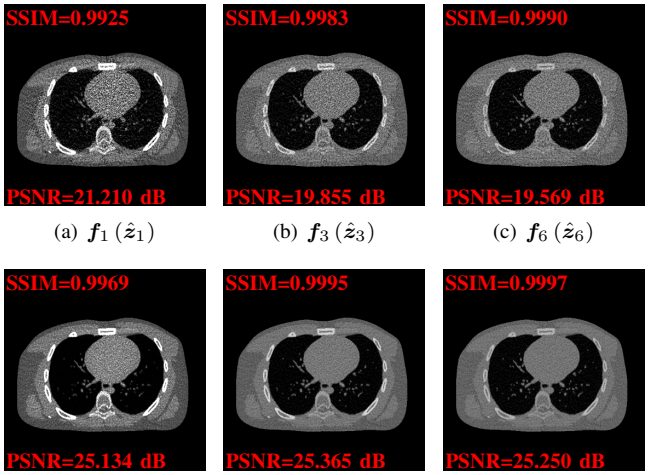
3) Reconstruction Results: The images reconstructed by different approaches for the XCAT phantom at 40, 80 and 140 keV (respectively from left to right columns) with zoomed-in areas marked by yellow boxes are shown in Fig. 5 in Hounsfield unit value. Their respective chosen optimal parameter β corresponds to the highest PSNR in Fig. 6 below.



(a) $\hat{\boldsymbol{z}}_1$ (b) $\hat{\boldsymbol{z}}_3$

(c) $\hat{\boldsymbol{z}}_6$ (d) $\hat{\boldsymbol{z}}_{\text{all}}$

Fig. 3. Latent variables $\hat{\boldsymbol{z}}_1$, $\hat{\boldsymbol{z}}_3$ and $\hat{\boldsymbol{z}}_6$ obtained from (26), and $\hat{\boldsymbol{z}}_{\text{all}}$ obtained from (25) (using the scout reconstructions $\hat{\boldsymbol{x}}_k$).



(a) $\boldsymbol{f}_1(\hat{\boldsymbol{z}}_1)$ (b) $\boldsymbol{f}_3(\hat{\boldsymbol{z}}_3)$ (c) $\boldsymbol{f}_6(\hat{\boldsymbol{z}}_6)$

(d) $\boldsymbol{f}_1(\hat{\boldsymbol{z}}_{\text{all}})$ (e) $\boldsymbol{f}_3(\hat{\boldsymbol{z}}_{\text{all}})$ (f) $\boldsymbol{f}_6(\hat{\boldsymbol{z}}_{\text{all}})$

Fig. 4. Generated images $\boldsymbol{f}_1(\hat{\boldsymbol{z}}_1)$, $\boldsymbol{f}_3(\hat{\boldsymbol{z}}_3)$ and $\boldsymbol{f}_6(\hat{\boldsymbol{z}}_6)$ obtained from (26), and $\boldsymbol{f}_k(\hat{\boldsymbol{z}}_{\text{all}})$ obtained from (25) (using the scout reconstructions $\hat{\boldsymbol{x}}_k$).

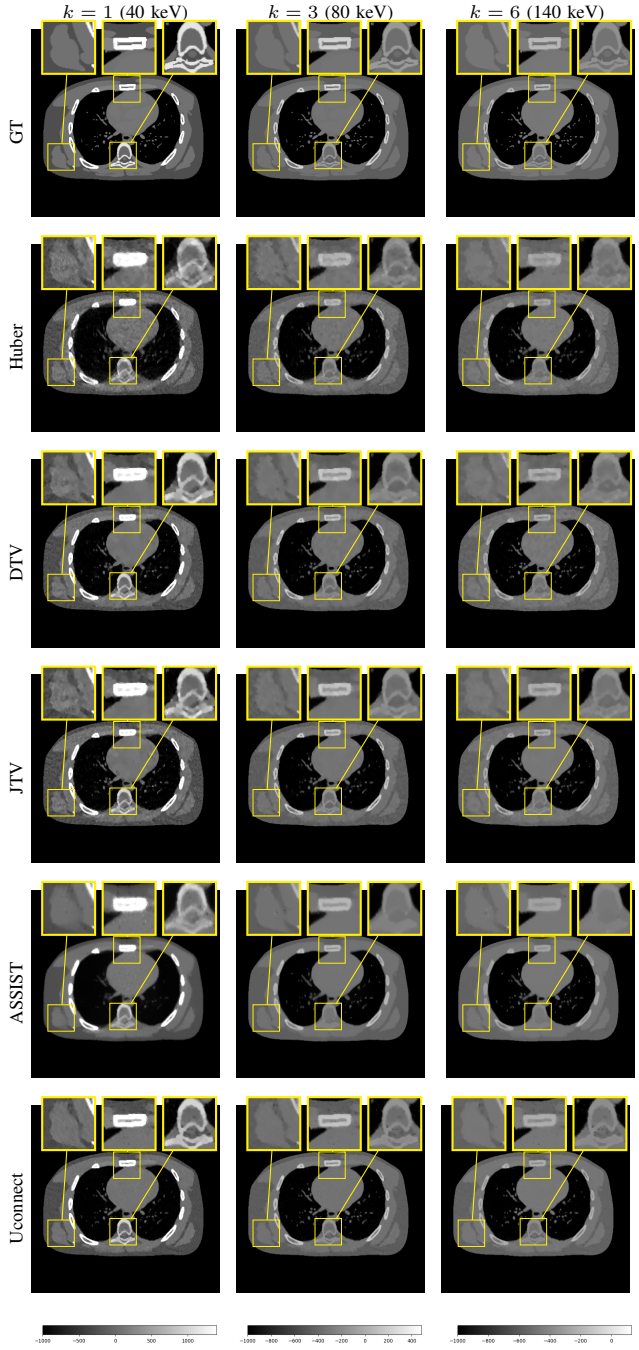


Fig. 5. GT and reconstructed images at $E_1 = 40$, $E_2 = 80$ and $E_3 = 140$ keV using different methods.

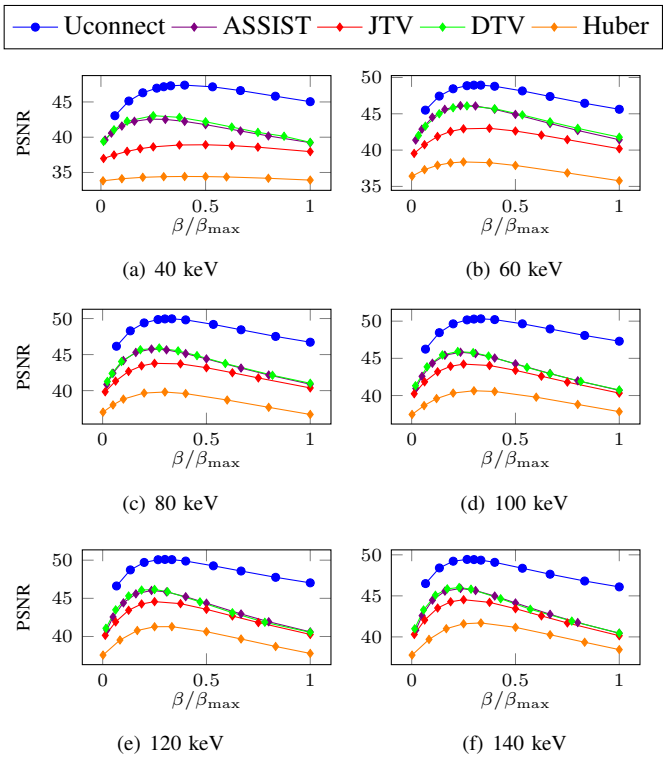
It is visually evident that the results of Huber, DTV and JTV remain noisy and tend to over-smooth bone structures. ASSIST tends to oversmooth the entire image especially at 140 keV, despite good denoising performance. In contrast, our proposed method Uconnect is capable of both preserving features and reducing noise.

We also evaluated quantitatively the aforementioned methods via the index PSNR and SSIM, using the GT images as the reference. Images were reconstructed with different values of the regularization β, from $\beta = \beta_{\min}$ to $\beta_{\max}$, where $\beta_{\min}$ and $\beta_{\max}$ are different for each method due to the variation in order of magnitude of the different penalties.

Fig. 6 shows the PSNR values of the reconstructed images at each of the 6 energy bins using different approaches. The images using Uconnect show the highest PSNR values at all energies, with gains between 2.5 and 5.5 dB in PSNR compared to the best performance by the other approaches at each energy. Similar results in terms of SSIM are observed (Fig. 7). These quantitative results are consistent with the above visual comparison.

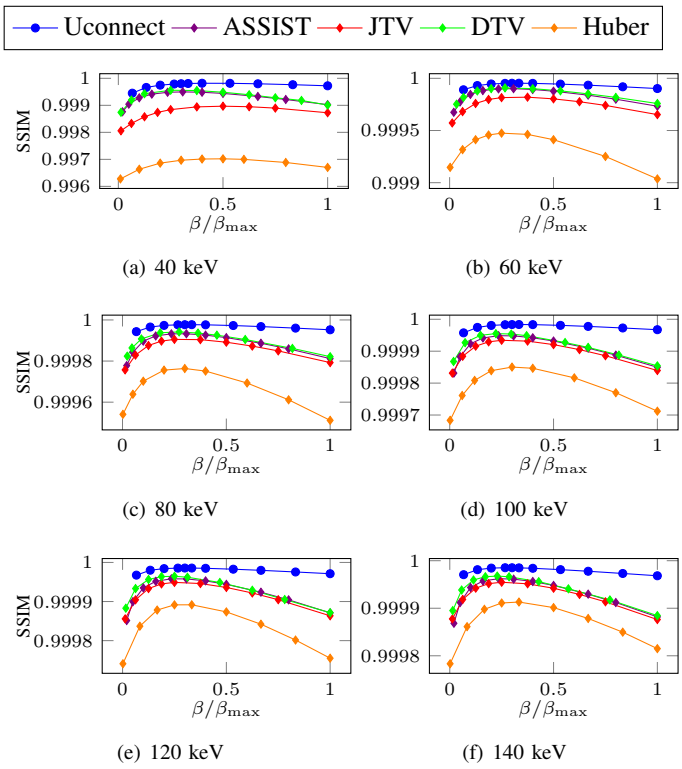
B. Patient Data

1) Data Preparation, Training, and Setup: 2 collections of real patient data images of size 512×512 from 2 different patients were obtained by the Philips IQon Spectral CT scanner from the Poitiers University Hospital, Poitiers, France. The Philips IQon Spectral CT is a dual energy scanner that produces 2 images at 80 kVp and 120 kVp from which 6 VMIs were obtained by interpolation (same energies as in Section III-A). $L = 360$ slices of the first dataset were used to train the $\boldsymbol{f}_k$s throughout 500 epochs. The batch size was set to 5 and the learning rate was set to 0.0002.



(a) 40 keV (b) 60 keV (c) 80 keV (d) 100 keV (e) 120 keV (f) 140 keV

Fig. 6. PSNR values of the reconstructed images using different methods.



(a) 40 keV (b) 60 keV (c) 80 keV (d) 100 keV (e) 120 keV (f) 140 keV

Fig. 7. SSIM values of the reconstructed images using different methods.

Owing to the fact that these images do not represent GTs attenuation of real patients, reconstructing from simulated measurements $\{\boldsymbol{y}_k\}$ with Poisson noise added, as we did with the XCAT phantom, was not considered. Therefore we assessed the different penalties for a denoising task with a least-squares data-fidelity term, i.e.

$$L_k(\boldsymbol{x}_k) = \frac{1}{2}\left\|\boldsymbol{x}_k - \boldsymbol{x}_k^{\text{noisy}}\right\|_2^2 \tag{27}$$

where $\boldsymbol{x}_k^{\text{noisy}}$ is the noisy image simulated as

$$\boldsymbol{x}_k^{\text{noisy}} = \boldsymbol{x}_k^{\text{patient}} + \boldsymbol{\varepsilon}_k, \quad \boldsymbol{\varepsilon}_k \sim \mathcal{N}\left(\boldsymbol{0},\, \boldsymbol{\Sigma}_k^2\right), \tag{28}$$

and $\boldsymbol{x}_k^{\text{patient}}$ a patient image from the second dataset, and $\boldsymbol{\Sigma}_k^2 = \operatorname{diag}\left\{\sigma_{1,k}^2, \ldots, \sigma_{J,k}^2\right\}$. The Uconnect approach for denoising corresponds to solving (23). The other methods were implemented by using the same approach with different penalties.

We estimated realistic $\sigma_{j,k}$ values with a Monte-Carlo approach. We created $Q = 30$ independent measurements with random Poisson noise following (3) with $\boldsymbol{x}_k = \left\{\boldsymbol{x}_k^{\text{patient}}\right\}$ for each energy $k = 1, \ldots, K$, with intensity $\bar{h}_k = \frac{2}{3} \cdot 10^5$ for all k, from which we reconstructed the spectral images $\boldsymbol{x}_k^{[q]} = \left[x_{1,k}^{[q]}, \ldots, x_{J,k}^{[q]}\right]^\top$, $k = 1, \ldots, K$ and $q = 1, \ldots, Q$, by WLS (i.e., (6) with $\beta = 0$). The variance $\sigma_{j,k}^2$ at each pixel j for each energy k was estimated as

$$\sigma_{j,k}^2 = \frac{1}{Q-1}\sum_{q=1}^{Q}\left(x_{j,k}^{[q]} - \bar{x}_{j,k}\right)^2, \quad \bar{x}_{j,k} = \frac{1}{Q}\sum_{q=1}^{Q} x_{j,k}^{[q]} \tag{29}$$

The weights γ_k in our method (i.e., Eq. (19) and (23)) were defined as

$$\gamma_k = \frac{\frac{1}{\sum_{j=1}^{J}\sigma_{j,k}^2}}{\sum_{k=1}^{K}\frac{1}{\sum_{j=1}^{J}\sigma_{j,k}^2}} \tag{30}$$

in order to give more weight to the less noisy images. Although these weights are in principle unknown, they can be estimated [57].

For the DL-based denoising method RED-CNN, a total of K networks need to be trained for each energy $k = 1, \ldots, K$. The training dataset for each energy k consists of a collection of $L = 360$ pairs $\left\{\left(\boldsymbol{x}_{k,l}^{\text{noisy}}, \boldsymbol{x}_{k,l}^{\text{patient}}\right)\right\}_{l=1}^{L}$, where the noisy input image $\boldsymbol{x}_{k,l}^{\text{noisy}}$ is simulated as

$$\boldsymbol{x}_{k,l}^{\text{noisy}} = \boldsymbol{x}_{k,l}^{\text{patient}} + \boldsymbol{\varepsilon}_{k,l}, \quad \boldsymbol{\varepsilon}_{k,l} \sim \mathcal{N}\left(\boldsymbol{0},\, \boldsymbol{\Sigma}_{k,l}^2\right), \tag{31}$$

where the l^{th} patient image from the first dataset $\boldsymbol{x}_{k,l}^{\text{patient}}$ serves as the corresponding target image during training, and $\boldsymbol{\Sigma}_{k,l}^2 = \operatorname{diag}\left\{\sigma_{1,k,l}^2, \ldots, \sigma_{J,k,l}^2\right\}$. The estimation of $\sigma_{j,k,l}^2$ follows the same procedure as described in (32) for each l, employing

$Q = 30$ independent measurements in a Monte-Carlo approach. The patch size in patch-based training of RED-CNN was set to 55×55, with an initial base learning rate set at 10^{-5}, gradually decreasing to 10^{-8} over the course of 10000 epochs.

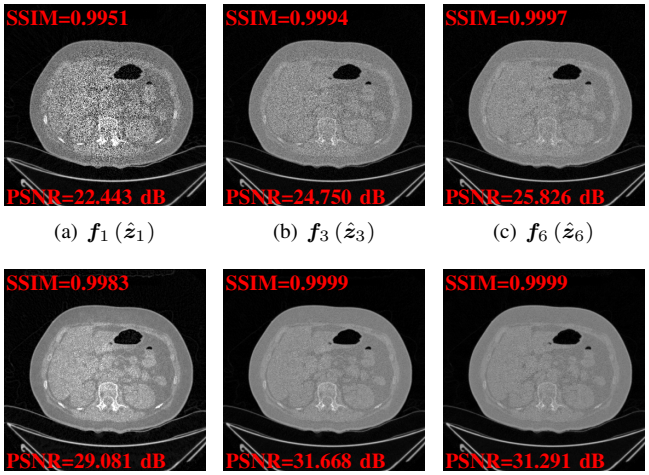
The prior image $\boldsymbol{x}^{\text{prior}}$ for DTV and ASSIST was defined as the weighted average across energy bins, i.e.,

$$\boldsymbol{x}^{\text{prior}} = \frac{1}{K}\sum_{k=1}^{K} \gamma_k \boldsymbol{x}_k^{\text{noisy}} \,. \tag{32}$$

In ASSIST, tensor units were constructed by searching 12 most similar patches of size 24×24 with a stride of 12 in the search window of 20×20. We used the same parameter setting as with the simulated XCAT phantom datasets, $\eta = 0.7$ in (12) for DTV, and for Uconnect, α in (20) was set manually to 0.4 and N in Algorithm 2 was set to 50.

2) Pre-evaluation: We proceeded with the same evaluation as in Section III-A2, but using $\boldsymbol{x}_k^{\text{noisy}}$ instead of $\hat{\boldsymbol{x}}_k$ in (25) and (26), and using $\boldsymbol{x}_k^{\text{patient}}$ as the reference for SSIM and PSNR.

Fig. 8 displays the obtained images $\boldsymbol{f}_k(\hat{\boldsymbol{z}}_{\text{all}})$ and $\boldsymbol{f}_k(\hat{\boldsymbol{z}}_k)$, $k = 1, 3, 6$ with their corresponding SSIM and PSNR. Similarly to the results of the simulated images, each $\boldsymbol{f}_k(\hat{\boldsymbol{z}}_{\text{all}})$ is less noisy and has higher SSIM and PSNR compared to the corresponding $\boldsymbol{f}_k(\hat{\boldsymbol{z}}_k)$.



(a) $\boldsymbol{f}_1(\hat{\boldsymbol{z}}_1)$ (b) $\boldsymbol{f}_3(\hat{\boldsymbol{z}}_3)$ (c) $\boldsymbol{f}_6(\hat{\boldsymbol{z}}_6)$

(d) $\boldsymbol{f}_1(\hat{\boldsymbol{z}}_{\text{all}})$ (e) $\boldsymbol{f}_3(\hat{\boldsymbol{z}}_{\text{all}})$ (f) $\boldsymbol{f}_6(\hat{\boldsymbol{z}}_{\text{all}})$

Fig. 8. $\boldsymbol{f}_1(\hat{\boldsymbol{z}}_1)$, $\boldsymbol{f}_3(\hat{\boldsymbol{z}}_3)$ and $\boldsymbol{f}_6(\hat{\boldsymbol{z}}_6)$ obtained from (26), and $\boldsymbol{f}_k(\hat{\boldsymbol{z}}_{\text{all}})$ obtained from (25) (using the noisy images $\boldsymbol{x}_k^{\text{noisy}}$).

3) Denoising Results: Fig. 9 shows the reference images $\boldsymbol{x}_k^{\text{patient}}$, the noisy images $\boldsymbol{x}_k^{\text{noisy}}$ and denoised images at 40, 80 and 140 keV with optimal β according to the PSNR shown in Fig. 10. From the noisy images $\boldsymbol{x}_k^{\text{noisy}}$, we notice that noise is predominantly concentrated in the central regions of the images for all energies, which correspond to structures with high attenuation coefficients. It is noticeable that the results of Huber, DTV and JTV suffer from severe noise although oversmoothing appears in the periphery of the images. ASSIST shows limited capability in distinguishing low-contrast structures especially at 40 keV, and its noise suppression on fine details is inadequate. RED-CNN effectively eliminates most noise at 40 keV but fails to preserve the edges of the

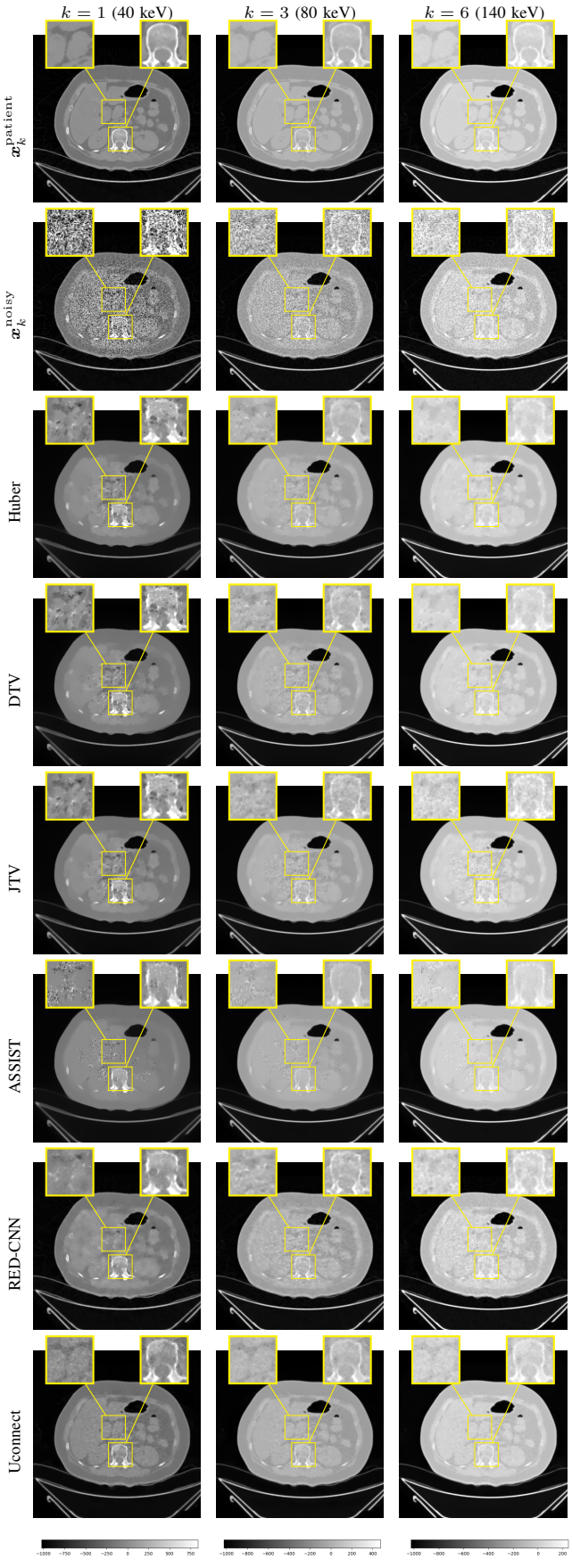


Fig. 9. GT, noisy images, and denoised images at $E_1 = 40$, $E_2 = 80$ and $E_3 = 140$ keV using different methods.

tissues. Besides, severe noise persists at other energies. On the other hand, the images denoised with Uconnect exhibit the least amount of noise and maintain the highest level of structural and edge preservation across all energies.

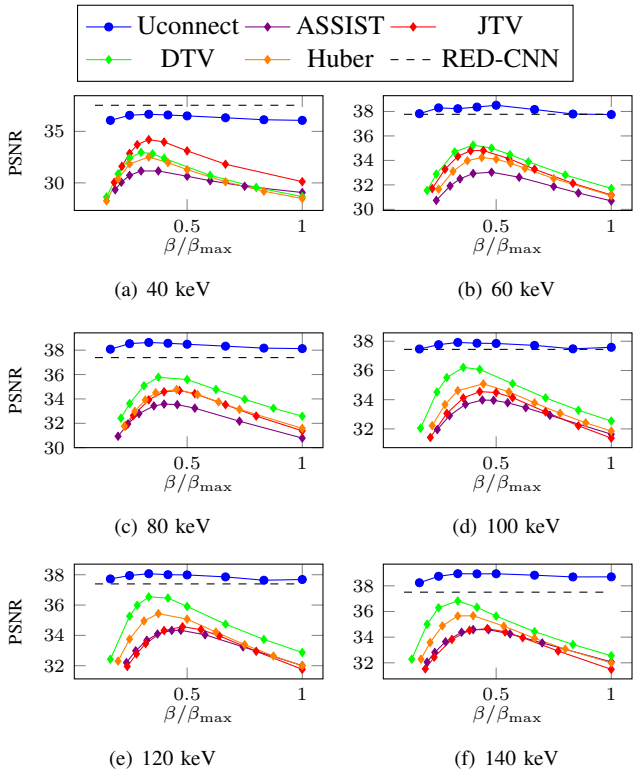


Fig. 10. PSNR values of the denoised images using different methods. As RED-CNN is independent of β due to the absence of a penalty term, we depict its evaluation values using a horizontal black dashed line.

We carried out the same evaluations, i.e., PSNR and SSIM, as in Section III-A3. Fig. 10 and Fig. 11 respectively show the PSNR and SSIM values (using $\boldsymbol{x}_k^{\text{patient}}$, $k = 1, \ldots, 6$, as reference) at each of the 6 energy bins for each method for a range of β-values. Uconnect exhibits significantly superior PSNR and SSIM values compared to Huber, DTV, JTV, and ASSIST. RED-CNN and Uconnect exhibit comparable SSIM values across all energy levels. In terms of PSNR, while RED-CNN shows a higher value than Uconnect at 40 keV, Uconnect surpasses RED-CNN at other energy levels. Taking into account the qualitative comparison illustrated in Fig. 9, it becomes evident that Uconnect outperforms all other methods considering all energies.

IV. Discussion

We proposed a novel method that takes the form of a trained penalty to solve the problem of synergistic reconstruction for spectral CT. In this paper, we used the pretrained U-Nets to "connect" images at all energies to a reference $\boldsymbol{z}$, so that the training can be supervised. We considered as reference the image at the lowest energy. Our U-Net based regularization, referred to as Uconnect, showed enhanced performance for both image reconstruction and image denoising tasks. The mappings $\{\boldsymbol{f}_k\}$ provide regularization (i) by their training, as they are trained on "clean" images (as initially suggested in

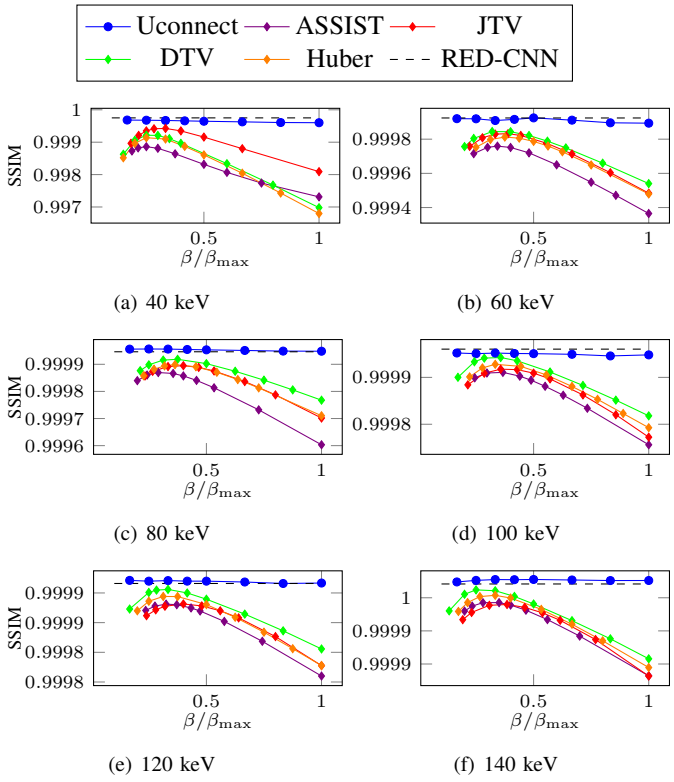


Fig. 11. SSIM values of the denoised images using different methods. As RED-CNN is independent of β due to the absence of a penalty term, we depict its evaluation values using a horizontal black dashed line.

[48]), and (ii) by sharing the information across channels, in a similar fashion as TDiL but without patch extraction. Note that our method can be generalized to a single architecture with K outputs, each of which corresponding to an energy bin, in order to further leverage the inter-energy information during training.

As pointed out in Section II-D1, the pretrained mappings $\{\boldsymbol{f}_k\}$ do not exactly reflect the physics, as $\boldsymbol{z}$ represents an attenuation image. Alternatively, models such as W-GAN or VAE, where the latent space is "trained" to encode the physical properties of the patient, should be used [46]. These models have been applied in PET/CT [58] and PET/MRI [59]. Similar models may also be used to generate material images and could be used for one-step material decomposition. However unsupervised training with a large number of channels remains challenging.

The results of the proposed approach can be potentially improved by fine-tuning all parameters, including α and $\{\gamma_k\}$ in (19), which we have not yet investigated. Additionally, we can explore other regularizers for the reference $\boldsymbol{z}$ in our method, such as by replacing Huber with TV or DTV.

An major limit to our validation is that our phantom data were based on VMIs, which correspond to an oversimplified setting. In addition, VMIs are artificially correlated, which may bias the results. Further validation should include real raw PCCT projection data and models trained from PCCT binned images.

Finally, our comparison can be expanded beyond other synergistic reconstruction methods, such as TDiL [24] and

SOUL-Net [34], a recent work using an unrolling architecture.

V. Conclusion

In this work, we used a CNN to derive a novel synergistic penalty term for spectral CT reconstruction. The proposed approach is a combination on the CNN representation technique [48] and multichannel dictionary learning [24].

The presented results demonstrate that Uconnect has the ability to reduce noise while preserving structures. Both visual inspection and quantitative measurements show significant improvements for our proposed method in synergistic spectral CT reconstruction. While there is certainly room for improvement, we believe that the proposed methodological framework provides a solid foundation for future research in this field. Further refinement of our approach may be achieved by unrolling our architecture, a development that warrants further investigation.

VI. Acknowledgments

All authors declare that they have no known conflicts of interest in terms of competing financial interests or personal relationships that could have an influence or are relevant to the work reported in this paper.

References

[1] S. S. Hsieh, S. Leng, K. Rajendran, S. Tao, and C. H. McCollough, "Photon counting CT: clinical applications and future developments," *IEEE transactions on radiation and plasma medical sciences*, vol. 5, no. 4, pp. 441–452, 2020.

[2] P. Lv, X. Z. Lin, K. Chen, and J. Gao, "Spectral CT in patients with small HCC: investigation of image quality and diagnostic accuracy," *European radiology*, vol. 22, pp. 2117–2124, 2012.

[3] Y. Long and J. A. Fessler, "Multi-material decomposition using statistical image reconstruction for spectral CT," *IEEE transactions on medical imaging*, vol. 33, no. 8, pp. 1614–1626, 2014.

[4] W. Wu, P. Chen, S. Wang, V. Vardhanabhuti, F. Liu, and H. Yu, "Image-domain material decomposition for spectral CT using a generalized dictionary learning," *IEEE transactions on radiation and plasma medical sciences*, vol. 5, no. 4, pp. 537–547, 2020.

[5] D. Morhard, C. Fink, A. Graser, M. F. Reiser, C. Becker, and T. R. Johnson, "Cervical and cranial computed tomographic angiography with automated bone removal: dual energy computed tomography versus standard computed tomography," *Investigative radiology*, vol. 44, no. 5, pp. 293–297, 2009.

[6] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[7] X.-Q. Zhang and J. Froment, "Total variation based fourier reconstruction and regularization for computer tomography," in *IEEE Nuclear Science Symposium Conference Record, 2005*, vol. 4. IEEE, 2005, pp. 2332–2336.

[8] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Physics in Medicine & Biology*, vol. 53, no. 17, p. 4777, 2008.

[9] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose x-ray CT reconstruction via dictionary learning," *IEEE transactions on medical imaging*, vol. 31, no. 9, pp. 1682–1697, 2012.

[10] S. Niu, Y. Gao, Z. Bian, J. Huang, W. Chen, G. Yu, Z. Liang, and J. Ma, "Sparse-view x-ray CT reconstruction via total generalized variation regularization," *Physics in Medicine & Biology*, vol. 59, no. 12, p. 2997, 2014.

[11] G. Sapiro and D. L. Ringach, "Anisotropic diffusion of multivalued images with applications to color filtering," *IEEE transactions on image processing*, vol. 5, no. 11, pp. 1582–1586, 1996.

[12] P. Blomgren and T. F. Chan, "Color TV: total variation methods for restoration of vector-valued images," *IEEE transactions on image processing*, vol. 7, no. 3, pp. 304–309, 1998.

[13] S. Lefkimmiatis, A. Roussos, M. Unser, and P. Maragos, "Convex generalizations of total variation based on the structure tensor with applications to inverse problems," in *International Conference on Scale Space and Variational Methods in Computer Vision*. Springer, 2013, pp. 48–60.

[14] 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.

[15] 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.

[16] H. Gao, H. Yu, S. Osher, and G. Wang, "Multi-energy CT based on a prior rank, intensity and sparsity model (PRISM)," *Inverse problems*, vol. 27, no. 11, p. 115012, 2011.

[17] O. Semerci, N. Hao, M. E. Kilmer, and E. L. Miller, "Tensor-based formulation and nuclear norm regularization for multienergy computed tomography," *IEEE Transactions on Image Processing*, vol. 23, no. 4, pp. 1678–1693, 2014.

[18] S. Niu, G. Yu, J. Ma, and J. Wang, "Nonlocal low-rank and sparse matrix decomposition for spectral CT reconstruction," *Inverse problems*, vol. 34, no. 2, p. 024003, 2018.

[19] W. Xia, W. Wu, S. Niu, F. Liu, J. Zhou, H. Yu, G. Wang, and Y. Zhang, "Spectral CT reconstruction—ASSIST: Aided by self-similarity in image-spectral tensors," *IEEE Transactions on Computational Imaging*, vol. 5, no. 3, pp. 420–436, 2019.

[20] E. Cueva, A. Meaney, S. Siltanen, and M. J. Ehrhardt, "Synergistic multi-spectral CT reconstruction with directional total variation," *Philosophical Transactions of the Royal Society A*, vol. 379, no. 2204, p. 20200198, 2021.

[21] Y. Zhang, Y. Xi, Q. Yang, W. Cong, J. Zhou, and G. Wang, "Spectral CT reconstruction with image sparsity and spectral mean," *IEEE transactions on computational imaging*, vol. 2, no. 4, pp. 510–523, 2016.

[22] L. Yao, D. Zeng, G. Chen, Y. Liao, S. Li, Y. Zhang, Y. Wang, X. Tao, S. Niu, Q. Lv *et al.*, "Multi-energy computed tomography reconstruction using a nonlocal spectral similarity model," *Physics in Medicine & Biology*, vol. 64, no. 3, p. 035018, 2019.

[23] S. Wang, W. Wu, J. Feng, F. Liu, and H. Yu, "Low-dose spectral CT reconstruction based on image-gradient l0-norm and adaptive spectral PICCS," *Physics in Medicine & Biology*, vol. 65, no. 24, p. 245005, 2020.

[24] 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, 2016.

[25] W. Wu, Y. Zhang, Q. Wang, F. Liu, P. Chen, and H. Yu, "Low-dose spectral CT reconstruction using image gradient ℓ_0–norm and tensor dictionary," *Applied Mathematical Modelling*, vol. 63, pp. 538–557, 2018.

[26] X. Li, X. Sun, Y. Zhang, J. Pan, and P. Chen, "Tensor dictionary learning with an enhanced sparsity constraint for sparse-view spectral CT reconstruction," in *Photonics*, vol. 9, no. 1. MDPI, 2022, p. 35.

[27] A. Perelli, S. A. Garcia, A. Bousse, J.-P. Tasu, N. Efthimiadis, and D. Visvikis, "Multi-channel convolutional analysis operator learning for dual-energy CT reconstruction," *Physics in Medicine & Biology*, vol. 67, no. 6, p. 065001, 2022.

[28] A. Bousse, V. S. S. Kandarpa, S. Rit, A. Perelli, M. Li, G. Wang, J. Zhou, and G. Wang, "Systematic review on learning-based spectral CT," *IEEE Transactions on Radiation and Plasma Medical Sciences*, 2023. [Online]. Available: https://arxiv.org/abs/2304.07588

[29] G. Wang, J. C. Ye, and B. De Man, "Deep learning for tomographic image reconstruction," *Nature Machine Intelligence*, vol. 2, no. 12, pp. 737–748, 2020.

[30] E. Ahishakiye, M. Bastiaan Van Gijzen, J. Tumwiine, R. Wario, and J. Obungoloch, "A survey on deep learning in medical image reconstruction," *Intelligent Medicine*, vol. 1, no. 03, pp. 118–127, 2021.

[31] W. Xia, H. Shan, G. Wang, and Y. Zhang, "Physics-/model-based and data-driven methods for low-dose computed tomography: A survey," *IEEE Signal Processing Magazine*, vol. 40, no. 2, pp. 89–100, 2023.

[32] 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.

[33] W. Wu, D. Hu, C. Niu, L. V. Broeke, A. P. Butler, P. Cao, J. Atlas, A. Chernoglazov, V. Vardhanabhuti, and G. Wang, "Deep learning based spectral CT imaging," *Neural Networks*, vol. 144, pp. 342–358, 2021.

[34] X. Chen, W. Xia, Z. Yang, H. Chen, Y. Liu, J. Zhou, and Y. Zhang, "SOUL-net: A sparse and low-rank unrolling network for spectral CT image reconstruction," *arXiv preprint arXiv:2207.12639*, 2022.

[35] Z. Wang, A. Bousse, F. Vermet, N. J. Pinton, J. Froment, B. Vedel, J.-P. Tasu, and D. Visvikis, "Synergistic multi-energy reconstruction with a deep penalty "connecting the energies"," in *IEEE Nuclear Science Symposium Medical Imaging Conference and Room Temperature Semiconductor Conference*, 2022. [Online]. Available: https://hal.ehesp.fr/IBNM/hal-03955092v1

[36] I. A. Elbakri and J. A. Fessler, "Statistical image reconstruction for polyenergetic X-ray computed tomography," *IEEE transactions on medical imaging*, vol. 21, no. 2, pp. 89–99, 2002.

[37] P. J. Huber, "Robust statistics," in *International encyclopedia of statistical science*. Springer, 2011, pp. 1248–1251.

[38] H. Gao, J.-F. Cai, Z. Shen, and H. Zhao, "Robust principal component analysis-based four-dimensional computed tomography," *Physics in Medicine & Biology*, vol. 56, no. 11, p. 3181, 2011.

[39] L. Bungert, D. A. Coomes, M. J. Ehrhardt, J. Rasch, R. Reisenhofer, and C.-B. Schönlieb, "Blind image fusion for hyperspectral imaging with the directional total variation," *Inverse Problems*, vol. 34, no. 4, p. 044003, 2018.

[40] G.-H. Chen, J. Tang, and S. Leng, "Prior image constrained compressed sensing (PICCS): a method to accurately reconstruct dynamic CT images from highly undersampled projection data sets," *Medical physics*, vol. 35, no. 2, pp. 660–663, 2008.

[41] Z. Yu, S. Leng, Z. Li, and C. H. McCollough, "Spectral prior image constrained compressed sensing (spectral PICCS) for photon-counting computed tomography," *Physics in Medicine & Biology*, vol. 61, no. 18, p. 6707, 2016.

[42] V. P. Sudarshan, Z. Chen, and S. P. Awate, "Joint PET+MRI patch-based dictionary for bayesian random field PET reconstruction," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 338–346.

[43] V. P. Sudarshan, G. F. Egan, Z. Chen, and S. P. Awate, "Joint PET-MRI image reconstruction using a patch-based joint-dictionary prior," *Medical image analysis*, vol. 62, p. 101669, 2020.

[44] P. Song, L. Weizman, J. F. Mota, Y. C. Eldar, and M. R. Rodrigues, "Coupled dictionary learning for multi-contrast MRI reconstruction," *IEEE transactions on medical imaging*, vol. 39, no. 3, pp. 621–633, 2019.

[45] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*. PMLR, 2017, pp. 214–223.

[46] M. Duff, N. D. Campbell, and M. J. Ehrhardt, "Regularising inverse problems with generative machine learning models," *arXiv preprint arXiv:2107.11191*, 2021.

[47] 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.

[48] K. Gong, J. Guan, K. Kim, X. Zhang, J. Yang, Y. Seo, G. El Fakhri, J. Qi, and Q. Li, "Iterative PET image reconstruction using convolutional neural network representation," *IEEE transactions on medical imaging*, vol. 38, no. 3, pp. 675–685, 2018.

[49] C.-K. Liu, C.-C. Liu, C.-H. Yang, and H.-M. Huang, "Generation of brain dual-energy CT from single-energy CT using deep learning," *Journal of Digital Imaging*, vol. 34, no. 1, pp. 149–161, 2021.

[50] C. Zhu, R. H. Byrd, P. Lu, and J. Nocedal, "Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization," *ACM Transactions on mathematical software (TOMS)*, vol. 23, no. 4, pp. 550–560, 1997.

[51] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of mathematical imaging and vision*, vol. 40, pp. 120–145, 2011.

[52] K. Lange, D. R. Hunter, and I. Yang, "Optimization transfer using surrogate objective functions," *Journal of computational and graphical statistics*, vol. 9, no. 1, pp. 1–20, 2000.

[53] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[54] E. Y. Sidky, J. H. Jørgelognsen, and X. Pan, "Convex optimization problem prototyping for image reconstruction in computed tomography with the Chambolle–Pock algorithm," *Physics in Medicine & Biology*, vol. 57, no. 10, p. 3065, 2012.

[55] W. P. Segars, M. Mahesh, T. J. Beck, E. C. Frey, and B. M. Tsui, "Realistic CT simulation using the 4D XCAT phantom," *Medical physics*, vol. 35, no. 8, pp. 3800–3808, 2008.

[56] W. van Aarle, W. J. Palenstijn, J. Cant, E. Janssens, F. Bleichrodt, A. Dabravolski, J. D. Beenhouwer, K. J. Batenburg, and J. Sijbers, "Fast and flexible x-ray tomography using the astra toolbox," *Opt. Express*, vol. 24, no. 22, pp. 25 129–25 147, Oct 2016. [Online]. Available: http://opg.optica.org/oe/abstract.cfm?URI=oe-24-22-25129

[57] J. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): applications to tomography," *IEEE Transactions on Image Processing*, vol. 5, no. 3, pp. 493–506, 1996.

[58] N. J. Pinton, A. Bousse, Z. Wang, C. Cheze-Le-Rest, V. Maxim, C. Comtat, F. Sureau, and D. Visvikis, "Synergistic PET/CT reconstruction using a joint generative model," in *International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (accepted)*, 2023.

[59] V. Gautier, C. Comtat, F. Sureau, A. Bousse, L. Friot-Giroux, V. Maxim, and B. Sixou, "VAE constrained MR guided PET reconstruction," in *International Conference on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine (accepted)*, 2023.