Full-Spectrum-Knowledge-Aware Unsupervised Network for Photon-counting CT Imaging

Danyang Li[a], Zheng Duan[a], Dong Zeng[a], Zhaoying Bian[a], and Jianhua Ma[a]

[a]School of Biomedical Engineering, Southern Medical University, Guangzhou 510515, China

ABSTRACT

Deep learning (DL) based methods have been widely adopted in computed tomography (CT) field. And they also show a great potential in photon-counting CT (PCCT) imaging field. They usually require a large quantity of paired data to train networks. However, it is time-consuming and expensive to collect such large-scale PCCT dataset. In addition, lots of energy-integrating detector (EID) data are not yet included in the DL-based PCCT reconstruction network training. In this work, to address the issue of limited PCCT data and take advantage of labeled EID data, we propose a novel unsupervised full-spectrum-knowledge-aware DL-based network (FSANet), which contains supervised and unsupervised networks, to produce high-quality PCCT images. Specifically, the supervised network is trained based on paired EID dataset and serves as the prior knowledge to regularize the unsupervised PCCT network training. Moreover, a data-fidelity term for characterizing the PCCT image characteristics is constructed as a self-supervised term. Finally, we train the PCCT network with the prior knowledge and self-supervised terms following an unsupervised learning strategy. Numerical studies on synthesized clinical data are conducted to validate and evaluate the performance of the presented FSANet method, qualitatively and quantitatively. The experimental results demonstrate that presented FSANet method significantly improves the PCCT image quality in the case of limited photon counts.

Keywords: Photon-counting CT, full-spectrum-knowledge-aware, unsupervised learning.

1. INTRODUCTION

Comparing with the conventional computed tomography (CT), photon-counting CT (PCCT) can obtain multiple measurements of the scanned object at multi-energy bins and provide abundant energy-dependent material-specific information. Due to the energy discrimination capability, PCCT can effectively improve contrast-to-noise ratio, increase the dose efficiency and reduce electronic noise.[1,2]

However, the collected PCCT measurements in the narrow energy bins are corrupted by serious quantum noise,[3] and the PCCT image quality degrades obviously due to the limited photons. To solve this problem, many statistical iteration reconstruction (SIR) methods have been proposed in the past decades. The main idea of the SIR is to construct a reconstruction model with data fidelity and regularization terms, where the first term incorporates the statistical property of X-ray photons and the second term provides the prior information of the desired PCCT images. For example, Rigie *et al.* introduced a total nuclear variation regularization to leverage similar gradient information and improve the image quality.[4] Kim *et al.* developed a patch-based low-rank regularization to maintain the image structures and reduce noise.[5] Semerci *et al.* combined a tensor nuclear norm and a total variation regularization to suppress noise.[6] Zhang *et al.* proposed to deliver inner spectrum correlation information and constructed a tensor-based dictionary learning strategy.[7] Niu *et al.* considered the self-similarity of the spectral CT images and proposed a non-local low-rank and sparse matrix decomposition method.[3] Wu *et al.* proposed to encourage the similarity of spectral CT images by utilizing a cube-based tensor regularization.[8] Recently, Zeng *et al.* proposed a full-spectrum-knowledge-aware tensor by imposing the global correlation, piecewise smooth and latent full-spectrum properties of PCCT images.[9] These methods have been shown great potential in preserving image details and suppressing noise. However, there remain some challenges in practice. First, the SIR methods usually utilize fan-beam geometry and the computation costs will be a

---

burden for cone-beam geometry. Second, SIR methods are sensitive for the hyper-parameters, and appropriate parameters selection is needed for different clinical applications.

Recently, deep learning (DL) technology has been widely adopted in CT imaging field. In spectral CT imaging field, Lu *et al.* utilized DL-based method for material decomposition.[10] Fang *et al.* proposed to remove ring artifacts for PCCT data by using DL-based method.[11] Wu *et al.* employed a DL-based method for reconstructing PCCT images.[12] It is shown that the DL-based methods achieved competing results compared with the SIR methods. However, the current DL-based methods need a large quantity of paired data (i.e., noisy and high-quality data) to obtain a desired model by supervised training strategy. Moreover, collecting large-scale spectral CT data is time-consuming, and the PCCT data in clinics is hard to be obtained. By the way, a number of energy-integrating detector (EID) data, which is easily to be obtained, is not yet included in training the DL-based method for PCCT imaging.
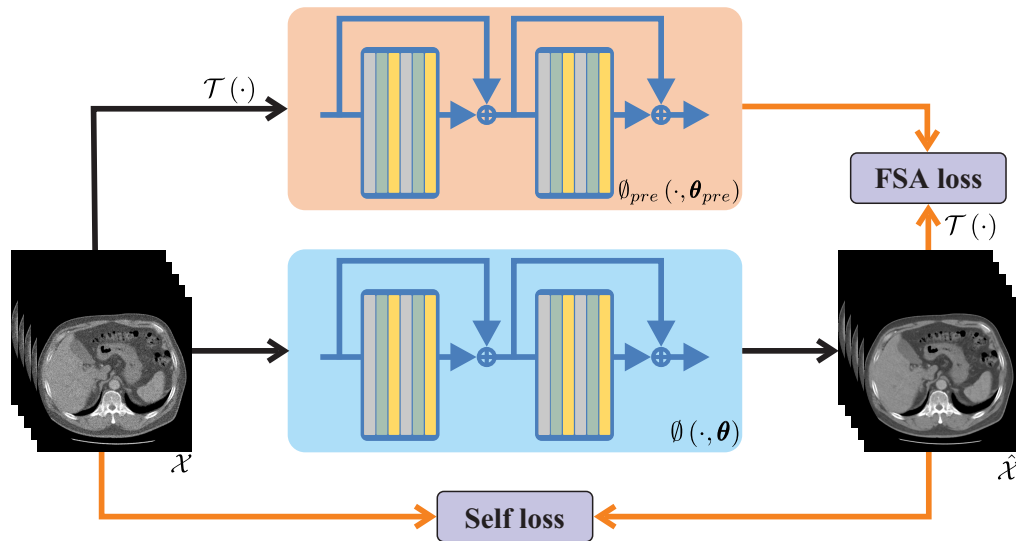

Figure 1. Illustration of the presented FSANet method for PCCT image recovery. The pipeline with black arrows denote the data flow of the PCCT images. The pipeline with orange arrows denote calculations of loss functions. The supervised and unsupervised networks have the same architecture. It should be noted that the optimization for the parameters of the unsupervised network follows an unsupervised strategy, and only noisy PCCT images are involved during its training period.

Therefore, we present an unsupervised DL-based method in the image domain by utilizing the prior information of the paired EID dataset (i.e., low-dose images/high-quality ones). Specifically, we first initialize supervised and unsupervised networks for EID and PCCT images, respectively. Then, the supervised network is trained on well paired EID dataset and serves as the prior knowledge to regularize the unsupervised PCCT network training. Moreover, a data-fidelity term for characterizing the PCCT image characteristics is constructed as the self-supervised loss. Finally, with the prior knowledge and self-supervised terms, we can train the network for PCCT images following an unsupervised learning strategy. We call the presented DL-based model as full-spectrum-knowledge-aware DL-based network, shorten as "FSANet". We evaluate the presented FSANet and other reconstruction methods on synthesized clinical data. Experimental results demonstrate that the presented FSANet outperforms the competing methods in terms of qualitative and quantitative metrics.

2. METHODS

Considering the spatial and energy dimensions of the spectral data, the PCCT imaging model can be expressed as follows:

$$\mathcal{Y} = \mathcal{A}\mathcal{X}^* + \varepsilon, \tag{1}$$

where $\mathcal{Y} = \{y_n, n \leq N\}$ and $\mathcal{X}^* = \{x_n^*, n \leq N\}$ are the measurements and desired PCCT images along the multi-energy bins, and N is the total number of the energy bins. $\mathcal{A}$ is the linear projection operator for PCCT imaging, and ε denotes the noise in the projection domain.

The images reconstructed from $\mathcal{Y}$ suffer from noise and artifacts. To improve image quality, DL-based methods are fed with the PCCT images and produce the denoised ones. It can be expressed as follows:

$$\hat{\mathcal{X}} = \emptyset_{DL}(\mathcal{X}, \boldsymbol{\theta}_{DL}), \tag{2}$$

where $\emptyset_{DL}$ represents the network of the DL-based method with parameters $\boldsymbol{\theta}_{DL}$. $\hat{\mathcal{X}}$ are the estimated PCCT images by the network. $\mathcal{X} = (\mathcal{A}^{\mathcal{T}}\mathcal{A})^{-1}\mathcal{A}^T\mathcal{Y}$ are the network inputs which are directly reconstructed by the filtered back projection (FBP) algorithm.[13] Followed by the supervised strategy, the network parameters $\boldsymbol{\theta}_{DL}$ are optimized by minimizing the loss function between the target images $\mathcal{X}^{target}$ and $\hat{\mathcal{X}}$, which can be expressed as follows:

$$\hat{\boldsymbol{\theta}}_{DL} = \arg\min_{\boldsymbol{\theta}_{DL}} \mathcal{L}\left(\hat{\mathcal{X}}, \mathcal{X}^{target}\right), \tag{3}$$

where $\mathcal{L}$ is the user defined loss function and $\mathcal{X}^{target}$ are the FBP-reconstructed images from the high-dose measurements.

In order to utilize the prior information of the network pre-trained on paired EID dataset, we present a full-spectrum-knowledge-aware (FSA) loss function, as follows:

$$\mathcal{L}_{FSA}\left(\mathcal{X}\right) = \|\mathcal{T}\left(\emptyset\left(\mathcal{X}, \boldsymbol{\theta}\right)\right) - \emptyset_{pre}\left(\mathcal{T}\left(\mathcal{X}\right), \boldsymbol{\theta}_{pre}\right)\|_1, \tag{4}$$

where $\emptyset_{pre}$ denotes a pre-trained network for EID images with parameters $\boldsymbol{\theta}_{pre}$. $\emptyset\left(\cdot, \boldsymbol{\theta}\right)$ is the network for PCCT images with parameters $\boldsymbol{\theta}$. $\mathcal{T}\left(\cdot\right)$ is an operator to transform the PCCT images to the EID images. $\|\cdot\|_1$ is L_1 norm. Moreover, we also construct a self-supervised loss to encourage the data fidelity of the PCCT images, as follows:

$$\mathcal{L}_{Self}\left(\mathcal{X}\right) = \|\emptyset\left(\mathcal{X}, \boldsymbol{\theta}\right) - \mathcal{X}\|_1. \tag{5}$$

In summary, the total loss function of the presented model is expressed as follows:

$$\mathcal{L}\left(\mathcal{X}\right) = \mathcal{L}_{Self}\left(\mathcal{X}\right) + \alpha\mathcal{L}_{FSA}\left(\mathcal{X}\right), \tag{6}$$

where α is a hyper-parameter of the loss function. It can be seen that the presented model engages an unsupervised learning strategy where only noisy PCCT images is involved in the training stage. Fig. 1 illustrates the presented FSANet method for PCCT image recovery. Finally, we utilize Adam[14] to optimize the parameters in the network.

3. RESULTS

3.1 Implementation Details

The network $\emptyset_{pre}$ are trained on paired EID dataset collected from conventional CT in local hospital with the WGAN loss.[15] The PCCT image datasets involved in this study are simulated from EID images by segmenting the soft and bone tissues. The noisy data is generated by adding Poisson noise in the projection domain. The X-ray spectrum with 120 kVp tube voltage and 1.6 mm Al filtration is generated by SPEKTR toolbox.[16] Five energy bins are determined by the five thresholds: 25, 50, 60, 70 and 85 keV. The simulation imaging parameters are listed as follows: 816 parallel X-ray beams and 1160 projection views over 360° are adopted, and source-to-detector and source-to-center distances are 1040.0 and 570.0 mm, respectively. The network trained by supervised strategy, called "Supervised Net", serves as the upper-bound performance of the presented FSANet. Moreover, the filtered back projection (FBP) method with a ramp filter and a tensor-based dictionary learning regularization (TDL) method are the compared methods for the presented method.

In experiments, we simulate 3000 cases to establish the whole dataset, and randomly select 2000, 500 and 500 cases for training, validation and testing datasets, respectively. In training period, the learning rate, batch size and training epoch are $1e^{-4}$, 6 and 2000, respectively. A modified residual network[17] is selected as the backbone network for the supervised and unsupervised networks of the presented FSANet method. Both networks are implemented in Python with PyTorch package on a NVIDIA Tesla K40c GPU.
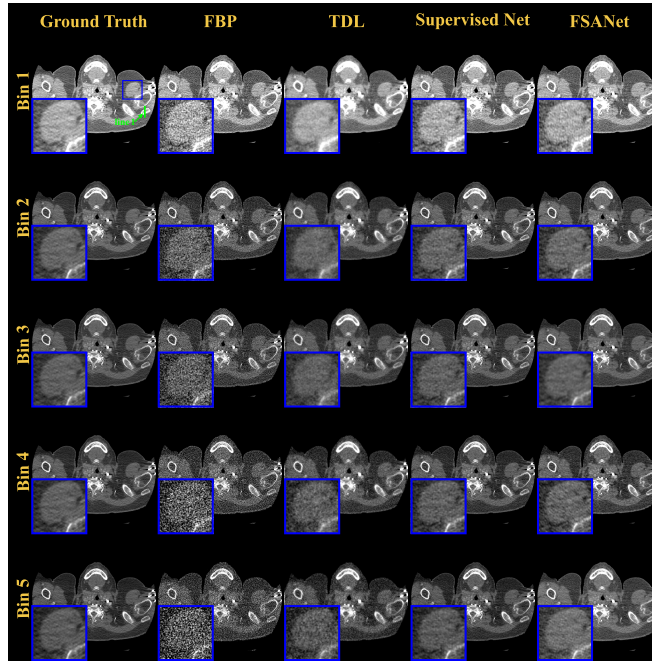
Figure 2. Results of the presented and compared methods on Case 1. The display windows from Bin 1 to 5 are [0.005, 0.008], [0.0028, 0.005], [0.002, 0.004], [0.0015, 0.0035] and [0.0015,0.0028] mm^{-1}, respectively. Zoomed ROIs indicated by the blue boxes are displayed for better visualization.

3.2 Qualitative Analysis

Fig. 2 illustrates the visual comparisons of the presented and compared methods on Case 1. The images at normal dose are chosen as the ground truth. It can be observed that FBP algorithm suffers from noise. TDL effectively removes the noise-induced artifacts and improves the image quality, but losses the image resolution. On the contrary, the presented FSANet produces more remarkable results closing to the Supervised Net and the ground truth in terms of noise reduction and structure preservation. Moreover, zoomed in regions-of-interest (ROIs) by the blue boxes in Fig. 2 is selected for better visual inspection. It can be seen that TDL smooths the structure edges, but the Supervised Net and presented FSANet methods maintain the image details.

Fig. 3 shows the results of different methods on Case 2. Similar to Case 1, FBP induces noise to the images, TDL is prone to produce blurry results, and the presented FSANet avoids over smoothing and preserves structure details. The zoomed in ROIs indicated by the red boxes in Fig. 3 also demonstrate the advanced performance of the presented FSANet method. Fig. 4(a) and (b) show the profiles indicated by the green and orange lines in the Fig. 2 and Fig. 3, respectively. From the results, we can observe that the presented FSANet produces the closet results to the ground truth compared with the FBP and TDL methods.

Table 1. Quantitative measurements on the reconstruction results on the testing dataset from the different methods.

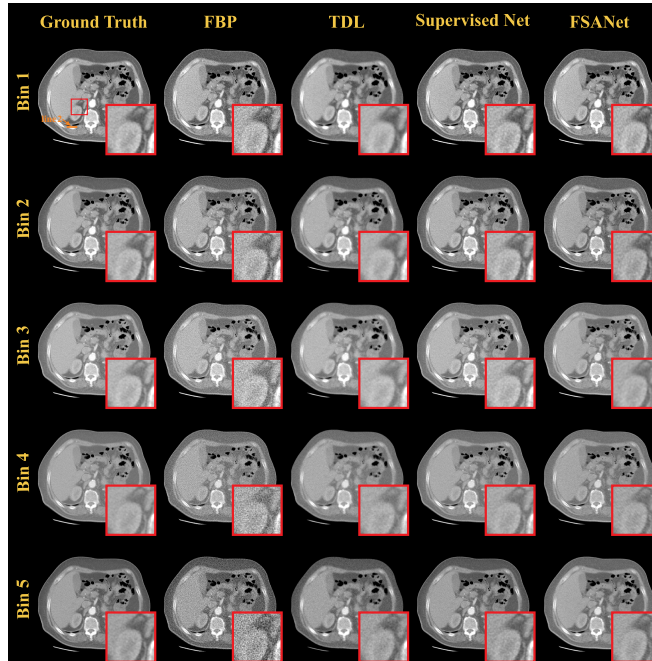|	PSNR (dB)	RMSE ($\times 10^{-5}$)
FBP	28.36 ± 2.88	12.5 ± 0.73
TDL	35.37 ± 5.70	7.62 ± 4.324
Supervised Net	$\mathbf{38.14 \pm 2.60}$	$\mathbf{3.78 \pm 0.433}$
GMM-3DTV	$\mathbf{37.44 \pm 2.25}$	$\mathbf{4.46 \pm 0.788}$

Figure 3. Results of the presented and compared methods on Case 2. The display windows from Bin 1 to 5 are [0.0015, 0.0025], [0.0009, 0.0018], [0.0007, 0.0015], [0.0006, 0.0012] and [0.0005, 0.0010] mm^{-1}, respectively. Zoomed ROIs indicated by the red boxes are displayed for better visualization.
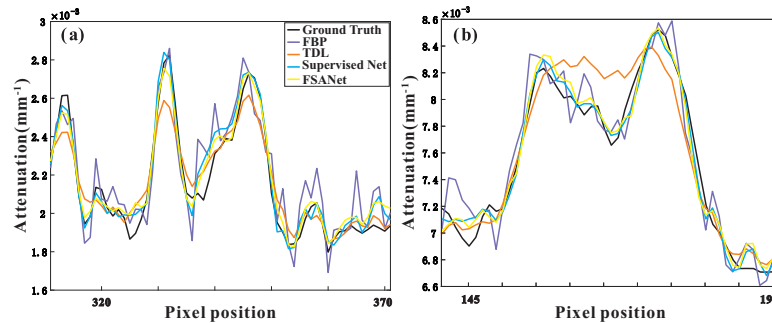

Figure 4. Profiles of the green and orange lines in Bin 1 for Case 1 and 2, respectively. (a) are the profiles of the results from different methods on Case 1, respectively. (b) are the profiles of the results from different methods on Case 2, respectively.

3.3 Quantitative Analysis

In this study, peak signal-to-noise (PSNR) and root-mean-square-error (RMSE) are utilized to quantify the performances of different methods. Tab. 1 lists the quantitative measurements of the results from different methods on the whole testing dataset. From the results, it can be seen that the presented FSANet achieves better results among all metrics, compared with FBP and TDL methods. Therefore, the qualitative and quantitative results demonstrate that the presented FSANet method achieves superior results to the FBP and TDL methods.

4. DISCUSSION AND CONCLUSION

DL-based methods have been shown promising performance in conventional CT imaging, and it has also inspired the application in PCCT imaging. However, most of them are supervise-based and need a large quantity of paired training dataset, which is hard to be obtained for PCCT. To address this intrinsic limitation, in this work, we presented a DL-based PCCT denoising method with an unsupervised learning strategy, called "FSANet".

Specifically, we first trained a denoising network on paired EID dataset and served it as a prior for PCCT images. Then, we constructed this prior network as a training loss to regularize the PCCT network learning in an unsupervised manner. Moreover, a self-supervised loss of the noisy PCCT images is introduced to promote data-fidelity of the PCCT images. Finally, with the mentioned two loss terms, we obtained the presented FSANet method. Simulation experiments demonstrated the feasibility and effectiveness of the presented FSANet method. In the future, clinical patient studies would be involved to further demonstrate the denosing performance of the presented FSANet method.

REFERENCES

[1] M. J. Willemink, M. Persson, A. Pourmorteza, N. J. Pelc, and D. Fleischmann, "Photon-counting CT: Technical principles and clinical prospects," *Radiology*, vol. 289, no. 2, pp. 293-312, 2018.

[2] K. Taguchi, J. S. Iwanczyk, "Vision 20/20: Single photon counting x-ray detectors in medical imaging," *Medical Physics*, vol. 40, no. 10, p. 100901, 2013.

[3] 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.

[4] D. S. Rigie and P. J. La Rivire, "Joint reconstruction of multi-channel, spectral CT data via constrained total nuclear variation minimization," *Physical in Medical & Biology*, vol. 60, no. 5, pp. 1741-1762, 2015.

[5] K. Kim and *et al.*, "Sparse-view spectral CT reconstruction using spectral patch-based low-rank penalty," *IEEE Transactions on Medical Imaging*, vol. 34, no. 3, pp. 748-760, 2015.

[6] 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.

[7] 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.

[8] W. Wu, F. Liu, Y. Zhang, Q. Wang, and H. Yu, "Non-local low-rank cube-based tensor factorization for spectral CT reconstruction," *IEEE Transactions on Medical Imaging*, vol. 38, no. 4, pp. 1079-1093, 2019.

[9] D. Zeng, L. Yao, Y. Ge, S. Li, Q. Xie, H. Zhang, and *et al.*, "Full-spectrum-knowledge-aware tensor model for energy-resolved CT iterative reconstruction," *IEEE Transactions on Medical Imaging*, vol. 39, pp. 2831-2843, 2020.

[10] Y. Lu, M. Kowarschik, X. Huang, Y. Xia, J. Choi, S. Chen, and *et al.* "A learning-based material decomposition pipeline for multi-energy X-ray imaging," *Medical physics*, vol. 46, pp. 689-703, 2019.

[11] W. Fang, L. Li, and Z. Chen, "Removing ring artefacts for photon-counting detectors using neural networks in different domains," *IEEE Access*, vol. 8, pp. 42447-42457, 2020.

[12] W. Wu, D. Hu, C. Niu, L. Broeke, A. Butler, P. Cao, and *et al.*, "Deep learning based spectral CT imaging," *Neural Networks*, vol. 144, pp. 342-358, 2021.

[13] G. L. Zeng, "Medical image reconstruction: a conceptual tutorial," Springer, 2010.

[14] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint*, arXiv: 1412.6980, 2014

[15] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN." [Online]. Available: https://arxiv.org/abs/1701.07875

[16] J. Punnoose, J. Xu, A. Sisniega, W. Zbijewski, and J. H. Siewerdsen, "Technical note: Spektr 3.0-A computational tool for X-ray spectrum modeling and analysis," *Medical Physics*, vol. 43, no. 8Part1, pp. 4711-4717, 2016.

[17] K. He, X. Zhang, S. Ren, *et al.*, "Deep residual learning for image recognition," *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.