Improved low-count quantitative PET reconstruction with a variational neural network

Hongki Lim, Student Member, IEEE, Il Yong Chun, Member, IEEE, Yuni K. Dewaraja, Member, IEEE, and Jeffrey A. Fessler, Fellow, IEEE

Abstract—Image reconstruction in low-count PET is particularly challenging because gammas from natural radioactivity in Lu-based crystals cause high random fractions that lower the measurement signal-to-noise-ratio (SNR). In model-based image reconstruction (MBIR), using more iterations of an unregularized method may increase the noise, so incorporating regularization into the image reconstruction is desirable to control the noise. New regularization methods based on learned convolutional operators are emerging in MBIR. We modify the architecture of a variational neural network, BCD-Net, for PET MBIR, and demonstrate the efficacy of the trained BCD-Net using XCAT phantom data that simulates the low true coincidence counts with high random fractions typical for Y-90 PET patient imaging after Y-90 microsphere radioembolization. Numerical results show that the proposed BCD-Net significantly improves PET reconstruction performance compared to MBIR methods using non-trained regularizers, total variation (TV) and non-local means (NLM), and a non-MBIR method using a single forward pass deep neural network, U-Net. BCD-Net improved activity recovery for a hot sphere significantly and reduced noise, whereas non-trained regularizers had a trade-off between noise and quantification. BCD-Net improved CNR and RMSE by 43.4% (85.7%) and 12.9% (29.1%) compared to TV (NLM) regularized MBIR. Moreover, whereas the image reconstruction results show that the non-MBIR U-Net over-fits the training data, BCD-Net successfully generalizes to data that differs from training data. Improvements were also demonstrated for the clinically relevant phantom measurement data where we used training and testing datasets having very different activity distribution and count-level.

Index Terms—Variational neural network, Regularized model-based image reconstruction, Low-count quantitative PET, Y-90

I. INTRODUCTION

Image reconstruction in low-count PET is particularly challenging because dominant gammas from natural radioactivity in Lu-based crystals cause low signal-to-noise-ratio (SNR) measurements [1]. To accurately reconstruct images in low-count PET, regularized model-based image reconstruction (MBIR) solves the following variational problem consisting of 1) a data fidelity $f(x)$ that models the physical PET imaging model, and 2) a regularization term $R(x)$ that penalizes image roughness and controls noise [2]:

$$\hat{x} = \arg\min_{x \geq 0} f(x) + R(x)$$

$$= \arg\min_{x \geq 0} 1^T (Ax + \bar{r}) - y^T \log(Ax + \bar{r}) + R(x). \quad (1)$$

This work was supported by grant R01 EB022075, awarded by the National Institute of Biomedical Imaging and Bioengineering, National Institute of Health, U.S. Department of Health and Human Services.

Hongki Lim, Il Yong Chun, and Jeffrey A. Fessler are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48109 USA (email: [hongki, fessler, iychun]@umich.edu).

Yuni K. Dewaraja is with the Department of Radiology, University of Michigan, Ann Arbor, MI 48109 USA (yuni@med.umich.edu).

This work has been submitted to the IEEE for possible publication. Copyright may be transferred without notice, after which this version may no longer be accessible.

Here, $f(x)$ is the Poisson negative log-likelihood for measurement $y$ and estimated measurement $\hat{y}(x) = Ax + \bar{r}$, the matrix $A$ denotes the system model, and $\bar{r}$ denotes the mean background events such as scatter and random coincidences. Recently, applying learned regularizers to $R(x)$ is emerging for MBIR [3].

While there is much ongoing research on machine learning or deep-learning techniques applied to CT [4]–[8] and MRI [9]–[13] reconstruction problems, only a few studies have applied these techniques to PET. Most past PET studies use deep learning in image space without exploiting the physical imaging model in (1). For example, [14] applied a deep neural network (NN) mapping between reconstructed PET images with normal dose and reduced dose and [15] applied a multilayer perceptron mapping between reconstructed images using maximum a posteriori algorithm and a reference (true) image, however, their framework uses the acquisition data only to form the initial image. Therefore, the reconstruction quality depends greatly on the training dataset and information from atypical imaging situations and variable noise levels (that are not part of the training set) may not be recovered well. A recent work [16] trained a NN to reconstruct an image directly from sinogram, however, its framework has similar limitation with [14]. Recently [17], [18] proposed a PET MBIR framework using a deep-learning based regularizer. Our proposed MBIR framework, BCD-Net, also uses a regularizer that penalizes differences between the unknown image and “denoised” images given by regression neural network in a recurrent manner. In particular, whereas [17], [18] trained only a single image denoising NN, the proposed method is a recurrent framework that is composed of multiple trained NNs. This recurrent framework enables NNs in the later stages to learn how to recover fine details. Our proposed BCD-Net is also distinct from [17], [18] in that denoising NNs are derived by variational (optimization) formulation with a mathematical motivation (whereas, for the trained regularizer, [17], [18] brought U-Net [19] and DnCNN [20] developed for medical image segmentation and general Gaussian denoising) and characterized by less parameters, thereby avoiding over-fitting and generalizing well to unseen data especially when training samples are limited (see Section IV).

Variational NNs [8]–[11], [21], [22] are a broad family of methods that originate from unrolling algorithm for solving an optimization problem and BCD-Net [23] is a specific example of a variational NN. BCD-Net is constructed by unfolding block coordinate descent (BCD) MBIR algorithm using “learned” convolutional analysis operators [24], [25], and significantly improved image recovery accuracy in extreme imaging applications, e.g., highly undersampled MRI.
denoising low-SNR images, etc. A preliminary version of this paper was presented at the 2018 Nuclear Science Symposium and Medical Imaging Conference [26]. We significantly extended this work by applying our proposed method to the measurement data with newly developed techniques. We also added detailed analysis on our proposed method as well as comparisons to related works.

To show the efficacy of our proposed method BCD-Net in low-count PET imaging, we performed both digital phantom simulation and experimental measurement studies with activity distributions and count-rates that are relevant to clinical Y-90 PET imaging after liver radioembolization. Novel therapeutic applications have sparked growing interest in quantitative imaging of Y-90, an almost pure beta emitter that is widely used in internal radionuclide therapy. In addition to the FDA approved Y-90 microsphere radioembolization and Y-90 ibritumomab radioimmunotherapy, there are 50 active clinical trials for Y-90 labeled therapies (www.clinicaltrials.gov). However, the lack of gamma photons complicates imaging of Y-90; it involves SPECT via bremsstrahlung photons produced by the betas [27] or PET via a very low abundance positron via bremsstrahlung photons produced by the betas [27]. For Y-90 labeled therapies (www.clinicaltrials.gov). However, in the setting of Y-90 radioembolization and provides details in the simulation and measurement data. Section IV discusses what imaging factor affects generalization performance of BCD-Net most, and compares between BCD-Net and non-trained regularizers and single forward pass denoising NN.

Our proposed BCD-Net applies to PET imaging in general, particularly in other imaging situations that also have low counts. Using shorter scan times and lower tracer activity in diagnostic PET has cost benefits and reduces radiation exposure, but at the expense of reduced counts that makes traditional iterative reconstruction challenging.

Section II develops our proposed BCD-Net architecture for PET MBIR. Section II also explains the simulation studies in the setting of Y-90 radioembolization and provides details on how we perform the physical phantom measurement. Section III presents how the different reconstruction methods perform with the simulation and measurement data. Section IV discusses what imaging factor affects generalization performance of BCD-Net most, and compares between BCD-Net using shallow convolutional autoencoding denoiser and BCD-Net using deep U-Net denoiser. Section V concludes with future works.

II. METHODS

This section presents the problem formulation of the BCD-Net and gives a detailed derivation on how we get the final form of BCD-Net. We also provide several techniques for BCD-Net that we specifically devised for PET measurement data where each measurement has different count-level (and noise-level). Then we review the related works that we compare with BCD-Net such as MBIR methods using conventional non-trained regularizers and single forward pass denoising (non-MBIR) NN. This section also describes the simulation setting and details on measurement data and what evaluation metrics are used to assess the efficacy of each reconstruction algorithm.

A. BCD algorithm for MBIR using “learned” convolutional regularization

Conventional PET regularizers penalize differences between neighboring pixels [29]. That approach is equivalent to assuming that convolving the image with the [1-1] finite difference filters with different directions produces a sparse output. Using such “hand-crafted” filters is unlikely to be the best approach. A recent trend is to use training data to learn filters $c_k$ that produce sparse outputs when convolved with images of interest [24], [25]. Such learned filters can be used to define a regularizer that prefers images having sparse outputs, as follows [30]:

$$R(x) = \min_{\{z_k\}} \beta \left( \sum_{k=1}^{K} \|c_k \ast x - z_k\|^2_2 + \alpha_k \|z_k\|_1 \right).$$

where $\beta$ is regularization parameter, $\{c_k \in R^R : k = 1, \ldots, K\}$ is a set of convolutional filters, $\{z_k \in R^{n_p} : k = 1, \ldots, K\}$ is a set of sparse codes, $\{\alpha_k \in R : k = 1, \ldots, K\}$ is a set of thresholding parameters controlling the sparsity of $\{z_k\}$, $n_p$ is the number of image voxels, and $R$ and $K$ is the size and number of learned filters, respectively. BCD-Net is inspired by this type of “learned” regularizer. Ultimately, we hope that the learned regularizer can better separate true signal from noisy components.

BCD algorithm solves (1) with regularizer (2) by alternatingly updating $\{z_k\}$ and $x$:

$$\{z_k^{(n+1)}\} = \arg\min_{\{z_k\}} \|c_k \ast x(n) - z_k\|^2_2 + \alpha_k \|z_k\|_1$$

$$x^{(n+1)} = \arg\min_{x} f(x) + \beta \left( \sum_{k=1}^{K} \|c_k \ast x - z_k^{(n+1)}\|^2_2 \right).$$

where $T(\cdot)$ is the element-wise soft thresholding operator: $T(t, q) := \text{sign}(t) \max(|t| - q, 0)$.

Assuming that learned filters $\{c_k\}$ satisfy the tight-frame condition, $\sum_{k=1}^{K} \|c_k \ast x\|_2^2 = \|x\|_2^2 \forall x$ [24], we rewrite the updates in (3)-(4) as follows:

$$u^{(n+1)} = \sum_{k=1}^{K} \tilde{c}_k \ast \left( T(c_k \ast x(n), \alpha_k) \right)$$

$$x^{(n+1)} = \arg\min_{x} f(x) + \beta \|x - u^{(n+1)}\|_2^2,$$

where $\tilde{c}_k$ denotes a rotated version of $c_k$. For efficient image reconstruction [9] in PET, we use EM-surrogate of Poisson log-likelihood function [31]:

$$f(x) + \beta \|x - u^{(n+1)}\|_2^2 = \sum_{i=1}^{n_x} \{ [Ax]_i + \bar{r}_i - y_i \log([Ax]_i + \bar{r}_i) + \beta \sum_{j=1}^{n_p} (x_j - u^{(n+1)}_j)^2 \}

\leq \sum_{j=1}^{n_p} \{ -e_j(x(n)) \langle x(n) \rangle \log(x_j) + \alpha_j x_j + \beta(x_j - u^{(n+1)}_j)^2 \}

= \sum_{j=1}^{n_p} Q_j(x_j)$$

A preliminary version of this paper was presented at the 2018 Nuclear Science Symposium and Medical Imaging Conference [26]. We significantly extended this work by applying our proposed method to the measurement data with newly developed techniques. We also added detailed analysis on our proposed method as well as comparisons to related works.
where \( c_j(x^{(n)}) = \sum_{i=1}^{n_d} a_{ij} \frac{y_i}{y_i(x^{(n)})} \) and \( n_d \) is the number of rays. Equating \( \frac{\partial Q_j(x_j)}{\partial x_j} \) to zero is equivalent to finding the root of the following quadratic formula:

\[
2 \beta x_j^2 + (a_j - 2 \beta u_j^{(n+1)}) x_j - \sum_{i=1}^{n_d} a_{ij} \frac{y_i}{y_i(x^{(n)})} = 0,
\]

where \( a_{ij} \) denotes an element of the system model at \( i \)th row and \( j \)th column, \( a_j = \sum_{i=1}^{n_d} a_{ij} \) and \( n' \) denotes \( n \)'th iteration in (6).

B. BCD-Net for PET MBIR and its training

To further denoising capability, we extend the convolutional autoencoder in [5] [23] by replacing \( \{c_k\} \) with separate decoding filters \( \{d_k\} \). We define the following updates for each layer:

\[
u^{(n+1)} = \sum_{k=1}^{K} d_k^{(n+1)} \ast \left( T\left(c_k^{(n+1)} \ast x^{(n)}, \alpha_k^{(n)} \right) \right)
\]

\[x^{(n+1)} = \arg \min_{x} f(x) + \beta \| x - u^{(n+1)} \|_2^2.
\]

One can further extend the convolutional autoencoder in (7) to a general regression NN, e.g., deep U-Net [19]. See Section IV.

We train the image denoising module in the form of (7) – specifically, consisting of encoding and decoding filters, and thresholding values \( \{c_k^{(n+1)}, d_k^{(n+1)}, \alpha_k^{(n)} \} \) – that “best” map between high-quality images (e.g., true images if available) and noisy images in the sense of mean squared error:

\[
\arg \min_{\{c_k, \{d_k\}, \{\alpha_k\}}} \sum_{l=1}^{L} \left\| x_{\text{true},l} - \sum_{k=1}^{K} d_k \ast \left( T\left(c_k \ast x_l^{(n)}, \alpha_k \right) \right) \right\|_2^2,
\]

where \( \{x_{\text{true}, l} \in \mathbb{R}^{n_p} : l = 1, \ldots, L \} \) is a set of true images and \( \{x_l^{(n)} \in \mathbb{R}^{n_p} : l = 1, \ldots, L \} \) is a set of images estimated by image reconstruction module in the \( n \)'th layer.

Fig. 1 shows the corresponding BCD-Net architecture. We name the \( u \) and \( x \) updates in (7)-(8) as following two modules: 1) image denoising module and 2) image reconstruction module. A layer refers to one loop of denoising and reconstruction module.

C. BCD-Net for measurement data in PET

1) Normalization and scaling scheme: Different PET images can have very different intensity values due to variations in scan time and activity, and it is important for trained methods to be able to generalize to a wide range of count levels. Towards this end, we implemented normalization and scaling techniques in BCD-Net. [18] extended [17] by implementing “local linear fitting” to ensure that the denoising NN output has similar intensity as the input patch from the current estimated image. Our approach is different in that we normalize and scale the image with a global approach, not a patch-based approach.

In particular, we modify the architecture in (7)-(8) as:

\[
u^{(n+1)} = \sum_{k=1}^{K} d_k^{(n+1)} \ast \left( T\left(c_k^{(n+1)} \ast g_1(x^{(n)}) \right) \right)
\]

\[x^{(n+1)} = \arg \min_{x} f(x) + \beta \| x - g_2(u^{(n+1)}) \|_2^2.
\]

The normalization function \( g_1(\cdot) \) is defined by \( g_1(v) := \frac{1}{\sum_j y_j} \) to satisfy \( 1 \leq \sum_j y_j \leq 1 \), and the scaling function \( g_2(\cdot) \) is defined by \( g_2(v) := s \nu \) with \( s = \arg \min_{x} f(s \nu) \). We solve the optimization problem over \( s \) using Newton’s method:

\[s^{(n+1)} = s^{(n)} - \frac{\nabla_s f(s^{(n)} \nu)}{\nabla^2 f(s^{(n)} \nu)} = s^{(n)} - \sum_{i=1}^{n_d} \frac{[Av]_i}{(Av)_i + ri}.
\]

We also apply this normalization technique when training the convolutional filters and thresholding values:

\[
\arg \min_{\{c_k, \{d_k\}, \{\alpha_k\}}} \sum_{l=1}^{L} \left\| g_1(x_{\text{true},l}) - \sum_{k=1}^{K} d_k \ast \left( T\left(c_k \ast g_1(x_l^{(n)}) \right) \right) \right\|_2^2.
\]

2) Adaptive regularization parameter scheme: The best regularization parameter value can also vary greatly between scans, depending on the count level. Therefore, instead of choosing one specific value for regularization parameter, we set the \( \beta \) value based on evaluation on current gradients of data-fidelity term and regularization term:

\[
\beta^{(n)} = \frac{\| \nabla \nu f(x^{(n)}) \|_2}{\| \nabla \nu R(x^{(n)}) \|_2} \cdot c
\]

\[
= \frac{\| a_j - e_j(x^{(n)}) \|_2}{\| 2(x^{(n+1)} - g_2(u^{(n+1)}) \|_2} \cdot c,
\]

where \( c \) is a constant specifying how we balance between the data-fidelity term and regularization term and \( n' \) denotes \( n \)'th iteration in (11).
D. Conventional MBIR methods: Non-trained regularizers

We compared the proposed BCD-Net with two MBIR methods using standard non-trained regularizers.

1) Total-variation (TV): TV regularization penalizes the sum of absolute value of differences between adjacent voxels:

$$ R(x) = \beta \|Cx\|_1, $$

where $C$ is finite differencing matrix. Recent work \cite{32} applied Primal-Dual Hybrid Gradient (PDHG) \cite{33} for PET MBIR using TV regularization and demonstrated that PDHG-TV is superior than clinical reconstruction (e.g., OS-EM) for low-count datasets in terms of several image quality evaluation metrics such as contrast recovery and variability.

2) Non-local means (NLM): NLM regularization minimizes the differences between nearby patches in image:

$$ R(x) = \beta \sum_{i,j \in S_t} p \left( \frac{\|N_ix - N_jx\|^2}{\sigma_f^2N_f} \right), $$

where $p(t)$ is a potential function of a scalar variable $t$, $S_t$ is the search neighborhood around the $i$th voxel, and $N_ix$ is a patch extraction operator at the $i$th voxel. We implemented Fair potential function for $p(t)$:

$$ p(t) = \sigma_f^2 \left( \sqrt{\frac{t}{\sigma_f^2N_f}} + \log \left( 1 + \frac{t}{\sigma_f^2N_f} \right) \right), $$

where $\sigma_f$ is a design parameter and $N_f$ is the number of voxels in the patch $N_ix$. Unlike conventional local filters that assume similarity between only adjacent voxels, NLM filters can average image intensities over distant voxels. As in \cite{34}, we used ADMM to accelerate algorithmic convergence with an adaptive penalty parameter selection method \cite{35}.

E. Conventional non-MBIR method: Denoising deep U-Net

Many related works \cite{5}, \cite{14}, \cite{36} use single image denoising (deep) NN (e.g., U-Net) as a post-reconstruction processing. We implemented non-recurrent (single forward pass) 3-D version of U-Net to compare with the proposed BCD-Net. The ‘encoder’ part of U-Net consists of multiple sets of 1) max pooling layer, 2) $3 \times 3 \times 3$ convolutional layer, 3) batch normalization (BN) layer, 4) ReLU layer and the ‘decoder’ part of U-Net consists of multiple sets of 1) upsampling with trilinear interpolation \cite{17}, 2) $3 \times 3 \times 3$ convolutional layer, 3) batch normalization (BN) layer, 4) ReLU layer. We used ReLU layer as the last step to enforce the non-negativity constraint on image \cite{17}.

F. Experimental setup: Digital phantom simulation and experimental measurement

1) Y-90 PET/CT XCAT simulations: We used XCAT \cite{37} phantom (Fig. 2) to simulate Y-90 PET following radioembolization. We set the image size to $128 \times 128 \times 100$ with a voxel size $4.0 \times 4.0 \times 4.0$ (mm$^3$) and chose 100 slices ranging from lung to liver. To simulate the extremely low count scan, typical for Y-90 PET, we set total true coincidences and random fraction based on numbers from patient PET imaging performed after radioembolization \cite{38}. To test the generalization capability of trained NNs, we changed all imaging factors between training and testing dataset. Here, imaging factors include activity distribution (shape and size of tumor and liver background, concentration ratio between hot and warm region) and count-level (total true coincidences and random fraction). Fig. 2 and Table I provide details on how we changed the testing dataset from the training dataset.

2) Y90 PET/CT physical phantom measurements and patient scan: For training NNs (BCD-Net and U-Net), we used measurement with sphere phantom (Fig. 3) where six ‘hot’ spheres (2.4, 8, 16, 30 and 113 mL, 0.5 MBq/ml) are placed in a ‘warm’ background (0.057 MBq/ml) with total activity of 0.65 GBq. The phantom was scanned for 40 (3 acquisitions) - 80 (1 acquisition) minutes on a Siemens Biograph mCT PET/CT. For testing NNs and other reconstruction algorithms, we used an anthropomorphic liver/lung torso phantom (Fig. 3) with total activity and distribution that is clinically realistic for imaging following radioembolization with Y-90 microspheres: 5% lung shunt, 1.17 MBq/mL in liver, 3 hepatic lesions (4 and 16 mL spheres, 29 mL ovoid) of 6.6 MBq/ml. The phantom with total activity of 1.9 GBq was scanned for 30 minutes on a Siemens Biograph mCT PET/CT. Fig 3 and Table II provide details on the count-level and activity distribution difference between training (sphere phantom) and testing (liver phantom) dataset. We also tested NNs with an actual Y-90 patient scan and the count-level information is provided in Table III.

We acquired all measurement data with time of flight information. The measurement data size is $400 \times 168 \times 621 \times 13$. The last dimension of measurement indicates the number of time bin. The reconstructed image size is $400 \times 400 \times 112$ with a voxel size $2.04 \times 2.04 \times 2.03$ (mm$^3$). To reconstruct the image with measurement data, we used SIEMENS time of flight system model (A in \cite{11}) along with manufacturer given attenuation/normalization correction, PSF modelling, and randoms/scatters estimation.

G. Evaluation metrics

For XCAT phantom simulation, we evaluated each reconstruction with activity recovery (AR) (VOI: hot sphere,
background), contrast recovery (CR) (VOI: cold spot), contrast to noise ratio (CNR), and root mean squared error (RMSE). For physical phantom measurement, we used AR, coefficient of variation (CV) \([29]\) (VOI: hot spheres), and CNR averaged over multiple hot spheres. For patient measurement, we used field of view (FOV) activity bias since the total activity in FOV is known (equal to the injected activity because the microspheres are trapped) while the activity distribution is unknown:

\[
\begin{align*}
\text{AR} & \text{ (%) } = \frac{\text{Estimated } C_{\text{VOI}}}{\text{True } C_{\text{VOI}}} \times 100(\%) \\
\text{CR} & \text{ (%) } = \left(1 - \frac{\text{Estimated } C_{\text{VOI}}}{\text{True } C_{\text{BKG}}} \right) \times 100(\%) \\
\text{CNR} & = \frac{C_{\text{Lesion}} - C_{\text{BKG}}}{\text{STD}_{\text{BKG}}} \\
\text{RMSE} & \text{ (%) } = \sqrt{\frac{\sum_j (x_{\text{true}}[j] - \hat{x}[j])^2}{J_{\text{FOV}}}} \times 100(\%) \\
\text{CV} & \text{ (%) } = \frac{\text{STD}_{\text{VOI}}}{C_{\text{VOI}}} \times 100(\%) \\
\text{FOV bias} & \text{ (%) } = \frac{\sum_j \hat{x}[j] - x_{\text{true}}[j]}{\sum_j x_{\text{true}}[j]} \times 100(\%)
\end{align*}
\]

where \(C_{\text{VOI}}\) is mean counts in the volume of interest (VOI), \(STD_{\text{BKG}}\) is standard deviation between voxel values in uniform background liver, and \(J_{\text{FOV}}\) is the total number of voxels in field of view (FOV).

### III. Results

**A. Experimental setup: Training BCD-Net and U-Net**

1) **BCD-Net**: We trained 3D convolutional filters and thresholding values in each layer with a stochastic gradient descent method using PyTorch \([40]\) deep-learning library. We trained a thirty-layer BCD-Net where each layer has \(K = 128\) sets of thresholding values and convolutional encoding/decoding filters for the simulation experiment and twenty-layer BCD-Net with \(K = 64\) (due to GPU memory capacity) for the measurement experiment. We set the size of each filter as \(3 \times 3 \times 3\) \((R = 3^3)\), and the initial thresholding values by sorting the initial estimate of image and getting a 10% largest value of sorted initial image. We used Adam optimization method \([41]\) to train the NN with learning rate of \(10^4\) for encoding filters, \(10^1\) for decoding filters, and \(10^6\) for thresholding values. We applied the learning rate decay scheme (e.g., decreasing the learning rate as a factor of 0.9 per 20 epochs). Due to the large size of 3D input, we set the batch size as 1. We used 300 epochs to train the denoising NN at each layer.

With the XCAT simulation data, we trained BCD-Net using five pairs \((L = 5)\) of 3D true image and 3D realizations (1 true image, 5 realizations). We generated multiple realizations to train the denoising NN to deal with the Poisson noise.

With measurement data, we used the sphere phantom data to train the BCD-Net. We used 4 pairs \((L = 4)\) of 3D true image and 4D measurements (1 true image, 4 scans) for training. We set \(c = 0.007\) in \([13]\). In measurement experiment, due to the memory capacity of GPUs, we downsample \(x^{(n)} \in \mathbb{R}^{400 \times 400 \times 112}\) with a factor of 2 before feeding into the denoising module and upsample \(u^{(n)} \in \mathbb{R}^{200 \times 200 \times 112}\) before feeding into the image reconstruction module. The final output image \(\hat{x} \in \mathbb{R}^{400 \times 400 \times 112}\) is from the reconstruction module.

2) **U-Net**: For training U-Net, we used identical training dataset that we used for training BCD-Net. We also used Adam optimization method to train the NN with learning rate of \(10^{-2}\). We scaled the image generated by U-Net with \(g_2(\cdot)\) described in \([12]\). We set the number of convolutional filter channels of first layer of encoder as 48 for simulation dataset and 32 for measurement dataset.

### B. Experimental setup: Reconstruction and testing trained NN

With simulation dataset, we compare the proposed method BCD-Net to the standard EM (1 subset), TV-based MBIR with PDHG algorithm (PDHG-TV), NLM-based MBIR with ADMM algorithm (ADMM-NLM), and single forward pass denoising U-Net. For BCD-Net and U-Net, we used 20 EM algorithm iterations to get the initial image. We report averaged evaluation metrics over 3 realizations. With measurement dataset, we compare the proposed method BCD-Net to the standard EM (1 subset), OSEM (1 iteration and 21 subsets) with post-filtering which is identical to reconstruction setting used in our clinic, and denoising U-Net. We used 10 EM algorithm iterations to get the initial image for BCD-Net and U-Net.

1) **Reconstruction (testing) results on simulation data**: We used 50 iterations for EM and 200 iterations for PDHG-TV, ADMM-NLM, and 5 iterations for the reconstruction module \([5]\) at each layer of BCD-Net. When reporting evaluation results, we selected the iteration number for EM to obtain the highest CNR. For each regularizer, we finely tuned the regularization parameter \(\beta\) (within range \([2^{-15}, 2^{15}]\)) to achieve highest CNR and lowest RMSE values. For NLM, we additionally tuned the window and search sizes.

2) **Reconstruction (testing) results on measurement data**: We tested on the liver-torso phantom measurement data and patient measurement data with trained BCD-Net (and U-Net) using sphere phantom data. We used 2 iterations for the reconstruction module \([11]\) at each layer of BCD-Net.

### C. Results: Reconstruction (testing) on simulation data

The proposed variational NN, BCD-Net, significantly improves overall reconstruction performance over the other non-trained regularization methods. See Table \(IV\) and Fig. \(2\). Table \(IV\) shows that BCD-Net achieves best results in most
evaluate metrics. In particular, BCD-Net improves CNR and RMSE by 43.4%/8.7%/8.3% and 12.9%/29.1%/26.9% compared to PDHG-TV/ADMM-NLM/U-Net. Fig. 2 shows that reconstructed image using BCD-Net is closest to the true image whereas PDHG-TV and U-Net exceedingly blur in cold region and ADMM-NLM is noisy in uniform region. The profiles in Fig. 3 also illustrate that PDHG-TV and U-Net underestimate the hot region, whereas profile of BCD-Net is very close to true value in both hot and cold region.

D. Results: Reconstruction (testing) on measurement data

1) Phantom study: Similar to the result of simulation data, BCD-Net improves overall reconstruction performance over the other reconstruction methods. See Fig. 4 and Table V.

When reporting evaluation metrics and visualizing the images, we selected the iteration number for EM and BCD-Net to obtain the highest CNR. Table V shows that BCD-Net improves AR/CV/CNR at multiple spheres by 9.3%/37.5%/22.3% compared to OSEM. Even though U-Net improves CNR more than BCD-Net, image generated by U-Net looks unnatural and the shape of liver is round-shape like sphere phantom showing that U-Net is over-fitting to the training data.

2) Patient study: Because of the unknown true activity distribution, we quantitatively evaluated each reconstruction method with FOV activity bias. BCD-Net improves the FOV bias by 44.9% and 67.0% compared to OSEM and U-Net. U-Net generates very artificial image that resembles the sphere phantom activity distribution. See Fig. 5.

We report CN and CV value using average of three hot lesions.

TABLE IV

| METHOD | AR (Lesion) | AR (Liver) | CR (Cold Spot) | CNR | RMSE |
|--------|-------------|------------|----------------|-----|------|
| EM     | 77.6        | 99.5       | 49.4           | 1   | 7.0  |
| PDHG-TV| 73.8        | 90.9       | 64.5           | 6.9 | 6.5  |
| ADMM-NLM| 77.6       | 88.2       | 68.0           | 5.3 | 7.9  |
| U-Net  | 72.5        | 85.1       | 55.7           | 9.1 | 7.7  |
| BCD-Net| 83.7        | 90.4       | 70.2           | 9.9 | 5.6  |

TABLE V

| METHOD | AR (29ml) | AR (16ml) | AR (4ml) | AR (Liver) | CV | CNR |
|--------|-----------|-----------|----------|------------|----|-----|
| EM     | 74.9      | 66.3      | 49.1     | 88.5       | 35.2| 6.1 |
| OSEM w/ post-filtering | 83.9 | 80.1 | 47.5 | 90.4 | 40.9 | 5.6 |
| U-Net  | 72.9      | 62.1      | 35.9     | 46.5       | 12.7| 7.2 |
| BCD-Net| 97.1      | 80.3      | 54.0     | 87.2       | 25.6| 6.8 |

We report CN and CV value using average of three hot lesions.
**IV. Discussion**

In this study we showed the efficacy of trained regularization method BCD-Net on both qualitative and quantitative Y-90 PET/CT imaging and compared between conventional non-trained regularizers and other training-based denoising method. The proposed regularization method uses convolutional filters to lift estimated signals and thresholding operations to remove unwanted signals. Moreover, the recurrent framework of BCD-Net enables one to train the filters and thresholding values to deal with the different image roughness.

**TABLE VI**

| Patient Measurement Data Results: Evaluation Metrics for Reconstructed Images from Different Reconstruction Methods |
|---------------------------------------------------------------|
| **FOV bias**        | **OSEM w/ filter** | **U-Net** | **BCD-Net** |
|---------------------|--------------------|-----------|-------------|
| OSEM w/ filter      | -19.2              | -33.0     | -10.6       |
| U-Net               |                     |           |             |
| BCD-Net             |                     |           |             |
TABLE VII
IMPACT OF IMAGING VARIABLE ON GENERALIZATION CAPABILITY OF
BCD-Net.

| Changed imaging variable | Training | Testing | RMSE | Drop (%) |
|--------------------------|----------|---------|------|----------|
| Identical                | -        | -       | 4.74 | -        |
| Shape and size           | See Fig 4| 5.49    | 15.9 |          |
| Concentration ratio      | 9:1      | 4:1     | 5.55 | 17.1     |
| Concentration ratio      | 1:7:1    | 4:1     | 5.81 | 22.5     |
| Truths Count-level       | $2 \times 10^5$ | $5 \times 10^5$ | 5.01 | 5.7      |
| Truths Count-level       | $11 \times 10^5$ | $9 \times 10^5$ | 5.71 | 20.5     |

at its each layer. We experimentally demonstrate its capability of generalization with simulation and measurement data. In the XCAT PET/CT simulation with activity distributions and count-rates mimicking Y-90 PET imaging, AR in the volume of interest was significantly underestimated with standard reconstruction and other MBIR methods using non-trained regularization, however approached the true activity with the proposed regularization method. Improvements were also demonstrated for the measurement data where we used training and testing datasets having very different activity distribution and count-level.

We investigated if the recurrent BCD-Net combined with U-Net denoisers [17] (denoising module in (7) is replaced by U-Net) is better than the proposed BCD-Net using convolutional autoencoder denoiser (7). At each BCD-Net layer, U-Net in this experiment has about 22.6 million trainable parameters, whereas the denoiser in (7) has only 7040 ($3 \times 3 \times 3 \times 128 \times 2 + 128$) trainable parameters. In Fig. 6 RMSE value of U-Net-based BCD-Net is lower than that of the proposed BCD-Net using (7) in training dataset, however, it is opposite in testing data. This result demonstrates that the shallow convolutional autoencoders in (7) has better generalization capability over the deep U-Net denoisers, when the training samples are limited. Fig. 6 also compares between reconstructed images from BCD-Net with the convolutional autoencoders (7) and that using U-Nets, and it shows that the visual quality of (7)-based BCD-Net is better than that of U-Net-based BCD-Net in that U-Net denoisers generates artificial cold region around the hot region. This over-fitting issue with U-Net based regularizer could be moderated if one trained the NN with more diverse and large dataset; however, medical imaging community often finds it hard to collect large enough datasets compared to the computer vision community [42], therefore it is important for a reconstruction method to ensure the generalization to unseen dataset.

We tested which imaging variable impacts most on the generalization performance of the proposed BCD-Net. Table VII shows that how BCD-Net performs when training and testing data had same activity distribution and count-level (only difference is Poisson noise) and how the performance of BCD-Net is degraded when each imaging variable is changed between training and testing dataset. We changed one of three factors (shape and size of tumor and liver, concentration ratio, count-level) in training dataset compared to testing dataset. The result shows that generalization performance of the proposed BCD-Net depends largely on all of imaging
BCD-Net with larger size of filter exceedingly blurs image RMSE compared to smaller number and size of filters; however, it does not decrease testing performance. Plots show that the proposed BCD-Net achieves lower training RMSE when using larger number filters on performance. Plots show that the proposed BCD-Net, to the unseen dataset, when training dataset is small. The proposed BCD-Net achieves significant qualitative and quantitative improvements over the conventional MBIR methods using “hand-crafted” non-trained regularizers, TV and NLM. In particular, these conventional MBIR methods have a trade-off between noise and recovery accuracy, whereas the proposed BCD-Net improves AR for hot regions and reduced noise at the same time. Visual comparison of reconstructed images also shows that the proposed BCD-Net significantly improves PET image reconstruction performance compared to MBIR methods using non-trained regularizers and non-MBIR denoising U-Net.

Future work includes investigating performance of BCD-Net trained with end-to-end training principles and adaptive selection of trainable parameter numbers depending on the size of training dataset.

VI. ACKNOWLEDGMENT

We acknowledge Se Young Chun (UNIST) for providing NLM regularization codes. We also acknowledge Maurizio Conti and Deepak Bharkhada (SIEMENS Healthcare Molecular Imaging) for providing the forward/back projector for TOF measurement data. This work was supported by NIH-NIBIB grant R01EB022075.
