

Physics-Guided Deep Scatter Estimation by Weak Supervision for Quantitative SPECT

Hanvit Kim¹, Member, IEEE, Zongyu Li², Graduate Student Member, IEEE, Jiye Son,
 Jeffrey A. Fessler³, Fellow, IEEE, Yuni K. Dewaraja⁴, Member, IEEE,
 and Se Young Chun⁵, Member, IEEE

Abstract—Accurate scatter estimation is important in quantitative SPECT for improving image contrast and accuracy. With a large number of photon histories, Monte-Carlo (MC) simulation can yield accurate scatter estimation, but is computationally expensive. Recent deep learning-based approaches can yield accurate scatter estimates quickly, yet full MC simulation is still required to generate scatter estimates as ground truth labels for all training data. Here we propose a physics-guided weakly supervised training framework for fast and accurate scatter estimation in quantitative SPECT by using a 100× shorter MC simulation as weak labels and enhancing them with deep neural networks. Our weakly supervised approach also allows quick fine-tuning of the trained network to any new test data for further improved performance with an additional short MC simulation (weak label) for patient-specific scatter modelling. Our method was trained with 18 XCAT phantoms with diverse anatomies / activities and then was evaluated on 6 XCAT phantoms, 4 realistic virtual patient phantoms, 1 torso phantom and 3 clinical scans from 2 patients for

¹⁷⁷Lu SPECT with single / dual photopeaks (113, 208 keV). Our proposed weakly supervised method yielded comparable performance to the supervised counterpart in phantom experiments, but with significantly reduced computation in labeling. Our proposed method with patient-specific fine-tuning achieved more accurate scatter estimates than the supervised method in clinical scans. Our method with physics-guided weak supervision enables accurate deep scatter estimation in quantitative SPECT, while requiring much lower computation in labeling, enabling patient-specific fine-tuning capability in testing.

Index Terms—Quantitative SPECT, scatter estimation, weakly supervised learning, physics-based deep learning.

I. INTRODUCTION

QUANTITATIVE single photon emission computed tomography (SPECT) is important for clinical applications including dosimetry-guided treatment planning [1], [2], [3] for optimization of therapies such as peptide receptor radionuclide therapy (PRRT) for the treatment of neuroendocrine tumors with ¹⁷⁷Lu-DOTATATE [4], or radioembolization with ⁹⁰Y microspheres for the treatment of liver malignancies [5]. Scattered photons detected within an energy window degrade SPECT image quality by reducing the contrast in the reconstructed images and introducing additional uncertainties on activity distributions [6]. Scatter estimation has been an active research area for several decades for improving the quality of SPECT images [7], [8], [9], [10], [11]. Scatter estimation methods can be divided into three categories [7]: 1) multiple energy window-based methods [12], 2) Monte-Carlo simulation-based methods [2], [13], [14], and 3) deep learning-based methods [15].

Energy window-based estimation is used widely [16] due to its straightforward implementation with computational efficiency and wide applicability to various radionuclides. One drawback is the noise amplification due to relatively low abundance of counts in narrow scatter windows [8]. Filtering can reduce noise, but may introduce bias and other artifacts in the estimated scatter [8], [17]. Moreover, the photons recorded in the lower energy window may not originate from the same locations in the photopeak window or may have scattered several times, causing inaccurate scatter estimation [13].

The Monte-Carlo (MC) simulation-based scatter estimation has achieved accurate estimation performance [18], [19], [20], [21], [22], [23]. Since it simulates the transport of photons

Manuscript received 28 March 2023; accepted 19 April 2023. Date of publication 27 April 2023; date of current version 2 October 2023. This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Republic of Korea, under Grant NRF-2017R1D1A1B05035810; in part by the Korea Health Technology Research and Development Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health and Welfare, Republic of Korea, under Grant HI18C0316; in part by the Creative-Pioneering Researchers Program through Seoul National University; and in part by the National Institute of Health, USA, under Grant R01CA240706 and Grant 2 R01 EB022075. (Corresponding author: Se Young Chun.)

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the University of Michigan Institutional Review Board (IRB).

Hanvit Kim was with the Department of Electrical Engineering, Ulsan National Institute of Science and Technology (UNIST), Ulsan 44919, South Korea. He is now with the Digital Biomedical Research Division, Electronics and Telecommunications Research Institute, Daejeon 34129, South Korea (e-mail: coreavit@etri.re.kr).

Zongyu Li and Jeffrey A. Fessler are with the Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI 48104 USA (e-mail: zonyul@umich.edu; fessler@umich.edu).

Jiye Son was with the School of Electrical and Computer Engineering (ECE), UNIST, Ulsan 44919, South Korea. She is now with the Interdisciplinary Program for Bioengineering, Seoul National University (SNU), Seoul 08826, South Korea (e-mail: sonjy9712@snu.ac.kr).

Yuni K. Dewaraja is with the Department of Radiology, University of Michigan, Ann Arbor, MI 48104 USA (e-mail: yuni@umich.edu).

Se Young Chun is with the Department of ECE, INMC and IPAI, SNU, Seoul 08826, South Korea (e-mail: sychun@snu.ac.kr).

Digital Object Identifier 10.1109/TMI.2023.3270868

in patients [19], one must track enough photon histories to generate scatter estimates with low noise, hence accurate MC scatter modeling requires heavy computation. A few studies have attempted to mitigate the accuracy-computational efficiency trade-off [21], [24]. GPU acceleration can reduce MC simulation time [25], but unfortunately most current MC simulation codes for SPECT imaging systems including SIMIND do not support GPU. Moreover, GPU may not be the best choice for accurate MC simulations since they usually require double (FP64) precision computations while GPU is optimized for single (FP32) or half (FP16) precision computations. The effective scatter source estimation (ESSE) method [26] uses MC simulations to estimate a kernel to represent the scatter projection and has been used for ^{177}Lu SPECT, yielding promising results over energy window-based estimation on phantom studies [27], [28]. The ESSE method can be seen as a further approximation by simplifying the MC simulation as a convolution with a spatially-invariant learned kernel without using given attenuation information.

There has been much interest in deep learning (DL)-based reconstruction for emission tomography for the last few years [29], [30], [31], [32], [33], [34]. Recently, DL-based scatter estimation in SPECT was also investigated [15], [35]. Xiang et al. proposed a DL-based scatter estimation method for ^{90}Y SPECT that greatly reduces computational cost in testing while keeping the accuracy comparable to the MC method [15]. However, it still requires running enough MC simulation histories to generate scatters as ground truth labels for all training data. In other words, heavy computation in MC simulation-based method for one case is translated into even heavier computation in DL-based method for generating scatter estimates as ground truth labels before training.

This paper proposes a physics-guided deep scatter (PGDS) estimator using a weakly supervised framework that does not require computationally expensive full MC simulations for generating ground truth labels; in contrast, it uses $100\times$ shorter MC simulations to generate weak labels to train deep neural networks to learn the energy-window specific mapping from measurement to scatter estimates. Moreover, our weakly supervised framework enables quick fine-tuning of the trained network to incoming test data with a weak label for further improvement. After training the networks on 18 diverse XCAT phantoms for ^{177}Lu SPECT imaging, we evaluated our proposed methods on 6 XCAT phantoms, 4 realistic virtual patient phantoms and 1 torso phantom. Although ^{177}Lu emits two gamma-rays (208 keV (10%), 113 keV (6%)) suitable for SPECT imaging, in typical acquisition protocols only the higher energy gamma-ray is used because of the high downscatter associated with an acquisition window centered on the lower peak. However, due to the low ^{177}Lu gamma-ray yields it is desirable to increase count levels by imaging both gamma-rays, provided a suitable scatter correction can be performed. Hence, in our phantom studies we evaluate scatter estimation for both single (208 keV) and dual (113 keV, 208 keV) photopeaks acquisition. In addition to the phantom studies, we evaluated our methods on 3 clinical scans from patients who underwent ^{177}Lu PRRT with SPECT imaging

after therapy administration to determine pharmacokinetics for dosimetry.

Section II briefly reviews commonly known scatter estimation methods in SPECT imaging. Then, Section III describes our proposed weak-supervised training framework with a very short MC simulation, called physics-guided deep scatter estimation (PGDS). Section IV outlines the detailed description on data and experimental settings and then Section V reports quantitative evaluation results on diverse XCAT phantoms, realistic virtual patient phantoms as well as real torso phantom and real clinical scans for ^{177}Lu SPECT with single / dual photopeaks. We conclude our paper with discussions in Sections VI and VII.

II. BACKGROUND

A. SPECT Image Reconstruction With Scatter Estimation

The SPECT measurement model can be described as:

$$\mathbf{y} \sim \text{Poisson}(\mathbf{A}\mathbf{x} + \mathbf{s}), \quad (1)$$

where $\mathbf{y} \in \mathbb{R}^M$ is the vector of measured projections that follow independent Poisson distributions, $\mathbf{x} \in \mathbb{R}^N$ denotes the unknown activity of emitted (unscattered) primary photons, $\mathbf{s} \in \mathbb{R}^M$ denotes the scatters, and the matrix $\mathbf{A} \in \mathbb{R}^{M \times N}$ models the SPECT imaging system including attenuation.

Given a scatter estimate $\hat{\mathbf{s}}$, the unknown activity \mathbf{x} is reconstructed using maximum likelihood (ML) estimation:

$$\hat{\mathbf{x}} = \arg \max_{\mathbf{x} \geq 0} L(\mathbf{x}; \mathbf{A}, \mathbf{y}, \hat{\mathbf{s}}),$$

$$L(\mathbf{x}; \mathbf{A}, \mathbf{y}, \hat{\mathbf{s}}) \triangleq \mathbf{y}' \log(\mathbf{A}\mathbf{x} + \hat{\mathbf{s}}) - \mathbf{1}'_M (\mathbf{A}\mathbf{x} + \hat{\mathbf{s}}), \quad (2)$$

where $\mathbf{x} \geq \mathbf{0}$ is a non-negativity constraint on \mathbf{x} . The ordered-subset expectation maximization (OSEM) [36] with early stopping is a popular choice for solving (2) with the following update equation:

$$\hat{\mathbf{x}} \leftarrow \hat{\mathbf{x}} + \hat{\mathbf{x}} \oslash (\mathbf{A}'_k \mathbf{1}_k) \odot \nabla L(\hat{\mathbf{x}}; \mathbf{A}_k, \mathbf{y}_k, \hat{\mathbf{s}}_k) \quad (3)$$

where \odot denotes element-wise multiplication, \oslash denotes element-wise division, k is a subset index that is varying over sub-iterations, \mathbf{A}_k is a submatrix of \mathbf{A} for the k th subset and $\mathbf{y}_k, \hat{\mathbf{s}}_k$ are the subvectors of $\mathbf{y}, \hat{\mathbf{s}}$ for the k th subset, respectively. We denote the OSEM reconstructed image by $\hat{\mathbf{x}} = \text{OSEM}(\mathbf{A}, \mathbf{y}, \hat{\mathbf{s}})$.

As the true scatter vector \mathbf{s} depends on the unknown true image \mathbf{x} , it is challenging to estimate $\hat{\mathbf{s}}$. The following subsections review some of the commonly used scatter estimation methods.

B. Energy Window-Based Scatter Estimation

Energy window-based scatter estimation uses additional acquisitions in scatter windows. Besides the measurement \mathbf{y} in the photopeak window, two additional measurements $\mathbf{y}^{\text{lw}}, \mathbf{y}^{\text{uw}}$ are collected in the lower and upper narrow energy windows, respectively. Then, the triple energy window (TEW) method [12], [13], [16], [37] estimates the scatter in the photopeak window using the following formula:

$$\hat{\mathbf{s}}^{\text{TEW}} = \left(\mathbf{y}^{\text{lw}}/w_{\text{lw}} + \mathbf{y}^{\text{uw}}/w_{\text{uw}} \right) \times w_{\text{pw}}/2, \quad (4)$$

Algorithm 1 MC Simulation-Based Scatter Correction [21]

Measurement \mathbf{y} , attenuation map $\boldsymbol{\mu}$, forward model \mathbf{A} ,
 # of MC histories α , # of updates J are given.

- 1: Incorporate $\boldsymbol{\mu}$ into \mathbf{A} for attenuation correction.
- 2: Set the initial scatter $\hat{\mathbf{s}} = \mathbf{0}$.
- 3: $\hat{\mathbf{x}} = \text{OSEM}(\mathbf{A}, \mathbf{y}, \hat{\mathbf{s}})$
- 4: **for** $j = 1, \dots, J$ **do**
- 5: $[\hat{\mathbf{y}}^{\text{MC}}, \hat{\mathbf{s}}^{\text{MC}}] = \text{MC}(\hat{\mathbf{x}}, \boldsymbol{\mu}, \alpha)$
- 6: $\hat{\mathbf{s}} = (\mathbf{y} \oslash \hat{\mathbf{y}}^{\text{MC}}) \odot \hat{\mathbf{s}}^{\text{MC}}$
- 7: $\hat{\mathbf{x}} = \text{OSEM}(\mathbf{A}, \mathbf{y}, \hat{\mathbf{s}})$
- 8: **end for**
- 9: **return** $\hat{\mathbf{x}}, \hat{\mathbf{s}}$

where w_{lw} , w_{uw} , and w_{pw} denote the widths of lower, upper, and photopeak windows, respectively.

C. Monte-Carlo Simulation-Based Scatter Estimation

Monte-Carlo (MC) simulation has been used estimate the scatter component by modeling the physics of photon transport in the patients and the SPECT camera [38] for several radionuclides such as ^{99m}Tc [18], [22], ^{166}Ho [39], ^{90}Y [20], [21], ^{131}I [2], [40], and recently ^{177}Lu [41].

MC-based scatter estimation assumes that the ratio of the unknown scatter component s to the measured total projection \mathbf{y} is proportional to the ratio of those simulated by MC [21], i.e.,

$$s \approx (\mathbf{y} \oslash \mathbf{y}^{\text{MC}}) \odot s^{\text{MC}}, \quad (5)$$

when the MC simulation is performed with very large photon histories. The unknown image \mathbf{x} is estimated by the OSEM with some estimated scatter $\hat{\mathbf{s}}$ (i.e., $\hat{\mathbf{x}} = \text{OSEM}(\mathbf{A}, \mathbf{y}, \hat{\mathbf{s}})$), and then the estimate of the unknown scatter s is updated using (5) and the MC simulation (SIMIND [42] in this work) with the estimated image $\hat{\mathbf{x}}$ and the corresponding attenuation map $\boldsymbol{\mu}$:

$$[\hat{\mathbf{y}}^{\text{MC}}, \hat{\mathbf{s}}^{\text{MC}}] = \text{MC}(\hat{\mathbf{x}}, \boldsymbol{\mu}, \alpha), \quad (6)$$

where α denotes the number of MC histories. Algorithm 1 and Fig. 1-A1 illustrate such alternating updates for $\hat{\mathbf{x}}$ and $\hat{\mathbf{s}}$.

While MC simulation-based scatter estimation is more accurate than the TEW method for a wide range of isotopes in SPECT and does not require additional acquisition windows, MC takes much more time, mainly due to the MC simulation step in the line 5 of Algorithm 1. For example, 500 million histories per projection were simulated per iteration using SIMIND and usually 2-3 iterations (J) are needed [21]; thus it took about 16-24 hours per one scatter correction using 12 cores (or 192-288 hours per patient using a single core).

D. Supervised Deep Learning-Based Scatter Estimation

Xiang et al. [15] proposed a deep neural network (DNN) for ^{90}Y SPECT scatter estimation. With the measurement \mathbf{y} and the projected attenuation map \mathbf{p} , the deep scatter estimation network $\mathbf{h}(\cdot)$ is

$$\hat{\mathbf{s}}^{\text{DNN}} = \mathbf{h}(\mathbf{y}; \mathbf{p}). \quad (7)$$

Training the DNN $\mathbf{h}(\cdot)$ requires a training set with suitable inputs (a large set of measurement and attenuation map pairs) and labels (ground truth scatters). The work of [15] used digital phantoms (L true activities $\{\mathbf{x}_1^{\text{true}}, \dots, \mathbf{x}_L^{\text{true}}\}$) and long MC simulations to generate the necessary measurements for inputs ($\{\mathbf{y}_1^{\text{oracle}}, \dots, \mathbf{y}_L^{\text{oracle}}\}$) and oracle scatters for ground truth labels ($\{\mathbf{s}_1^{\text{oracle}}, \dots, \mathbf{s}_L^{\text{oracle}}\}$). This heavy computation with long MC simulations on all training data is the primary ‘‘cost’’ for labeling in supervised learning. As illustrated in Fig. 1-A2, the training of the DNN $\mathbf{h}(\cdot)$ minimizes the following mean squared error (MSE) loss:

$$\mathcal{L}_{\text{MSE}}(\mathbf{h}) = \sum_{l=1}^L \left\| \mathbf{h}(\mathbf{y}_l^{\text{oracle}}; \mathbf{p}_l) - \mathbf{s}_l^{\text{oracle}} \right\|_2^2. \quad (8)$$

We denote the DNN trained with the loss (8) by $\mathbf{h}_{\text{MSE}}(\cdot)$.

While the DNN-based method can yield accurate and fast scatter estimates in inference, it is computationally expensive to generate ground truth scatter labels. In [15], 1 billion histories per projection were used for generating ground truth scatter labels. For 10 digital phantoms, it took about 133 hours using 12 CPU cores. Training with clinical scans would take even longer time due to using 2-3 iterations of long MC simulations in Algorithm 1. Thus, reducing labeling computation cost seems important in clinical applications with diverse, new imaging protocols, more patients and new diseases.

III. METHOD

This section describes our physics-guided deep scatter (PGDS) estimator using weakly supervised learning for 2D projection-wise scatter estimation in 3D quantitative SPECT. While supervised training for a scatter estimating DNN requires time-demanding MC simulations to generate scatters as ground truth labels [15], our proposed method uses weak labels from ‘‘short MC’’ simulations that are two orders of magnitude faster. Our scatter generation module is the same as the supervised method (7) and the scaling modules are the DNNs to indirectly enhance weak labels. They are jointly trained under weak supervision and we use only the trained scatter generation module for inference.

A. Physics-Guidance by Short MC Simulation

Inspired by the MC scatter estimation pipeline in Algorithm 1, we define the ‘‘physics-guidance’’ that is the key ingredient for our weakly supervised training of the scatter generation module. First, we rearrange the terms in (5) as:

$$(\mathbf{y}^{\text{MC}} \oslash \mathbf{s}^{\text{MC}}) \odot \hat{\mathbf{s}} \Rightarrow \mathbf{r}^{\text{MC}} \odot \hat{\mathbf{s}}, \quad (9)$$

where \mathbf{r}^{MC} denotes the proposed ‘‘physics-guidance’’ ratio vector that can be used for predicting the measurement \mathbf{y} using the scatter estimate $\hat{\mathbf{s}}$. An accurate ratio \mathbf{r}^{MC} can facilitate training of scatter generation module with a given measurement \mathbf{y} , but still requires long time MC simulation.

An approximate physics-guidance, denoted by $\mathbf{r}^{\text{short}}$, can be quickly generated by running a ‘‘short MC’’ simulation and

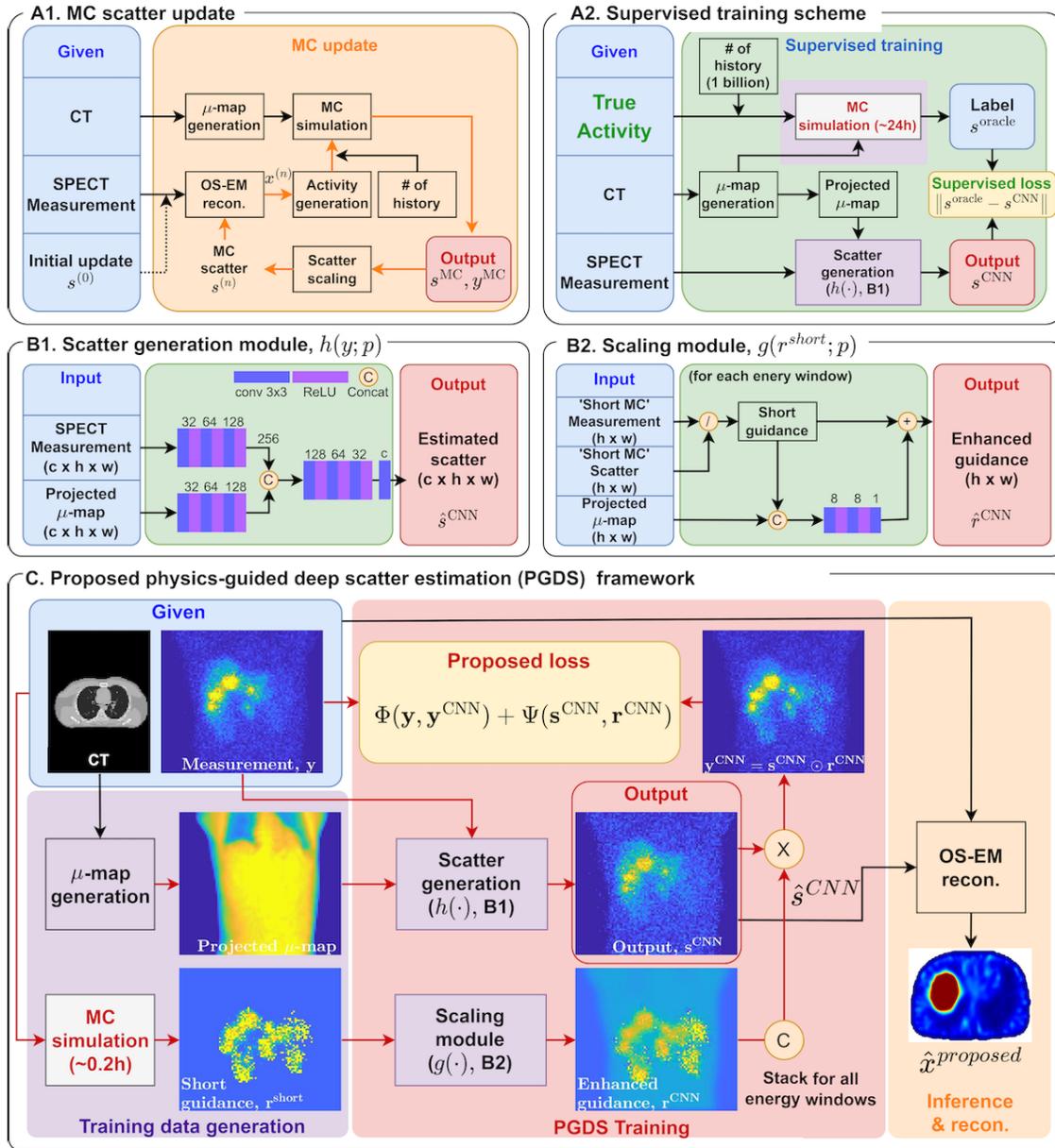


Fig. 1. (A1) Flowchart illustrating the MC scatter estimation pipeline [21]. “1 update” denotes one cycle of scatter estimation and reconstruction using the current estimated image to yield the next estimated image. (A2) Supervised training pipeline of scatter generation module with the MC simulation for generating ground truth scatters [15]. (B1) Network structure for the scatter generation module $h(\cdot)$ to estimate scatters from the measurements and its corresponding projected attenuation maps in all energy windows simultaneously. (B2) Network structure for the scaling module $g(\cdot)$ to enhance approximate physics-guidance with the projected attenuation map in each energy window. (C) Proposed PGDS with the scatter generation module and scaling modules by weak supervision (short MC simulation) for generating approximate physics-guidances.

then enhanced by the scaling module $g(\cdot)$ with a projected attenuation map p to predict r^{MC} as follows:

$$r^{MC} \approx \hat{r}^{CNN} = g(r^{short}, p). \quad (10)$$

However, we do not generate r^{MC} as labels. Instead, we propose a weakly supervised training framework for $g(\cdot)$ with the input r^{short} and the given measurement y .

B. Physics-Guided Deep Scatter Estimation (PGDS)

The scatter generation module in (7) and the scaling modules in (10) can predict the (total) measurement via (9):

$$y \approx \hat{y}^{CNN} \equiv \hat{r}^{CNN} \odot \hat{s}^{CNN} = g(r^{short}, p) \odot h(y, p). \quad (11)$$

Fig. 1-B1 illustrates the scatter generation module $h(y; p)$ whose output is the estimated scatter \hat{s}^{CNN} and inputs are the measurement y and the projected attenuation map p . Fig. 1-B2 illustrates the scaling module $g(\hat{r}^{short}; p)$ that was trained to learn scaling factors to enhance the approximate physics-guidance in each energy window. For multiple energy windows (say c energy windows), we jointly trained one scatter generation module and c separate scaling modules. The outputs of each scaling module were stacked channel-wise to have the same channel size as whole energy windows.

For a set of measurements $\{y_1, \dots, y_L\}$, approximate physics-guidances \hat{r}_l^{short} (scatters \hat{s}_l^{short} as weak labels) are generated via short MC simulations. Then, all modules are

Algorithm 2 Physics-Guided Deep Scatter Estimation (PGDS)

Train set: measurement, attenuation map $\{y_l, \mu_l\}, l = 1, \dots, L$, forward model A , # of MC histories β , # of epoch e

- 1: Initialize h, g and set $\hat{s} = \mathbf{0}, \forall l$
- 2: $\hat{x}_l = \text{OSEM}(A, y_l, \hat{s}_l), \forall l$
- 3: $[\hat{y}_l^{\text{short}}, \hat{s}_l^{\text{short}}] = \text{MC}(\hat{x}_l, \mu_l, \beta), \forall l$
- 4: Pre-compute $\hat{r}_l^{\text{short}} = \hat{y}_l^{\text{short}} \oslash \hat{s}_l^{\text{short}}, p_l = A\mu_l, \forall l$
- 5: **for** $i = 1, \dots, e$ **do**
- 6: $\hat{s}_l^{\text{CNN}} = h(y_l; p_l), \forall l$
- 7: $\hat{r}_l^{\text{CNN}} = g(\hat{r}_l^{\text{short}}, p_l), \forall l$
- 8: $\hat{y}_l^{\text{CNN}} = \hat{r}_l^{\text{CNN}} \odot \hat{s}_l^{\text{CNN}}, \forall l$
- 9: $[h, g] \leftarrow \arg \min$ Eq. (12) with respect to $[h, g]$
- 10: **end for**
- 11: **return** $[h_{\text{PGDS}}, g_{\text{PGDS}}] \leftarrow [h, g]$

jointly trained by minimizing our weakly supervised loss between the predicted projections \hat{y}_l^{CNN} and the measured / simulated projections y_l via (11) as follows:

$$\mathcal{L}_{\text{PGDS}}(g, h) = \sum_{l=1}^L \Phi(y_l, \hat{y}_l^{\text{CNN}}) + \Psi(\hat{r}_l^{\text{CNN}}, \hat{s}_l^{\text{CNN}}) \quad (12)$$

where $\Phi(\hat{y}_l^{\text{CNN}}, y_l)$ is the Poisson negative log-likelihood $(\hat{y}_l^{\text{CNN}} - y_l \log \hat{y}_l^{\text{CNN}})$ and $\Psi(\hat{r}_l^{\text{CNN}}, \hat{s}_l^{\text{CNN}})$ is a regularizer to avoid undesirable solutions $g(r^{\text{short}}, p) = 1, h(y; p) = y$:

$$\Psi(\cdot, \cdot) = c_1 \left\| \hat{r}_l^{\text{CNN}} - r_l^{\text{short}} \right\|_1 + c_2 \left\| \hat{s}_l^{\text{CNN}} - s_l^{\text{short}} \right\|_1.$$

We chose $c_1, c_2 = 0.01$ from the studies in Section V-E.

Algorithm 2 summarizes the labeling / training procedure of our PGDS including OSEM reconstruction and short MC simulation. For weak labels, short MC simulations are performed for all measurements with reduced MC histories, which are 100 times faster than full MC simulations (i.e., β in Algorithm 2 $\ll \alpha$ in Algorithm 1). Then, for the measurement y and its projected attenuation p , the scatter is estimated by

$$\hat{s}^{\text{PGDS}} = h_{\text{PGDS}}(y; p). \quad (13)$$

Fig. 1-C illustrates our weakly supervised framework.

Algorithm 3 Patient Specific Fine-Tuning (PGDS-T)

Test set: measurement y , attenuation map μ , forward model A , # of MC histories β , # of epoch e_t

- 1: set $\hat{s} = \mathbf{0}, [h, g] = [h_{\text{PGDS}}, g_{\text{PGDS}}]$
- 2: $\hat{x} = \text{OSEM}(A, y, \hat{s})$
- 3: $[\hat{y}^{\text{short}}, \hat{s}^{\text{short}}] = \text{MC}(\hat{x}, \mu, \beta)$
- 4: Pre-compute $\hat{r}^{\text{short}} = \hat{y}^{\text{short}} \oslash \hat{s}^{\text{short}}, p = A\mu$
- 5: **for** $i = 1, \dots, e_t$ **do**
- 6: $\hat{s}^{\text{CNN}} = h(y; p)$
- 7: $\hat{r}^{\text{CNN}} = g(\hat{r}^{\text{short}}, p)$
- 8: $\hat{y}^{\text{CNN}} = \hat{r}^{\text{CNN}} \odot \hat{s}^{\text{CNN}}$
- 9: $[h, g] \leftarrow \arg \min$ Eq. (12) with respect to $[h, g]$
- 10: **end for**
- 11: **return** $[h_{\text{PGDS-T}}, g_{\text{PGDS-T}}] \leftarrow [h, g]$

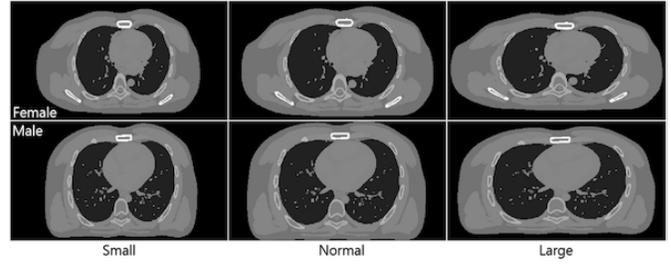


Fig. 2. Examples for 6 different anatomical structures of XCAT phantoms corresponding to 3 sizes (small, normal, large) \times 2 genders (male, female).

C. Patient-Specific Fine-Tuning (PGDS-T)

Our proposed method also supports the fine-tuning of the trained networks $[h_{\text{PGDS}}, g_{\text{PGDS}}]$ to a given measurement with weak label via additional short MC simulation. Algorithm 3 describes the details of fine-tuning our networks using a new test measurement y . Fine-tuning can be useful in inference stage by adapting the trained network in (13) to the test data and then estimating the scatter using (13). We found that our model trained on a synthetic dataset can be fine-tuned to well perform on clinical patient dataset. We set $e = 100$ in Algorithm 2 and $e_t = 30$ in Algorithm 3.

IV. EXPERIMENT SETUP**A. Synthetic, Realistic, and Real Datasets**

1) *Digital XCAT Phantom Dataset (Training, Testing)*: ^{177}Lu SPECT scans were simulated using the SIMIND MC simulator [42] with emulating imaging following ^{177}Lu DOTATATE PRRT a total of 24 3D XCAT phantoms [43] with diverse anatomies and activities (4 activity types \times 6 anatomy types). Anatomy differences include gender (male, female) and size (normal, small with scaling factors [0.85, 0.95] and large with scaling factors [1.1, 0.9]) as illustrated in Fig. 2. Activity differences are summarized in Table I, having different activity ratios in organs, different lesion volumes and activities. All 24 XCAT phantom simulations were divided into 18 simulations for training and 6 simulations for testing.

Our SIMIND MC simulator models full photon transport in the object and camera, including collimator septal penetration and scatter effect in the object, collimator, NaI(Tl) crystal, and backscatter layer. We configured the SIMIND simulator to accurately emulate ^{177}Lu imaging in the clinic with Siemens Intevo Bold SPECT/CT system equipped with a medium-energy (ME) collimator, a 15 mm crystal, acquisition windows for the two photopeaks centered at 113 and 208 keV with a 20% window width, two 10% side window width (only for TEW), 128 projections over 360° , and the reconstructed image size of $128 \times 128 \times 81$ voxels ($4.8 \times 4.8 \times 4.8 \text{ mm}^3$).

For supervised training, ground truth scatters from the true activity images were generated as labels by using the SIMIND MC simulation with 1 billion histories per projection as suggested in [15] (denoted as ‘Oracle’). In contrast, for our proposed weakly supervised training, approximate scatter estimates for the given measurements were generated as weak labels using Algorithm 1 with 1 update and ‘Short MC’

TABLE I

ACTIVITY RATIOS AND LESION VOLUMES IN XCAT PHANTOMS FOR TRAINING (TYPE 1-3) AND TESTING (TYPE 4)

Phantom	Activity ratio					Lesion volumes [mL]
	kidney	liver	lung	bkg	lesion	
1	3	1	0	0.15	6	22, 22, 32, 37.2, 99
2	1	1	0	0.07	3	7.8, 15, 23, 68
3	1	1	0	0.08	7	5.4, 9.2, 10.2, 34.3
4	2	1	0	0.10	10	5.6,12.4,12.8,29.7,34.6

TABLE II

ACTIVITY RATIOS AND LESION VOLUMES IN NON-UNIFORM VIRTUAL PATIENT PHANTOMS FOR TESTING

Type	Activity ratio				Activity ratio (Volume [mL])
	kidney	liver	lung	bkg	
VP1	1.4	1	0	0.12	2.9 (3.9), 2.8 (93.0)
VP2	1.1	1	0	0.08	2.1 (1.7), 2.4 (2.2), 1.9 (25.1)
VP3	1.6	1	0	0.14	3.1 (7.0), 3.7 (8.7), 3.6 (12.2), 2.8 (12.7), 3.3 (15.2), 3.4 (59.3)
VP4	1.3	1	0	0.08	3.0 (9.8), 3.0 (15.7), 2.8 (18.2), 3.3 (65.3), 3.3 (152.2)

simulation with 10 million histories. The former took 24 hours per phantom for supervised training, while the latter took 15 minutes (0.24 hours per phantom) for weakly supervised training. All MC simulations were performed using 16 threads on a Mac Pro.

2) Non-Uniform Virtual Patient Phantom Dataset (Testing):

Four clinically relevant phantoms (of size $512 \times 512 \times 130$ voxels, $0.98 \times 0.98 \times 3\text{mm}^3$) were used in the simulation. For each phantom, the organ masks (liver, kidney, and spleen) and lesion masks were delineated by a radiology technologist and a radiologist at the University of Michigan (UM), respectively, on diagnostic CT images where patients underwent ^{177}Lu DOTATATE therapy with UM Institutional Review Board (IRB) approval for retrospective analysis. The activity ratios were computed from the post-therapy ^{177}Lu SPECT/CT scan, but the phantom was defined by ^{68}Ga PET/CT images ($200 \times 200 \times 577$ voxels, $4.073 \times 4.073 \times 2\text{mm}^3$) to exploit the higher spatial resolution. PET is acquired ahead of time to determine the eligibility of ^{177}Lu SPECT DOTATATE therapy and expected to be similar to ^{177}Lu SPECT activity distributions (see Table II).

The center slices of PET images covering from lung to kidney were extracted, which were typical regions imaged by ^{177}Lu SPECT (SPECT field-of-view (FOV)). Density maps were generated from the CT images using the bi-linear conversion from a prior calibration experiment using a phantom with 16 tissue equivalent rods. The activity maps were registered into CT image spaces before the SIMIND MC simulation.

3) Torso Phantom Dataset (Testing):

A liver/lung torso phantom (Data Spectrum) was modified to include “lesion” inserts (a 30mL ovoid shape and a 16mL sphere) in the liver. ^{177}Lu activities in the form of a chloride solution were carefully measured [44] and used to fill the liver and inserts to achieve an insert-to-liver activity concentration ratio of 3.8:1, representative of liver lesions encountered when imaging after ^{177}Lu DOTATATE therapy. No activity was used in the lungs or the phantom compartment outside of the liver. The total activity in the phantom at scanning time was 118 MBq.

The SPECT acquisition used 20% windows at 113 keV and 208 keV photopeaks and adjacent 10% lower and upper scatter windows. Other acquisition parameters were the same as those used for patient imaging except that a prolonged acquisition of 275 sec/view was used to make the count-level (noise) representative of patient imaging.

4) *Clinical Patient Dataset (Testing)*: In addition to the above digital phantom data and virtual patient phantom data, we tested our method on clinical SPECT/CT datasets from 2 patients. Two patients underwent the cycle 1 of standard (7.4 GBq/cycle) ^{177}Lu -DOTATATE PRRT and then imaged by the previously described Siemens Intevo Bold SPECT/CT system with approval by the UM IRB. The patient provided written informed consent. Manufacture recommended settings of 20% acquisition window at 208 keV with adjacent 10% scatter windows, 256×256 matrix (downsampled to 128×128 in reconstruction) and 60×2 views were used. For one of the clinical scans, two 25 seconds/view SPECT scans were performed on the system described above at day 0 and day 1, respectively. Non-contrast, free breathing CT scan at the reference timepoint was performed at 120 kVp, 80 mAs at the reference and 15 mAs at all other timepoints. The patient image was upsampled using bilinear interpolation to have the same size with the virtual patient phantom.

B. Implementation Details

The same network structure as Xiang et al. [15] was used in this work as illustrated in Fig. 1-B. The scaling module $g(\cdot)$ has 3 convolution layers following ReLU (Rectified Linear Unit) activation for each layer. The input of the scaling module was approximate physics-guidance from “short” MC and projected μ -map. A skip-connection was used between the input and the output of the scaling module. Multiple scaling modules were used to account for different energy windows. The Adam optimizer was used for training and fine-tuning [45] with the parameters of $\beta_1 = 0.9$, $\beta_2 = 0.999$, batch size 80, 100 epochs, and learning rate 1×10^{-4} with decaying by 0.2 every 50 epochs. For fine-tuning, batch size, number of epochs and learning rate were set to 20, 30 and 1×10^{-5} , respectively.

Five independent noise realizations were used for each XCAT phantom for training, so a total of 11,520 projections ($18 \text{ XCAT phantoms} \times 5 \text{ realizations} \times 128 \text{ projections}$) were used for 90% training and 10% validation sets. The measurements and projected attenuation maps were not normalized. However, since the SIMIND normalizes its output to be 1 Mbq per 1 second in FOV, short MC result was normalized to have the same maximum value of the given measurement. Approximate physics-guidance from short MC was normalized by 99% percentile per each energy window. The training time for the proposed weakly supervised learning method was 1.5 hours on an NVIDIA RTX 3060Ti GPU.

All compared scatter estimation methods are summarized:

- **Oracle** True scatter generated along with measurement using MC simulation with 1 billion histories.
- **No scatter** All scatter components were set to zero.
- **TEW** TEW scatter estimation using (4).

- **Short MC** MC scatter estimation using Algorithm 1 with 1 update and 10 million photon histories.
- **Full MC** MC scatter estimation using Algorithm 1 with 3 updates and 1 billion histories per each MC simulation.
- **Supervised** DL scatter estimation trained with the ground truth scatters (labels) from Oracle.
- **PGDS** Proposed DL scatter estimation trained using Algorithm 2 using Short MC for weak labels.
- **PGDS-T** Proposed DL scatter estimation fine-tuned on test data using Algorithm 3 with Short MC for weak labels.

Since both methods use the same scatter estimation network layers that were trained differently, ours and previous supervised methods take the same inference time for scatter estimates. The reconstruction times for all methods are the same once the scatter estimates are generated since image reconstruction in all the above methods was using OSEM (20 iterations, 4 subsets), which takes only about 30 seconds for all iterations in each reconstruction step. We performed two reconstructions: one with single photopeak window (208 keV) and the other with dual photopeak windows (113, 208 keV) centered on the 2 dominant gamma-rays associated with ^{177}Lu . The latter summed two reconstructed images from two energy windows with the weights proportional to primary counts [46]. These weights can be estimated from measurements [47] or point source MC simulation [48].

Estimated scatter components were not blurred in the XCAT phantom dataset to account for piece-wise uniform activity. Hanning window filtering was applied in the non-uniform virtual patient and the clinical datasets, which is similar to the process on the clinical workstation [13].

C. Evaluation Metrics

The following metrics were used to compare our proposed method with other scatter estimation methods for quantitative SPECT imaging. Recovery coefficient (RC) on volume-of-interest (VOI) is a widely used metric that can assess the quantitative accuracy for hot spheres and is defined as

$$\text{RC}_{\text{VOI}} = \frac{\sum_{n \in \Omega_{\text{VOI}}} \hat{x}[n]}{\sum_{n \in \Omega_{\text{VOI}}} x[n]}$$

where $\hat{x}[n]$, $x[n]$ denote the n th voxel value of the reconstructed, true activity, respectively, and Ω_{VOI} is an index set for a given VOI. All reconstructed images were normalized to have the same total activity with the true activity images on FOV without post-filtering. Residual count error (RCE) can assess the quantitative accuracy for cold regions with zero activity [49] and is defined as

$$\text{RCE}_{\text{VOI}} = \frac{1}{|\Omega_{\text{VOI}}|} \sum_{n \in \Omega_{\text{VOI}}} \hat{x}[n] / \frac{1}{|\Omega_{\text{BKG}}|} \sum_{n \in \Omega_{\text{BKG}}} x[n]$$

where Ω_{BKG} is an index set for a chosen background region and $|\cdot|$ is the number of elements in a set.

Background noise was measured using multiple realizations (say K realizations) as follows. Firstly, the reconstructed volume \hat{x}_k for the k th realization is obtained for all realizations. Then, the mean and variance volumes are computed as

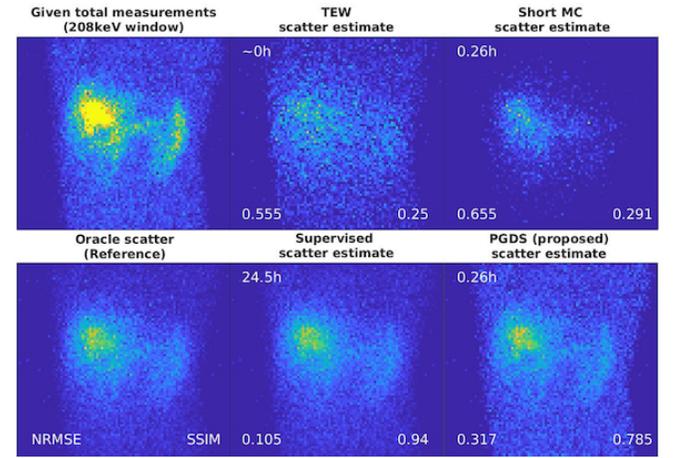


Fig. 3. MC simulated measurement ('projections') and its oracle scatter for test XCAT phantom (Phantom #4 in Table I) are shown in the leftmost column. The scatter estimates by TEW, Short MC, Supervised and our PGDS are shown with computation time, NRMSE and SSIM on the top left, bottom left and bottom right of each scatter estimate, respectively.

$\bar{\hat{x}} = \frac{1}{K} \sum_{k=1}^K \hat{x}_k$, $\text{Var}(\hat{x}) = \frac{1}{K-1} \sum_{k=1}^K (\hat{x}_k - \bar{\hat{x}})^2$, respectively. Finally, the background noise is given by

$$\text{Noise} = \frac{\sqrt{\frac{1}{|\Omega_{\text{BKG}}|} \sum_{n \in \Omega_{\text{BKG}}} \text{Var}(\hat{x})[n]}}{\frac{1}{|\Omega_{\text{BKG}}|} \sum_{n \in \Omega_{\text{BKG}}} \bar{\hat{x}}[n]}$$

Additionally, contrast recovery (CR) for the spherical lesions in the torso phantom experiment is calculated by:

$$\text{CR} = 100 \times \frac{C/C_{\text{BKG}} - 1}{R - 1}$$

where C is the mean counts for the lesion inserts, C_{BKG} is the mean counts of the background (normal liver) and R is the true lesion-to-normal liver activity concentration ratio ($R = 3.8$ in this study). The contrast-to-noise ratio for the lesion inserts is calculated as:

$$\text{CNR} = (C - C_{\text{BKG}}) / \text{STD}_{\text{BKG}}$$

where the background VOI is the normal liver (liver minus inserts). Normalized root mean squared error (NRMSE) was used for evaluating \hat{x} with respect to x and structural similarity index measure (SSIM) was also used [50]. For digital XCAT and virtual patient phantom datasets, true activity maps x were available and for the torso phantom and clinical patient datasets, the reconstructed volume with MC scatter correction was used as ground truth.

V. EXPERIMENT RESULTS

This section shows the experimental results on the test datasets from 6 XCAT phantoms, 4 realistic virtual patient phantoms and 3 clinical scans of 2 patients.

A. Digital XCAT Phantom Study

Fig. 3 shows the different scatter estimation results. Our proposed method yielded visually comparable images to those reconstructed with Oracle and Supervised methods and also visually better than images that were reconstructed using

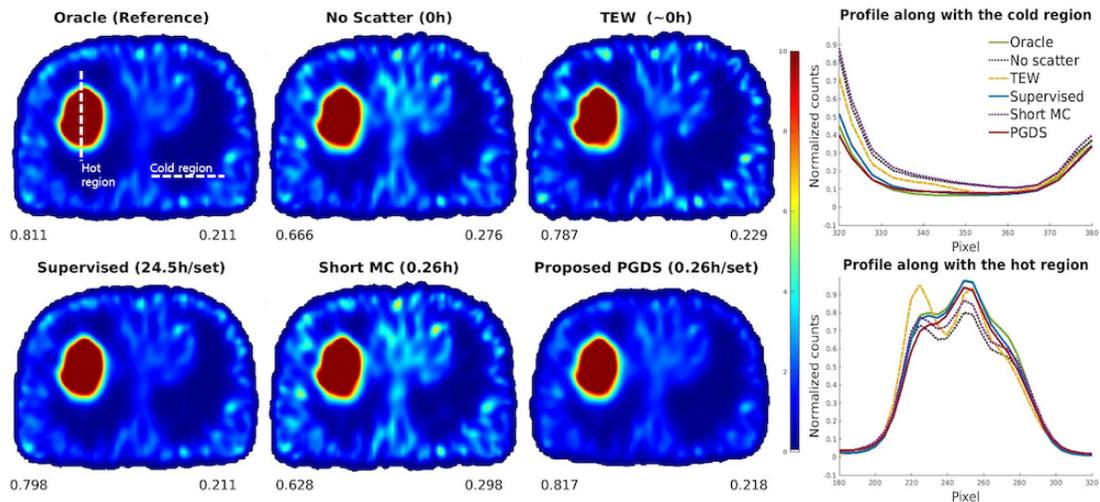


Fig. 4. Axial views of the reconstructed volumes and their normalized profiles (0 to 1) for digital XCAT phantom (Phantom #4 of Table I) using different scatter estimation methods; Oracle, No scatter, TEW, Supervised, Short MC and PGDS. RC for the hot region, RCE for the cold region and computation time are also reported on the bottom left, bottom right and top of each image, respectively. Our method yielded comparable performance to Oracle / Supervised with much faster computation.

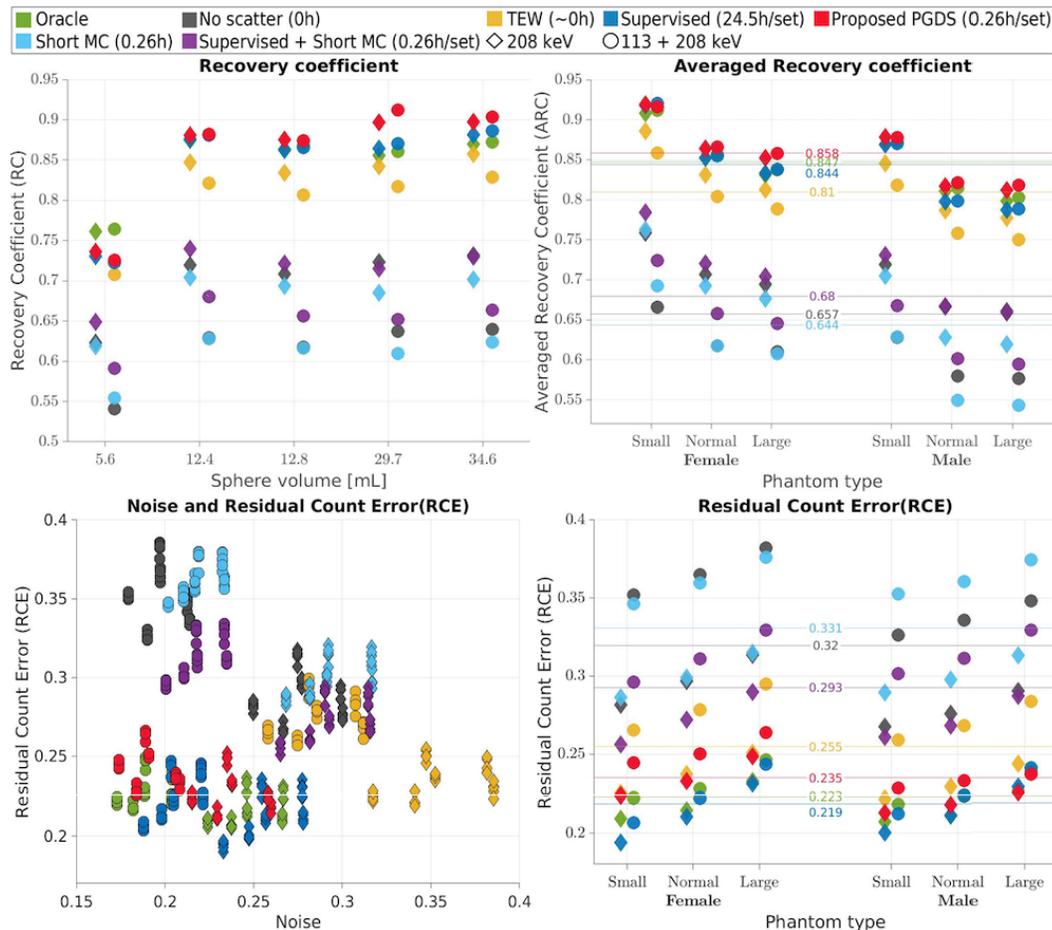


Fig. 5. Quantitative results of ^{177}Lu SPECT imaging on the test XCAT phantoms for single (208 keV, diamond) and dual (113, 208 keV, circle) photopeak windows using various scatter estimation methods. RC over sphere volumes, averaged RC over different anatomy types, RCE in lungs over background noise, and RCE over different anatomy types are reported with computation time in the legend. The values on the lines in the right column are the averages of RC and RCE over all phantom and window types. Our method yielded comparable results to Oracle and Supervised methods with much faster computation speed.

No Scatter, TEW and Short MC methods. Fig. 3 shows the estimated scatter components by different methods for our ^{177}Lu test XCAT phantom (Phantom #4 of Table I). TEW did not yield visually accurate scatter estimate while Supervised

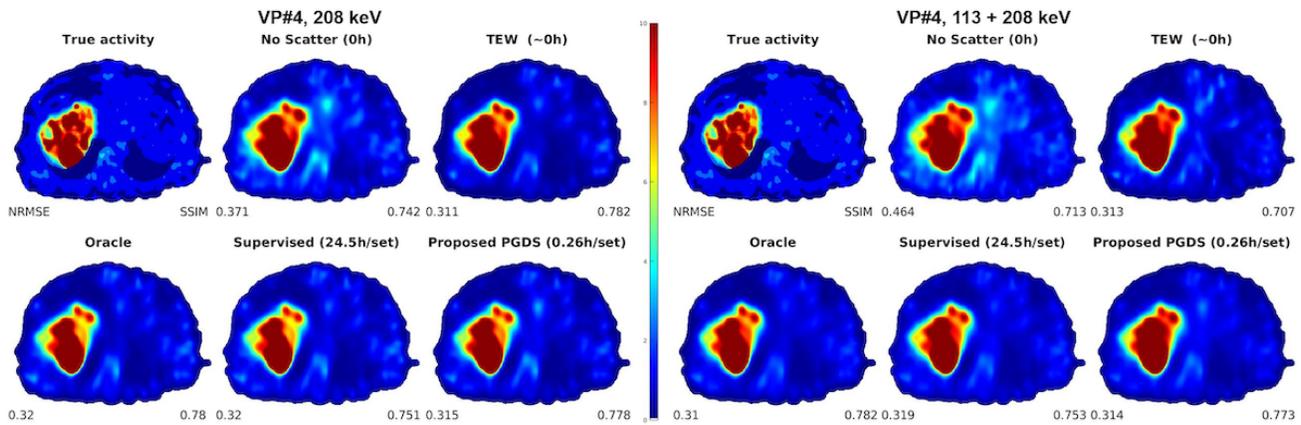


Fig. 6. Axial views of the true activity and the reconstructed images using various scatter estimation methods with NRMSE (bottom left in each image) and SSIM (bottom right in each image) for virtual patient phantom. For a single photopeak window, TEW, Supervised and our PGDS methods yielded similar quantitative and qualitative images to Oracle. For dual photopeak windows, all methods yielded similar images and quantitative values except for TEW, yielded worse SSIM than No scatter.

TABLE III

THE EVALUATION RESULTS OF OUR PROPOSED SCATTER ESTIMATION METHOD FOR ¹⁷⁷Lu ON VIRTUAL PATIENT PHANTOMS. OUR PGDS YIELDED COMPARABLE PERFORMANCE TO SUPERVISED METHOD AND IMPROVED VALUES OVER NO SCATTER AND TEW METHODS IN ALL METRICS. SEE TABLE II FOR MORE INFORMATION ON THE VIRTUAL PATIENT PHANTOMS INCLUDING ACTIVITY VOLUMES AND RATIOS. NOTE THAT RC SMALL AND RC LARGE DENOTE THE RCs FOR THE SMALLEST AND LARGEST LESIONS IN TABLE II, RESPECTIVELY. THE RCE VALUES WERE MEASURED IN THE LUNG REGION AND THE NOISE WAS CALCULATED IN THE BACKGROUND

208 keV						113 + 208 keV					
VP#1	Oracle	No scatter	TEW	Supervised	PGDS	VP#1	Oracle	No scatter	TEW	Supervised	PGDS
RC small	0.91	0.74	0.92	0.92	0.98	RC small	0.90	0.65	0.86	0.90	0.95
RC large	0.98	0.85	1.01	0.98	1.04	RC large	0.98	0.76	0.97	0.98	1.06
RCE	0.26	0.33	0.27	0.25	0.25	RCE	0.28	0.37	0.30	0.26	0.27
Noise	0.31	0.28	0.32	0.32	0.33	Noise	0.24	0.20	0.27	0.25	0.26
VP#2	Oracle	No scatter	TEW	Supervised	PGDS	VP#2	Oracle	No scatter	TEW	Supervised	PGDS
RC small	0.68	0.62	0.71	0.70	0.78	RC small	0.67	0.56	0.66	0.68	0.79
RC large	0.98	0.81	0.99	0.98	1.04	RC large	0.98	0.72	0.93	0.98	1.06
RCE	0.41	0.55	0.44	0.41	0.42	RCE	0.44	0.60	0.50	0.44	0.46
Noise	0.37	0.32	0.38	0.37	0.38	Noise	0.28	0.23	0.32	0.29	0.31
VP#3	Oracle	No scatter	TEW	Supervised	PGDS	VP#3	Oracle	No scatter	TEW	Supervised	PGDS
RC small	0.93	0.76	0.93	0.96	1.00	RC small	0.91	0.66	0.85	0.93	0.97
RC large	1.04	0.82	1.04	1.06	1.12	RC large	1.03	0.71	0.96	1.02	1.06
RCE	0.19	0.24	0.19	0.18	0.18	RCE	0.20	0.27	0.22	0.20	0.20
Noise	0.33	0.29	0.34	0.33	0.35	Noise	0.25	0.21	0.28	0.26	0.28
VP#4	Oracle	No scatter	TEW	Supervised	PGDS	VP#4	Oracle	No scatter	TEW	Supervised	PGDS
RC small	0.42	0.36	0.42	0.42	0.43	RC small	0.42	0.33	0.40	0.42	0.43
RC large	1.10	0.82	1.07	1.06	1.10	RC large	1.07	0.68	0.97	1.01	1.02
RCE	0.33	0.48	0.35	0.36	0.35	RCE	0.36	0.56	0.42	0.40	0.39
Noise	0.31	0.27	0.34	0.31	0.32	Noise	0.24	0.20	0.29	0.25	0.25

method yielded visually comparable scatter estimate to Oracle. Short MC yielded good number of counts in lesions and organs, but almost no count in background due to short photon simulations. The proposed method yielded visually comparable estimate to Oracle and Supervised methods with much faster computation, but with over-estimation in the background. For quantitative results, TEW and short MC yielded very low SSIM compared to Oracle, while our proposed PGDS yielded significantly enhanced NRMSE and SSIM over Short MC, as illustrated in Fig. 3. The supervised method slightly outperformed our PGDS in the background, but otherwise is comparable with our PGDS. However, note that Supervised method was trained with labels (Oracle) while our PGDS was trained with weak labels (Short MC), which explains the slight performance gap between Supervised and our PGDS methods.

Fig. 4 shows the axial views (slices) of XCAT reconstructed images and the line profiles over hot and cold regions for

different scatter estimation results that are illustrated in Fig. 3. Fig. 5 shows quantitative results of the reconstructed test XCAT phantoms with 208 keV and 113+208 keV energy windows using different scatter estimation methods, which are consistent with the results of Fig. 4. No scatter, Short MC and Supervised + Short MC yielded poor quantitative results with low RC and high RCE. TEW yielded improved RC and RCE over No scatter, Short MC and Supervised+Short MC, with high noise. Supervised method [15] had very similar performance to Oracle in all metrics, but with heavy computation for ground truth label generation with long MC simulations for all training data.

Our proposed PGDS yielded similar RCs and RCEs to those of Oracle and Supervised methods with small quantitative differences over different sizes of hot spheres and phantoms. For single and dual photopeak windows, Oracle, Supervised and our PGDS methods achieved consistently

TABLE IV

THE EVALUATION RESULTS OF OUR PROPOSED SCATTER ESTIMATION METHOD FOR ^{177}Lu TORSO PHANTOM MEASUREMENT. OUR PGDS WITH FINE-TUNING (PGDS-T) YIELDED COMPARABLE PERFORMANCE TO FULL MC AND SUPERVISED METHOD IN 208 KEV ENERGY WINDOW AND ACHIEVED IMPROVED VALUE IN 113+208 KEV ENERGY WINDOW. NOTE THAT THE BEST AND THE SECOND BEST SCORES EXCEPT FULL MC (AS A BENCHMARK) ARE LABELED AS BOLD AND UNDERLINED, RESPECTIVELY

208 keV							113 + 208 keV						
Metric	Full MC	No Scatter	TEW	Supervised	PGDS	PGDS-T	Metric	Full MC	No Scatter	TEW	Supervised	PGDS	PGDS-T
NRMSE	-	0.14	0.10	0.06	<u>0.05</u>	0.04	NRMSE	-	0.14	0.17	0.06	0.08	<u>0.07</u>
SSIM	-	0.98	0.92	<u>0.99</u>	<u>0.99</u>	1.00	SSIM	-	<u>0.98</u>	0.88	0.96	0.96	0.99
CR [%] (Sphere)	82.0	68.1	<u>75.8</u>	74.6	75.1	77.5	CR [%] (Sphere)	77.5	64.4	<u>74.8</u>	73.6	74.4	79.7
CR [%] (Ovoid)	82.6	70.4	77.2	<u>77.4</u>	77.2	79.3	CR [%] (Ovoid)	79.4	67.6	77.4	77.7	<u>77.8</u>	82.6
RCE	0.00	0.00	0.00	0.00	0.00	0.00	RCE	0.00	0.01	0.00	0.00	0.00	0.00
CNR (Sphere)	6.26	6.11	5.67	<u>6.22</u>	6.10	6.23	CNR (Sphere)	6.26	5.99	5.67	<u>6.17</u>	6.02	6.27
CNR (Ovoid)	6.31	6.32	5.77	6.45	6.27	<u>6.37</u>	CNR (Ovoid)	6.42	6.29	5.87	6.51	6.30	<u>6.49</u>

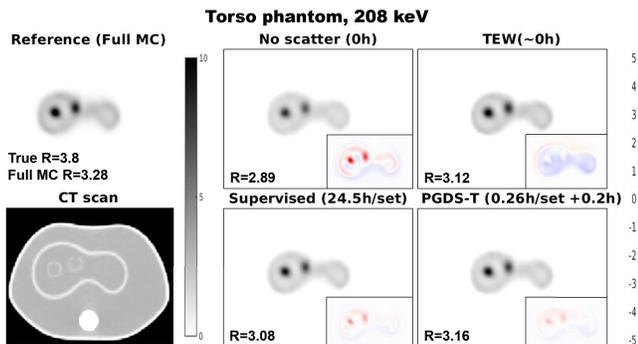


Fig. 7. The reconstructed images using various scatter estimation methods, the CT image of the torso phantom, and each absolute difference map for full MC images. The score R on the left bottom corner is the sphere-to-background ratio of the 16mL sphere (True R = 3.8). The sphere-to-background ratio can be converted into contrast recovery by using the equation in subsection IV-C. Note that the other quantitative evaluations are in Table IV.

small performance gaps in RC and RCE between single and dual energy windows and noise reduction using dual energy windows without compromising RCE performance. However, No scatter, TEW, Short MC, and Short MC+Supervised methods achieved noise reduction using dual energy windows with degraded performance in RCE. Moreover, including the 113 keV window for No scatter, TEW, Short MC, and Short MC+Supervised methods degraded the performance in RC and RCE, which suggests the importance of accurate scatter estimation when using multiple photopeak windows due to downscatter contributions.

B. Virtual Patient Phantom Study

Fig. 6 shows the SPECT reconstructed images for virtual patient phantom data with 208 keV and 113+208 keV along with NRMSE and SSIM. The noises of the reconstructed images with Oracle, Supervised and our PGDS methods were reduced in 113+208 keV as compared to those in 208 keV without compromising other quantitative values such as NRMSE and SSIM. However, the SSIM was degraded for No scatter and TEW methods when additionally using more counts in the 113 keV photopeak window. Quantitative evaluations for all 4 virtual patient phantoms are reported in Table III, showing consistent results as illustrated in Fig. 6.

Our proposed PGDS with computationally efficient weak labels via short MC yielded comparable performance in RC, RCE and noise to the Oracle and Supervised methods with heavy MC simulations for the test and training datasets, respectively.

C. Torso Phantom Study

Fig. 7 illustrates the SPECT reconstruction results of the torso phantom data. The images in the left column are the visualization of reference activity as well as the CT image. Table IV shows the quantitative evaluation results for both 208 keV and 113+208 keV energy windows. Note that MC-based scatter corrected images were used as ground truth for computing NRMSE and SSIM. To compute contrast recovery (CR) and contrast-to-noise ratio (CNR), CT annotated mask was used. Since the CR score is computed by using the true lesion-to-normal liver activity (R) which is 3.8, a value close to 100(%) implies a more accurate reconstruction performance. Our PGDS often yielded better performance than the Supervised method on CR but less performance on CNR due to noise. Moreover, our proposed method with fine-tuning (PGDS-T) achieved a comparable performance to or outperformed all other methods in both 208 keV and 113+208 keV energy windows by showing CR as 79.7 and 82.6 for the sphere and the ovoid, respectively.

D. Clinical Patient Study

Fig. 8 illustrates the SPECT reconstruction results of the clinical patient scan as well as the profile along the kidney region. Table V shows the quantitative evaluation results for 3 clinical scans from 2 patients. Note that MC-based scatter corrected images were used as ground truth for all evaluations (Full MC instead of Oracle). Our proposed method with fine-tuning (PGDS-T) outperformed all other methods including Supervised scatter estimation method. The proposed PGDS often yielded better performance than Supervised method - Our PGDS yielded 0.075 NRMSE and 0.960 SSIM, while Supervised method yielded 0.092 NRMSE and 0.956 SSIM. Our proposed PGDS-T further improved the reconstruction quality with 0.029 NRMSE and 0.996 SSIM using fine-tuning and additional weak label for a test measurement. Also, the absolute difference images (left) and relative

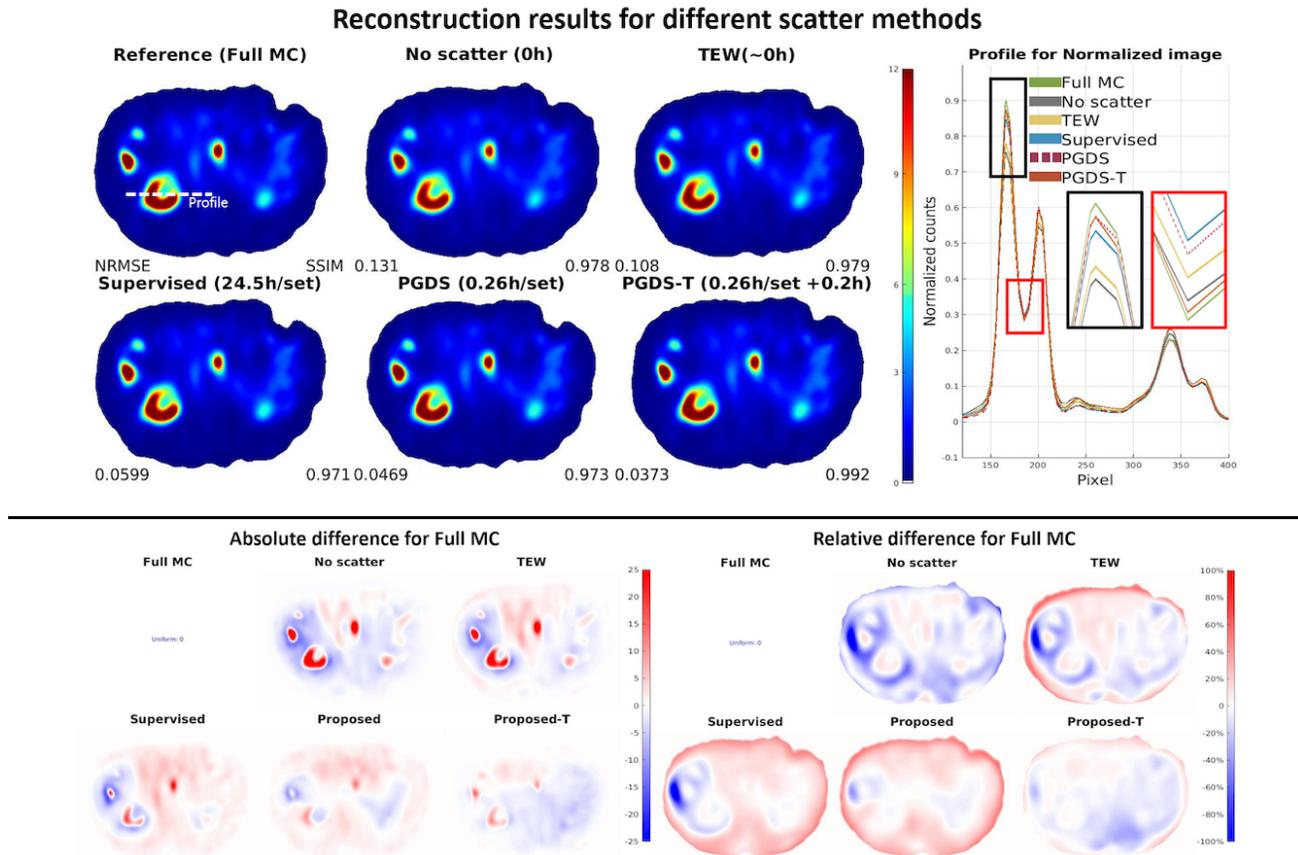


Fig. 8. Reconstructed images and profiles of reconstructed images for clinical patient data (P#1 Scan#2) using various scatter estimation methods. NRMSE and SSIM were evaluated with MC-based scatter corrected images as ground truth. Our proposed PGDS method yielded the highest contrast in the kidney region and the PGDS with fine-tuning (Proposed-T) yielded the lowest difference map with 0.0373 NRMSE while the Supervised method yielded 0.0599 NRMSE.

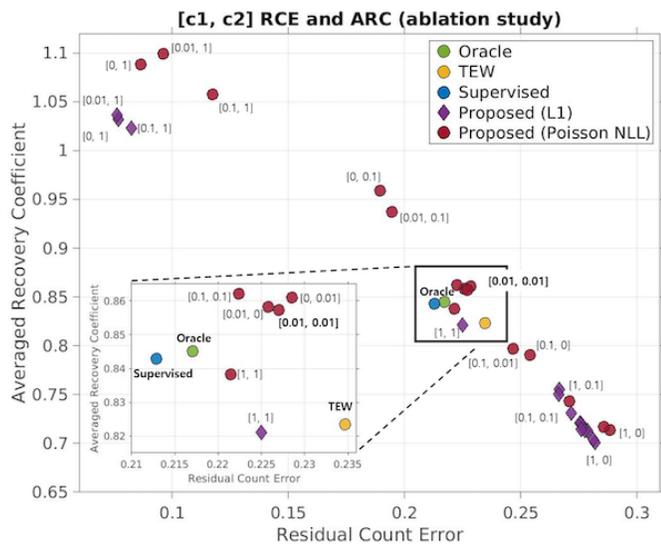


Fig. 9. Ablation study of PGDS for selecting the loss function (L1 or Poisson NLL), and the values of c_1 and c_2 in equation after (12). Each dot indicates the averaged performance over all phantom types.

difference images (right) show that our PGDS-T yielded the reconstructed image that is the closest to MC-based scatter corrected image.

TABLE V
QUANTITATIVE EVALUATION RESULTS OF VARIOUS SCATTER ESTIMATION METHODS FOR ^{177}Lu SPECT IMAGING ON 3 CLINICAL SCANS FROM 2 PATIENTS. THE RECONSTRUCTED IMAGE FROM THE FULL MC WAS USED AS THE GROUND TRUTH IN COMPARISON

NRMSE (SSIM)	No scatter	TEW	Supervised	PGDS	PGDS-T
P#1 Scan#1	0.140 (0.988)	0.155 (0.970)	0.079 (0.986)	0.078 (0.988)	0.028 (0.997)
P#1 Scan#2	0.131 (0.978)	0.108 (0.979)	0.060 (0.971)	0.047 (0.973)	0.037 (0.992)
P#2 Scan#1	0.176 (0.980)	0.165 (0.884)	0.092 (0.956)	0.075 (0.960)	0.029 (0.996)

E. Ablation Study

To evaluate the effectiveness of each component of proposed weakly supervised loss in (12), we performed an ablation study of for our proposed PGDS by using a different data-fitting term (Poisson negative log-likelihood vs. L1-norm (11) and by using grid search of c_1 and c_2 in various scales from 0 to 1. Fig. 9 shows the averaged performance over different phantom types. By comparing the two loss functions (11 and Poisson NLL), when the PGDS uses Poisson NLL loss, the lower RCE and higher RC was achieved than that of 11 loss. In Fig. 9 we enlarged the region to include the results with similar performance in average RC and RCE (the width and height of this region are 0.02 in RCE and 0.04 in RC). When the ratio

between c_1 and c_2 is 1:1, regardless of scales the reconstructed image qualities are very similar to each other (i.e., 0.01:0.01, 0.1:0.1 or 1:1). In contrast, when the ratio of c_1 to c_2 is much larger than 1, they yielded poor RCEs.

VI. DISCUSSION

Scatter estimation for SPECT using a supervised DL-based approach requires true scatter labels for training, which is very time consuming to generate; though once trained, it could be a good alternative to MC-based scatter estimation since it only takes 30 seconds for DL in inference while it takes 19 hours \times 3 updates for MC in inference. In this work, we propose a method to significantly reduce label cost (computation in MC simulation) by a factor of 100 (24.5 hours/set for label generation in supervised learning if true activities are available and 0.26 hour \times 1 update for weak label generation in our weakly supervised learning for the proposed PGDS). When true activities are not available, the computation time for generating training label is increased because Oracle must be replaced with Full MC that takes 19 hours \times 3 updates for each training set. It seems that the computation time gain is mostly in the labeling stage before the training stage, but this ‘computation gain’ in this stage can provide the advantage of training and retraining flexibility in clinical usage such as using different isotopes for other imaging, or patient-specific models. For different imaging systems with different isotopes, it is likely that one has to re-train the scatter estimation DNNs. Thus, our proposed method can save the overall time before and after training and hence more favorable compared to Supervised methods. Additionally, our PGDS has better generalization property to new data set only with short time fine-tuning.

To overcome the limitation of DL-based approaches that require labeled targets for training, we investigated several schemes before arriving at our proposed weakly supervised framework. One possible model was to directly multiply the small MC-based physical ratio by the output of the network without using the scaling network. Since the short MC yielded insufficient information on both the simulated measurement and simulated scatter, many of the ratios were 1. Unfortunately, it led to almost no back-propagation for training so the weakly supervised scheme was not able to yield sufficiently accurate results for clinical use. Another possible model was to replace the region of the value of 1 with the mean of the simulated ratio region, but it led to distorting the network output.

There were some options in designing the scaling network especially for handling multiple energy windows and using skip connections. We observed that using two scaling modules without a skip-connection showed very slow convergence, whereas using one scaling module with a skip-connection resulted in a failure by showing less scatter estimation accuracy. The other options were applying one scaling network for both energy windows without a skip-connection layer or applying two scaling networks for each energy window with a skip-connection layer. Since we aimed to preserve the physical interpretability, we chose the latter option that had two scaling modules (one for each energy window) each with a skip-connection.

Our proposed method uses a 2D projection-wise scatter estimation for 3D SPECT, but does not exploit any relationship between projections. Thus, exploring inter-projection dependencies may improve scatter estimation. We have shown that using dual-window in ^{177}Lu SPECT imaging with scatter dominant 113 keV window is possible by reducing noise while maintaining comparable performance in other quantification metrics thanks to accurate scatter estimation without much computation. Therefore, investigating inter-energy window information for reconstruction [33], [46], [47] and scatter estimation can be another future work for further improving performance.

VII. CONCLUSION

We proposed a novel DL scatter estimation method for SPECT reconstruction to overcome the high computing cost in generating training labels. We developed a weakly supervised training that requires only ‘‘short’’ runs of MC for generating weak labels efficiently. Our experimental results showed that our proposed approach yielded comparable or even improved performance to the existing methods such as supervised learning [15], yet is 100 times faster as demonstrated by experiments including synthetic XCAT phantom, virtual patient, torso phantom and clinical patient datasets.

REFERENCES

- [1] A. Delker et al., ‘‘Dosimetry for 177 Lu-DKFZ-PSMA-617: A new radiopharmaceutical for the treatment of metastatic prostate cancer,’’ *EJNMMI*, vol. 43, no. 1, pp. 42–51, 2016.
- [2] Y. K. Dewaraja et al., ‘‘MIRD pamphlet no. 23: Quantitative SPECT for patient-specific 3-dimensional dosimetry in internal radionuclide therapy,’’ *J. Nucl. Med.*, vol. 53, no. 8, pp. 1310–1325, Aug. 2012.
- [3] Z. Li, J. A. Fessler, J. K. Mikell, S. J. Wilderman, and Y. K. Dewaraja, ‘‘DblurDoseNet: A deep residual learning network for voxel radionuclide dosimetry compensating for single-photon emission computerized tomography imaging resolution,’’ *Med. Phys.*, vol. 49, no. 2, pp. 1216–1230, Feb. 2022.
- [4] D. J. Kwekkeboom et al., ‘‘Treatment with the radiolabeled somatostatin analog [^{177}Lu -DOTA 0 ,Tyr 3] octreotate: Toxicity, efficacy, and survival,’’ *J. Clin. Oncol.*, vol. 26, no. 13, pp. 2124–2130, 2008.
- [5] A. Kennedy, ‘‘Radioembolization of hepatic tumors,’’ *J. Gastrointestinal Oncol.*, vol. 5, no. 3, p. 178, 2014.
- [6] D. J. De Vries, M. A. King, E. J. Soares, B. M. Tsui, and C. E. Metz, ‘‘Effects of scatter subtraction on detection and quantitation in hepatic SPECT,’’ *J. Nucl. Med.*, vol. 40, no. 6, p. 1011, 1999.
- [7] H. Zaidi and K. F. Koral, ‘‘Scatter modelling and compensation in emission tomography,’’ *Eur. J. Nucl. Med. Mol. Imag.*, vol. 31, no. 5, pp. 761–782, May 2004.
- [8] B. F. Hutton, I. Buvat, and F. J. Beekman, ‘‘Review and current status of SPECT scatter correction,’’ *Phys. Med. Biol.*, vol. 56, no. 14, pp. R85–R112, Jul. 2011.
- [9] G. El Fakhri, I. Buvat, H. Benali, A. Todd-Pokropek, and R. Di Paola, ‘‘Relative impact of scatter, collimator response, attenuation, and finite spatial resolution corrections in cardiac SPECT,’’ *J. Nucl. Med.*, vol. 41, no. 8, pp. 1400–1408, 2000.
- [10] J. C. Dickson, ‘‘Quantitative SPECT: A survey of current practice in the U.K. nuclear medicine community,’’ *Nucl. Med. Commun.*, vol. 40, no. 10, pp. 986–994, 2019.
- [11] D. M. V. Huizing, M. Sinaasappel, M. C. Dekker, M. P. M. Stokkel, and B. J. de Wit-van der Veen, ‘‘ ^{177}Lu Lutetium SPECT/CT: Evaluation of collimator, photopeak and scatter correction,’’ *J. Appl. Clin. Med. Phys.*, vol. 21, no. 9, pp. 272–277, 2020.
- [12] K. Ogawa, Y. Harata, T. Ichihara, A. Kubo, and S. Hashimoto, ‘‘A practical method for position-dependent Compton-scatter correction in single photon emission CT,’’ *IEEE Trans. Med. Imag.*, vol. 10, no. 3, pp. 408–412, Sep. 1991.

- [13] M. Ljungberg, A. Celler, M. W. Konijnenberg, K. F. Eckerman, Y. K. Dewaraja, and K. Sjögreen-Gleisner, "MIRD pamphlet no. 26: Joint EANM/MIRD guidelines for quantitative ^{177}Lu SPECT applied for dosimetry of radiopharmaceutical therapy," *J. Nucl. Med.*, vol. 57, no. 1, pp. 151–162, Jan. 2016.
- [14] M. M. A. Dietze, S. Van Der Velden, M. G. E. H. Lam, M. A. Viergever, and H. W. A. M. De Jong, "Fast quantitative reconstruction with focusing collimators for liver SPECT," *EJNMMI Phys.*, vol. 5, no. 1, pp. 1–16, Dec. 2018.
- [15] H. Xiang, H. Lim, J. A. Fessler, and Y. K. Dewaraja, "A deep neural network for fast and accurate scatter estimation in quantitative SPECT/CT under challenging scatter conditions," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 47, no. 13, pp. 2956–2967, Dec. 2020.
- [16] J. Zeintl, A. H. Vija, A. Yahil, J. Hornegger, and T. Kuwert, "Quantitative accuracy of clinical ^{99m}Tc SPECT/CT using ordered-subset expectation maximization with 3-dimensional resolution recovery, attenuation, and scatter correction," *J. Nucl. Med.*, vol. 51, no. 6, pp. 921–928, Jun. 2010.
- [17] J. Hashimoto et al., "Scatter and attenuation correction in technetium-99m brain SPECT," *J. Nucl. Med.*, vol. 38, no. 1, pp. 157–162, 1997.
- [18] J. Xiao, T. C. De Wit, S. G. Staelens, and F. J. Beekman, "Evaluation of 3D Monte Carlo-based scatter correction for ^{99m}Tc cardiac perfusion SPECT," *J. Nucl. Med.*, vol. 47, no. 10, pp. 1662–1669, 2006.
- [19] M. Ljungberg, S. Strand, and M. A. King, "The SIMIND Monte Carlo program," in *Monte Carlo Calculation in Nuclear Medicine: Applications in Diagnostic Imaging*. USA: CRC Press, 2012, pp. 145–163.
- [20] M. Elschot, M. G. Lam, M. A. van den Bosch, M. A. Viergever, and H. W. de Jong, "Quantitative Monte Carlo-based ^{90}Y SPECT reconstruction," *J. Nucl. Med.*, vol. 54, no. 9, pp. 1557–1563, 2013.
- [21] Y. K. Dewaraja et al., "Improved quantitative ^{90}Y bremsstrahlung SPECT/CT reconstruction with Monte Carlo scatter modeling," *Med. Phys.*, vol. 44, no. 12, pp. 6364–6376, Dec. 2017.
- [22] F. J. Beekman, H. W. A. M. D. Jong, and S. V. Geloven, "Efficient fully 3-D iterative SPECT reconstruction with Monte Carlo-based scatter compensation," *IEEE Trans. Med. Imag.*, vol. 21, no. 8, pp. 867–877, Aug. 2002.
- [23] M. P. Nguyen, M. C. Goorden, R. M. Ramakers, and F. J. Beekman, "Efficient Monte-Carlo based system modelling for image reconstruction in preclinical pinhole SPECT," *Phys. Med. Biol.*, vol. 66, no. 12, Jun. 2021, Art. no. 125013.
- [24] H. Kim, S. Y. Chun, J. Fessler, and Y. Dewaraja, "On enhancing Monte-Carlo scatter correction for Y90 bremsstrahlung SPECT using guided filtering," *J. Nucl. Med.*, vol. 60, no. 1, p. 1348, 2019.
- [25] T. Rydén et al., "Fast GPU-based Monte Carlo code for SPECT/CT reconstructions generates improved ^{177}Lu images," *EJNMMI Phys.*, vol. 5, no. 1, pp. 1–12, Dec. 2018.
- [26] E. C. Frey and B. M. W. Tsui, "A new method for modeling the spatially-variant, object-dependent scatter response function in SPECT," in *Proc. IEEE Nucl. Sci. Symp. Conf. Rec.*, vol. 2, Nov. 1996, pp. 1082–1086.
- [27] R. de Nijs, V. Lagerburg, T. L. Klausen, and S. Holm, "Improving quantitative dosimetry in ^{177}Lu -DOTATATE SPECT by energy window-based scatter corrections," *Nucl. Med. Commun.*, vol. 35, no. 5, pp. 522–533, 2014.
- [28] E. K. Ghodoosi et al., "The effect of attenuation map, scatter energy window width, and volume of interest on the calibration factor calculation in quantitative ^{177}Lu SPECT imaging: Simulation and phantom study," *Phys. Medica*, vol. 56, pp. 74–80, Dec. 2018.
- [29] I. Shiri et al., "Deep-JASC: Joint attenuation and scatter correction in whole-body ^{18}F -FDG PET using a deep residual network," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 47, no. 11, pp. 2533–2548, Oct. 2020.
- [30] A. J. Ramon, Y. Yang, P. H. Pretorius, K. L. Johnson, M. A. King, and M. N. Wernick, "Improving diagnostic accuracy in low-dose SPECT myocardial perfusion imaging with convolutional denoising networks," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2893–2903, Sep. 2020.
- [31] A. Akhavanallaf, I. Shiri, H. Arabi, and H. Zaidi, "Whole-body voxel-based internal dosimetry using deep learning," *Eur. J. Nucl. Med. Mol. Imag.*, vol. 48, no. 3, pp. 670–682, Mar. 2021.
- [32] D. Hwang et al., "Generation of PET attenuation map for whole-body time-of-flight ^{18}F -FDG PET/MRI using a deep neural network trained with simultaneously reconstructed activity and attenuation maps," *J. Nucl. Med.*, vol. 60, no. 8, pp. 1183–1189, Aug. 2019.
- [33] H. Lim, J. A. Fessler, and Y. K. Dewaraja, "Joint dual photopeak Lu-177 SPECT image reconstruction with a stacked system model," in *Proc. EJNMMI*, vol. 45. Cham, Switzerland: Springer, 2018, pp. S95–S96.
- [34] Z. Li, Y. K. Dewaraja, and J. A. Fessler, "Training end-to-end unrolled iterative neural networks for SPECT image reconstruction," *IEEE Trans. Radiat. Plasma Med. Sci.*, vol. 7, no. 4, pp. 410–420, Apr. 2023.
- [35] H. Arabi, A. Akhavanallaf, A. Sanaat, I. Shiri, and H. Zaidi, "The promise of artificial intelligence and deep learning in PET and SPECT imaging," *Phys. Medica*, vol. 83, pp. 122–137, Mar. 2021.
- [36] H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," *IEEE Trans. Med. Imag.*, vol. 13, no. 4, pp. 601–609, Dec. 1994.
- [37] A. P. Robinson, J. Tipping, D. M. Cullen, and D. Hamilton, "The influence of triple energy window scatter correction on activity quantification for ^{177}Lu molecular radiotherapy," *Phys. Med. Biol.*, vol. 61, no. 14, pp. 5107–5127, Jul. 2016.
- [38] W. E. Bolch, "The Monte Carlo method in nuclear medicine: Current uses and future potential," *J. Nucl. Med.*, vol. 51, no. 3, pp. 337–339, Mar. 2010.
- [39] M. Elschot et al., "Quantitative Monte Carlo-based holmium-166 SPECT reconstruction," *Med. Phys.*, vol. 40, no. 11, Oct. 2013, Art. no. 112502.
- [40] Y. K. Dewaraja, M. Ljungberg, and J. A. Fessler, "3-D Monte Carlo-based scatter compensation in quantitative I-131 SPECT reconstruction," *IEEE Trans. Nucl. Sci.*, vol. 53, no. 1, pp. 181–188, Feb. 2006.
- [41] G. Kayal et al., "Generation of clinical ^{177}Lu SPECT/CT images based on Monte Carlo simulation with GATE," *Phys. Medica*, vol. 85, pp. 24–31, May 2021.
- [42] M. Ljungberg, S. E. Strand, and M. A. King, "The SIMIND Monte Carlo program," *Monte Carlo Calculation in Nuclear Medicine: Applications in Diagnostic Imaging*. Bristol, U.K.: IOP Publishing, 1998, pp. 63–145.
- [43] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. W. Tsui, "4D XCAT phantom for multimodality imaging research," *Med. Phys.*, vol. 37, no. 9, pp. 4902–4915, Sep. 2010.
- [44] D. E. Bergeron and J. T. Cessna, "An update on 'dose calibrator' settings for nuclides used in nuclear medicine," *Nucl. Med. Commun.*, vol. 39, no. 6, p. 500, 2018.
- [45] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
- [46] S. Y. Chun, M. P. Nguyen, T. Q. Phan, H. Kim, J. A. Fessler, and Y. K. Dewaraja, "Algorithms and analyses for joint spectral image reconstruction in Y-90 bremsstrahlung SPECT," *IEEE Trans. Med. Imag.*, vol. 39, no. 5, pp. 1369–1379, May 2020.
- [47] M. P. Nguyen, H. Kim, S. Y. Chun, J. A. Fessler, and Y. K. Dewaraja, "Joint spectral image reconstruction for Y-90 SPECT with multi-window acquisition," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. (NSS/MIC)*, Oct. 2015, pp. 1–4.
- [48] S. Y. Chun, H. Lim, J. A. Fessler, and Y. K. Dewaraja, "On parameter selection for joint spectral reconstruction in Y90 SPECT," in *Proc. IEEE Nucl. Sci. Symp. Med. Imag. Conf. Proc. (NSS/MIC)*, Nov. 2018, pp. 1–4.
- [49] M. E. Daube-Witherspoon et al., "PET performance measurements using the NEMA NU 2–2001 standard," *J. Nucl. Med.*, vol. 43, no. 10, pp. 1398–1409, Oct. 2002.
- [50] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.